Artificial Intelligence Policy: the Security Dimension

UDC: 323:327.172 DOI: https://doi.org/10.15421/172308

Radionova Iryna
Dr.Sc., Full Prof., https://orcid.org/0000-0002-8863-7609, radionova.irina@gmail.com
H. S. Skovoroda Kharkiv National Pedagogical University (Kharkiv, Ukraine)

Abstract

Many countries of the world, especially those considered as 'leaders', are accelerating the pace of development and application of artificial intelligence (AI) in the environment of intensifying global competition. The aim of the article is to define the security dimension in AI policy. We note that advanced AI can drive a profound change in the history of life on Earth. We find that national and international policies in the field of AI should take into account its potential transformative capacity, its dynamics and threats. The policies of the leading countries in the field of AI are beginning to be structured and formalized at national level, while at the same time the dangers of using AI are being identified and countermeasures are being initiated. The focus here is on the experience of the US, the EU, the People's Republic of China and the UK, as well as the realities of Ukraine. Among the main security dimensions, we highlight the threat to the survival of humankind; AI ethics; geopolitical competition; the threat of a major war; the interaction between private and public sectors in the field of development and application of AI; and the use of AI to strengthen autocracy or democracy. We argue that, same as for any technology, the consequences of use of AI are ambivalent. Interventions in the field of AI are carried out by fundamentally different agents. These agents are able to both act in concert and compete to varying degrees and in various combinations, focus on long-term, medium-term and short-term consequences. We establish that the social consequences of development and use of AI are the domain and responsibility of policymakers, which must take into account the opinions of experts and specialists in the field. Implementation of safety regulations requires significant time and resources. Governments need to continue expanding their abilities to monitor the safety of AI development and its impacts on society. At the same time, there emerges a paradoxical threat that the emphasis on security in the field of AI may lead to strengthening of the administrative and command apparatus of the state, to authoritarian and even totalitarian tendencies. We establish that international cooperation is required in the field of AI regulation, technology, ethics, stress test protocols, clarification of the risks of AI development and application in varying spheres of activity.

Keywords: authoritarianism, global competition, democracy, threat to survival of humankind, security dimension of AI policy, ethics of AI, strategy of development of AI in Ukraine, artificial intelligence (AI)

Політика у сфері штучного інтелекту: безпековий вимір

Радіонова Ірина
Харківський національний педагогічний університет імені Г.С. Сковороди (Харків, Україна)

Анотація

Країни світу, особливо країни-лідери, у загостреній глобальній конкуренції пришвидшують темпи розвитку та застосування штучного інтелекту (ШІ). Метою статті є визначення безпекового виміру політики у сфері ШІ. Зазначено, що просунутий ШІ може спровокувати глибоку зміну в історії життя на Землі. З'ясовано, що політика національного та міжнародного рівня у сфері ШІ повинна враховувати його потенційний трансформативний потенціал, динаміку та небезпеки. Політика країн-лідерів у сфері ШІ починає формалізуватися та структуруватися на національному рівні, при цьому ідентифікуються небезпеки застосування ШІ та ініціюється протидія ним. Зосереджено увагу на досвіді США, ЄС, КНР та Великої Британії, а також реаліях України. Серед основних безпекових вимірів визначено безпеку виживання людства; етику ШІ; геополітичну конкуренцію; загрозу великої війни; співвідношення приватного та державного секторів у сфері розробки та застосування ШІ; застосування ШІ з метою посилити автократію чи демократію. Аргументовано, що як будь-яка технологія ШІ амбівалентний за наслідками свого застосування. Інтервенції у сфері ШІ чинять принципово різні агенти дії. Вони можуть як діяти солідарно, так і конкурувати у різний ступінь і у різних комбінаціях, орієнтуватися на довгострокові, середньострокові та короткострокові наслідки. Встановлено, що соціальні наслідки використання та розвитку ШІ визначальною мірою є сферою діяльності та відповідальності політиків, котрі повинні враховувати думки експертів та спеціалістів у галузі. Впровадження безпекових норм потребує значного часу та ресурсів. Уряди потребують нарощування можливостей здійснювати моніторинг безпеки розвитку ШІ та його впливів на суспільство. При цьому виявлено загрозу, що у парадоксальний спосіб акцент на безпеці у сфері ШІ може призвести до посилення адміністративно-командного апарату держави, до авторитарних та навіть тоталітарних тенденцій. Доведено, що потрібна міжнародна співпраця у сфері регуляції ШІ, технологій, етики, протоколів стрес-тестів, прояснення ризиків розробки та застосування ШІ в різних областях діяльності.

Ключові слова: авторитаризм, глобальна конкуренція, демократія, безпека виживання людства, безпековий вимір політики у сфері ШІ, етика ШІ, стратегія розвитку ШІ в Україні, штучний інтелект (ШІ)

Стаття надійшла / Article arrived: 29.01.2023 Схвалено до друку / Accepted: 27.02.2023

Вступ. Розвиток штучного інтелекту (ШІ) останніми роками відбувається прискореними темпами. Asilomar AI Principles (2017) підкреслюють, що просунутий ШІ може спровокувати глибоку зміну в історії життя на Землі. Планування та керування розвитком ШІ потребує пропорційних обережності та ресурсів. Також зазначається, що суперінтелект необхідно розвивати лише на основі спільних етичних ідеалів на користь всього людства на противагу одній державі чи організації. У стосунках із ШІ культурне запізнення, тобто запізнення трансформації таких складових як цінності, переконання, закони, форми правління (У.Ф. Огборн) може призвести до катастрофічних небезпек та наслідків, масштаби яких ми не можемо прогнозувати. Н. Бостром пропонує гіпотезу «вразливого світу» (vulnerable world hypothesis) та попереджає про можливість «чорної кулі» (back ball) у результатах наукових досліджень, коли політики вже не встигатимуть за подіями (Bostrom, 2019).

Вітчизняні дослідники фокусують увагу на використанні засобів ШІ у самій політиці. При цьому І. Сабов (2018) підкреслює, що не існує комплексного дослідження застосування засобів ШІ у політиці. Більшість науковців опрацьовує певні аспекти використання ШІ. В.В. Костенко (2022) аналізує національні стратегії розвитку ШІ, з'ясовує їх цілі та можливості досягнення цих цілей. Значну увагу вітчизняні дослідники приділяють правовому регулюванню у сфері ШІ. Ю.В. Кривицький (2021) висвітлює ШІ як інструмент правової реформи. С. Лисенко (2020) розкриває необхідність та підсумовує досвід правового регулювання у сфері ШІ, розглядає можливість набуття ШІ статусу самостійного суб'єкта права, інформаційну безпеку технології блокчейн. І.І. Онищук (2020) формулює теоретико-прикладні та етичні засади правового регулювання технології ШІ.

Політика національного та міжнародного рівня у сфері ШІ повинна враховувати його потенційний трансформативний потенціал, динаміку та небезпеки. Методологія якісного аналізу небезпек дозволяє визначити можливі види ризиків, потенційні області ризику та фактори, що впливають на рівень ризику.

Метою статті є визначення безпекового виміру політики у сфері ШІ.

Результати дослідження. Політика країн-лідерів у сфері ШІ починає формалізуватися та структуруватися на національному рівні. Зараз більшість розвинутих країн вже має засадничі плани щодо розвитку ШІ. Вони можуть бути оформлені як цілісні національні стратегії, а можуть існувати у вигляді низки документів, присвячених окремим напрямам діяльності. Серед них: джерела та порядок фінансування, наукові дослідження, підготовка кадрів, рівень залучення державного та приватного секторів, розробка стандартів та нормативних вимог, законодавчої бази та інше. Не усі стратегії спрямовані на ідентифікацію та протидію небезпекам застосування ШІ (Костенко, 2022). Проте претенденти на роль світових лідерів у сфері ШІ фокусують увагу саме на цьому комплексному питанні.

Відразу зауважимо, що концептуалізація поняття ШІ не універсалізована. У проєкті Стратегії розвитку штучного інтелекту в Україні на 2022 – 2030 роки ШІ визначається як «система алгоритмів і програм для генерації нових знань і розв'язання творчих завдань, створених і контрольованих штучною свідомістю обчислювальної машини» (Щодо проєкту Стратегії, 2022, с. 84). Запланована опція деактивації системи, якщо вона поводиться непередбачувано. Розробники стратегії підкреслили, що надане розуміння штучної свідомості узгоджується з Принципами відповідального застосування ШІ Стратегії НАТО щодо ШІ.

Можна визначити чотири фактори, що є драйверами розвитку ШІ: hardware у формі чипів для тренування та виконання алгоритмів ШІ; біг дата як інформація для алгоритмів ШІ; дослідження та розвиток алгоритмів; комерційна екосистема ШІ (Ding, 2018, p. 4). Технології ШІ постійно розвиваються. Сьогодні йдеться про так званий «слабкий» (narrow or weak) ШІ, або ШІ з обмеженими можливостями. Такий ШІ може бути тренований задля виконання конкретного завдання і не може тренувати сам себе, або бути застосований без попереднього коригування і тренування до інших проблем. Саме такий вид ШІ ближчими роками буде все ширше використовуватися у різних сферах життя суспільства. Широке опитування вчених, які працюють у сфері ШІ, дало медіанний показник – ШІ досягне людського розумового рівня приблизно через 35 років (від 2018 року) (Future Proof, 2021, p. 24). І головне – на людському рівні ШІ не зупиниться. Зрозуміло, що технології такого рівня несуть людству як принципово нові можливості в усіх сферах життя, так і принципово нові небезпеки. Треба зазначити, що окремі безпекові виміри можуть визначатися не тільки на теоретичному рівні, а й на рівні розробки національної політики чи у сфері міжнародної співпраці. Підкреслимо, що для нас порядок розгляду окреслених вимірів обумовлений логікою викладу матеріалу, а не спробою побудувати ієрархію небезпек. На наш погляд, усі безпекові виміри є суттєво взаємопов'язаними та взаємозалежними, що призводить до їх посилення та ускладненої взаємогри. Відповідь на безпекові виклики застосування ШІ найперше повинні надати лідери у сфері розвитку ШІ. Ми розглянемо досвід США, КНР, ЄС, Великої Британії.

Засадниче безпекове питання щодо розвитку та застосування ШІ це безпека виживання людства. Людина, як найрозумніша жива істота, остаточно знищила велику кількість представників флори та фауни, призвела до екоциду. У XX столітті людина

довела, що безжально може масово вбивати собі подібних. Людина не оцінювала ступінь вразливість природи через свою діяльність доки не вдалася взнаки та шкода, якої вже було завдано навколишньому середовищу. Зараз вона створює інтелект, який може перевищити її здібності. Ми не можемо гарантувати собі, що розвинутий ШІ не перегляне цінність життя самої людини (як ми самі неодноразово робили), чи не вирішіть, що ми як вид є зайвими на планеті. У своїй назві Центр дослідження сфери ШІ у Стенфордському університеті акцентує безпекову людиноцентричність ШІ – Human-Centered Artificial Intelligence (HAI).

Етика ШІ також постає одним з безпекових вимірів. Все більше систем, інституцій та організацій у своїй роботі покладаються на алгоритми (фінансові інституції, системи охорони здоров'я та освіти, судова система, урядові організації, бізнес та інші). Вибір, який пропонується індивіду різними пошуковими системами, також роблять алгоритми. Концепт алгоритму, як і ШІ, не універсалізовано. Дискусії щодо етики алгоритмів розглядають алгоритми, які використовуються для перетворення даних на докази, що дають певний результат; які можуть ініціювати та мотивувати дії, які мають етичні наслідки. Алгоритмічна діяльність висуває етичні проблеми переконливості доказів; помилкових доказів; небажаних упереджень; несправедливих результатів; дискримінації; відповідальності за трансформаційні ефекти, що призводять до викликів для автономності та інформаційної конфіденційності; непрозорості прийнятих рішень; визначення морально відповідальних за наслідки діяльності, тригером якої були алгоритми (Tsamados et al., 2022, p. 216, 225-226). Ці проблеми на мікрорівні стосуються застосування конкретного алгоритму, а на макрорівні ставить етичні питання діяльності ШІ як такого.

Роблячи перші кроки у розробці глобальних етичних стандартів для ШІ Генеральна конференція ЮНЕСКО (листопад 2021 року) ухвалила основні напрями регулювання поведінки ШІ. Серед них: захист даних, соціальна оцінка і масове спостереження, контроль, захист навколишнього середовища. Рішення підтримали 193 країни, включаючи Україну (Щодо проєкту Стратегії, 2022, с. 89).

ЄС покладається перш за все на свої культурні уподобання та високі стандарти захисту від соціальних ризиків з боку ШІ. Усі ШІ стейкхолдери у своїй діяльності повинні керуватися основними вимогами ЄС: human agency та нагляд; міцність та безпека; конфіденційність та управління даними; прозорість; різноманітність, недискримінація та справедливість; суспільний добробут; екологія; звітність (EU guidelines, 2019, p. 3). ЄС наголошує на необхідності координації зусиль як на рівні ЄС загалом, так і на національному рівні країн-членів ЄС (EU guidelines, 2019, p. 6-7).

У 2022 році Конгрес США прийняв закон, що визначає регуляторний режим етики ШІ, як основну вимогу національної стратегії США (Artificial intelligence Capabilities and Transparency). Він оцінюється як «коротший» у порівнянні з законами ЄС, але є початком розробки етичних питань ШІ на рівні федерального законодавства (Griffin, 2022).

Велика Британія розглядає себе як глобального лідера в етичних стандартах для роботизованих систем та систем ШІ. Лідерство саме у цій найважливішій сфері може бути розширене на усю сферу регуляції ШІ. Вирішення такого завдання потребує значних зусиль щодо відповідального керування та регуляції. Уряд Великої Британії демонструє проактивну позицію у сфері розробки політики керування ШІ. Формується структура інституцій, вносяться зміни у законодавство задля реалізації амбітного завдання зробити з Великої Британії «найкраще місце у світі» для дослідження та застосування ШІ (Future Proof, 2021, p. 24).

Уряд КНР планує до 2025 року впровадити початкові безпекові закони та регуляції, етичні норми та оцінку можливостей контролю ШІ. До 2030 року КНР вдосконалить ці напрацювання, а також політику у сфері ШІ загалом. Дж. Дінг підкреслює, що розширених пояснень уряд КНР не надавав. Дж. Дінг наголошує закритість КНР у питаннях етичних лімітів застосування ШІ, що викликає стурбованість. Він також посилається на лімітовану присутність китайських вчених на міжнародних заходах присвячених регуляції ШІ. Разом з тим, є сигнали що Китай пов'язує своє лідерство у сфері ШІ з лідерством і в питаннях безпеки та етики його застосування (Ding, 2018, p. 20-21).

Треба відзначити наголос на етичному безпековому вимірі, який запропоновано для виконавчих органів управління та регулювання ШІ в Україні. Серед основних принципів пріоритет добробуту людини, заборона на заподіяння шкоди з ініціативи систем ШІ, підконтрольність людині, проєктована безпека даних (Щодо проєкту Стратегії, 2022, с. 90).

З етичним виміром безпеки перетинається загроза великої війни. Просунутий ШІ (advanced AI) може надати нові можливості першого удару, нові деструктивні можливості, що будуть спокушати розв'язати нову війну. В той самий час просунутий ШІ може зробити динаміку будь-якої кризи більш комплексною та непередбачуваною, з більшою швидкістю ескалації ніж та, якою люди можуть керувати. Відтак зросте ризик ненавмисної війни.

Окремої уваги та пересторога потребує застосування просунутого ШІ щодо ядерної зброї. Представники наукової спільноти Великої Британії оцінюють ризики останнього як надто великі. Тож, пропонують унеможливити застосування систем ШІ

у сферах комунікацій, командування, контролю щодо ядерної зброї. Велика Британія повинна ініціювати застосування цієї норми на міжнародному рівні через ризики випадкового застосування ядерної зброї (Future Proof, 2021, p. 27). Одночасно необхідно чітко визначити на міжнародному рівні що таке «летальні автономні системи озброєння» (lethal autonomous weapons systems). І знов таки контролювати та гарантувати неможливість для просунутого ШІ діяти у сфері застосування даної зброї автономно (Future Proof, 2021, p. 28-29).

Зазначені безпекові виміри фактично вже підіймають питання загострення глобальної конкуренції у сфері ШІ. Розвиток та застосування ШІ постають як фронтир глобальної конкуренції. Причому конкурують усі з усіма. А. Дафо попереджає про системну небезпеку ерозії цінностей через надзвичайне загострення змагання, що може спровокувати інструменталізацію загальнолюдських цінностей через боротьбу за більшу владу та багатство (Dafoe, 2018, p. 7-8).

Очевидно, що країни світу мають різні стартові можливості. Отож група лідерів, тих хто буде розробляти ШІ, є дуже вузькою. Списки країн лідерів відрізняються один від одного, але серед них США (глобальний лідер у сфері дослідження, застосування та розробки ШІ), КНР, ЄС (саме як союз), Велика Британія та Японія (Ding, 2018, p. 12). Зрозуміло, що кожна їхня дія впливає на долю всього людства. Принциповою за своєю сутністю є конкуренція між США та КНР. Наприклад, у 2017 році Білий дім заблокував придбання китайським інвестиційним фондом (пов'язаним з державою) американської компанії напівпровідників. Це сталося четвертий раз за всю історію США, коли американський президент заблокував корпоративне придбання через питання національної безпеки (Ding, 2018, p. 17). ЄС також уважно відстежує діяльність КНР. У тому ж 2017 році ЄС унеможливлює прямі закордонні інвестиції з боку компаній, які належать державі, у сфери критичних технологій, серед яких ШІ, робототехніка, напівпровідники, кібербезпека, космічні та ядерні технології; технології потенційного подвійного використання. Імпліцитно йдеться перш за все про китайські кампанії (Ding, 2018, p. 17).

КНР у лютому 2017 року прийняла мегапроєкт «Artificial Intelligence 2.0». У липні 2017 року Державна рада КНР проголосила національну стратегію розвитку ШІ до 2020 року («New Generation AI Development Plan»). (Ding, 2018, p. 7 - 10). Документи продемонстрували амбіцію КНР стати світовим лідером у сфері ШІ. За такої умови КНР перш за все орієнтується у своїй стратегії на США та білатеральну конкуренцію (Ding, 2018, p. 4). КНР визначає три стратегічні фази розвитку ШІ: спочатку країна наздоганяє лідерів світового розвитку у сфері ШІ (до 2020); потім стає одним з них (до 2025) і третя фаза – досягає першості у сфері інновацій ШІ (до 2030) (Ding, 2018, p. 10).

США та КНР очевидно будуть конкурувати один з одним у сфері ШІ. Однак безпека виживання людства призводить і до наукової співпраці цих країн. Так, навіть за умов міждержавного напруження США та КНР мають за підсумками 2021 року найбільшу кількість публікацій щодо ШІ, створених у колаборації (до речі наступною за кількістю групою є колаборації Китаю з Великою Британією) (Artificial Intelligence Index, 2022, p. 3).

Безсумнівно зростає глобальна конкуренція за людський капітал. США мають сталу історичну традицію спокуси та найму кращих фахівців та обдарованої молоді. Зараз звучать пропозиції змінити імміграційну політику на користь фахівців у сфері ШІ (див., наприклад: McCormic, 2021). КНР надає можливість талановитим студентам здобути освіту в кращих навчальних закладах світу, але водночас має жорсткі правила їхнього повернення у країну. У 2007 році КНР прийняла програму «Десяти тисяч талантів» (Ten Thousand Talents programme) зі значним фінансуванням для наймання кращих фахівців у сфері ШІ по всьому світу (Ding, 2018, p. 19). Що може протиставити Україна цьому полюванню на фахівців високого класу, як ми будемо захищати наш потужний кадровий потенціал – питання, на які потрібна найскоріша відповідь (кадровий потенціал України див.: Artificial Intelligence in Ukraine, 2021).

Наступний безпековий вимір безпосередньо пов'язаний з глобальною конкуренцією, оскільки визначає співвідношення приватного та державного секторів у сфері розробки та застосування ШІ. Одним з перших постає питання про готовність уряду тієї чи іншої країни відповідально використовувати ШІ у своїй діяльності. Обсяг статті не дозволяє докладно зупинитися на даному вимірі. Однак звернемо увагу на Oxford Government AI Readiness Index 2022. Слід зазначити, що Міністерство цифрової трансформації України ставить завдання найближчим часом увійти до топу 20 країн за цим індексом (Artificial Intelligence in Ukraine, 2021, с. 11).

ЄС не визначає себе у координатах приватний / державний сектор розвитку ШІ. Якщо стратегія США ґрунтується перш за все на ініціативах приватного сектору та саморегуляції, стратегія КНР на державному керівництві, координації державних та приватних інвестицій у розвиток ШІ (під наглядом держави), то ЄС розвиває «людино центричний» підхід до ШІ та орієнтується на ціннісну регуляцію розвитку сфери ШІ (human-centric AI) (EU guidelines, 2019, p. 3).

За даними Artificial Intelligence Index Report 2022 у 2021 році приватні інвестиції у ШІ склали близько 93,5 мільярдів доларів, що удвічі більше ніж у 2020 році. Причому важливо підкреслити, що чисельність

нових компаній, у які вкладаються гроші, постійно зменшується (Artificial Intelligence Index, 2022, p. 3). Тобто, лідери вже очевидні.

КНР використовує державний контроль за найбільшими хай-теками, а також протекціонізм щодо біг дата. КНР ґрунтується на підході техно-націоналізму, який захищає китайські компанії від конкуренції. Техно-націоналізм дозволяє компаніям, які не можуть конкурувати на глобальному ринку, бути успішними в Азії та Африці, а також домінувати на великому внутрішньому китайському ринку. Задля посилення протекціонізму КНР використовує оригінальні стандарти у сфері ШІ (smart manufacturing, cloud computing, industrial software, big data). У 2017 році був прийнятий закон у сфері кібербезпеки, який забороняє зберігання даних китайських користувачів поза межами Китаю (Ding, 2018, p. 18-19).

У проєкті Стратегії розвитку штучного інтелекту в Україні на 2022 – 2030 роки наголошується потреба в адекватній відповіді на виклики й загрози, що пов'язані з технологічним відставанням України від розвинутих держав світу (Щодо проєкту Стратегії, 2022, с. 82). Стратегія розкриває шляхи уникнення технологічної залежності України у сфері ШІ. «Україна значно відстає від провідних держав у темпах і обсягах упровадження розробок у сфері ШІ, але має необхідний потенціал фундаментальних знань для здійснення прориву в створенні цілковито нових технологій у сфері ШІ» (Щодо проєкту Стратегії, 2022, с. 88). Очевидно, що для цього потрібні державні програми та фінансування.

Головним у співвідношенні державного та приватного сектору є питання прозорості діяльності. Як співвідноситься вимога прозорості діяльності у сфері ШІ з питаннями інтелектуальної власності, комерційної тайни та іншими; як взагалі застосувати вимогу прозорості до діяльності надзвичайно складної системи ШІ (виникнення ефекту black box thinking). Універсальних відповідей на зазначені питання поки не існує.

Ще один безпековий вимір - застосування ШІ з метою посилення авторитаризму чи демократії. А. Дафо звертає увагу на небезпеку «міцного тоталітаризму» (robust totalitarianism) через посилення можливостей соціальних маніпуляцій, надзвичайне відстежування діяльності через повсюдні фізичні датчики та цифрові сліди (Dafoe, 2018, p. 7). В руках еліт та лідерів з'являються нові механізми зосередження влади та контролю. Досвід реагування на пандемію COVID-19 продемонстрував, як демократичні режими вдаються до авторитарних заходів у надзвичайних умовах. Не має підстав для впевненості, що демократії можуть встояти перед спокусою використати «чарівну паличку» ШІ задля нібито спільного блага, якщо ціною постануть авторитарні тенденції. Сам акцент на безпеці, у тому рахунку і безпеці у сфері ШІ (у парадоксальний спосіб), зазвичай означає посилення адміністративно-командного апарату держави, що призводить до авторитарних та навіть тоталітарних тенденцій.

КНР, як авторитарна країна, вже створила систему соціального кредиту («social credit system»). Система повинна постійно моніторити та оцінювати діяльність кожної громадянки, кожного громадянина та класифікувати рівень їхньої надійності. Бал довіри впливає на можливість отримати кредит, шанси отримати роботу, місце для дитини в системі освіти та інше. Саме ШІ дозволяє збирати та аналізувати біг дата для зазначеної системи.

Висновки. ШІ може пришвидшити економічне зростання, підвищити якість життя, допомогти розв'язати важливі завдання в різних областях діяльності. Але ризики (уявні та неуявні сьогодні), які він несе, є катастрофічними. Просунутий ШІ може не керуватися людськими цінностями, що призведе до вимирання людства, або зменшення цінності людини. Країни світу, особливо країни-лідери, у загостреній глобальній конкуренції пришвидшують темпи розвитку та застосування ШІ. Ми зосередили увагу на досвіді США, ЄС, КНР та Великої Британії, а також реаліях України. Як будь-яка технологія ШІ амбівалентний за наслідками свого застосування. Зазначимо, що соціальні наслідки використання та розвитку ШІ визначальною мірою є сферою діяльності та відповідальності політиків, котрі у своїх діяльності повинні враховувати думки експертів та спеціалістів у галузі. Впровадження безпекових норм потребує значного часу та ресурсів. Уряди потребують нарощування можливостей здійснювати моніторинг безпеки розвитку ШІ та його впливів на суспільство. Серед основних безпекових вимірів ми визначили безпеку виживання людства; етику ШІ; геополітичну конкуренцію; загрозу великої війни; співвідношення приватного та державного секторів у сфері розробки та застосування ШІ; застосування ШІ з метою посилити автократію чи демократію.

Керування ШІ повинно збільшити шанси того, що люди, які створюють та використовують ШІ, мають цілі, заохочення, світогляд, професійну підготовку, ресурси, підтримку та інституції необхідні для праці на користь людства (Dafoe, 2018, p. 6-7). Однак ми розуміємо, і про це свідчить історія, що означений ідеал не буде втілений у життя. Тож, інтервенції у сфері ШІ чинять принципово різні агенти дії: держави з різними формами політичного устрою і різними стратегіями розвитку, гіганти хай-тек бізнесу і стартапи, університети та наукові лабораторії, цивільні та військові, інші агенти. Вони можуть як діяти солідарно, так і конкурувати у різний ступінь і у різних комбінаціях, орієнтуватися на довгострокові, середньострокові та короткострокові наслідки. Означену мапу можна майже нескінченно ускладнювати. Відповідно потрібна міжнародна співпраця у сфері регуляції ШІ, технологій, етики,

протоколів стрес-тестів, прояснення ризиків розробки та застосування ШІ в різних областях діяльності. Новий імператив суспільної політики – «чини так, щоб наслідки твоєї діяльності узгоджувалися з продовженням автентичного людського життя на землі» (Йонас, 2001, с. 27).

БІБЛІОГРАФІЧНІ ПОСИЛАННЯ

Бостром, Н. (2020). *Суперінтелект. Стратегії і небезпеки розвитку розумних машин*. Пер. з англ. К.: Наш Формат.

Йонас, Г. (2001). *Принцип відповідальності. У пошуках етики для технологічної цивілізації*. Пер. з нім. К.: Лібра.

Костенко, О. В. (2022). Аналіз національних стратегій розвитку штучного інтелекту. *Інформація і право*, 2(41), 58-69. https:/doi.org/10.37750/2616-6798.2022.2(41).270365

Кривицький, Ю. В. (2021). Штучний інтелект як інструмент правової реформи: потенціал, тенденції та перспективи. *Науковий вісник Національної академії внутрішніх справ*, 2(119), 90-101. https://doi.org/10.33270/01211192.90

Кронівець, Т. М., & Тимошенко, Є. А. (2022). Правові та ціннісні особливості феномену штучного інтелекту як елементу правової дійсності. *Науковий вісник Херсонського державного університету. Серія Юридичні науки*, 2, 20-22. https://doi.org/10.32999/ksu2307-8049/2022-2-4

Лисенко, С. (2020). Майбутні тенденції інформаційної безпеки відповідно до технологій штучного інтелекту. *Публічне урядування*, 5(25), 151-163. https://doi.org/10.32689/2617-2224-2020-3(23)-151-163

Онищук, І. І. (2020). Правовое-регулювання технологій штучного інтелекту: теоретико-прикладні та етичні засади. *Наукові записки Інституту законодавства Верховної Ради України*, 3, 50-57. https://doi.org/10.32886/instzak.2020.03.06

Щодо проєкту Стратегії розвитку штучного інтелекту в Україні на 2022 – 2030 рр. (2022). *Artificial Intelligence*, 1, 75-157.

Artificial Intelligence in Ukraine. (2021). Ministry of Digital Transformation of Ukraine. Retrieved from https://niss.gov.ua/sites/default/files/2021-10/ai_ukraine_v2.pdf

Artificial Intelligence Index Report 2022. (2022). *HAI Stanford University Human-Centered Artificial Intillegence.* Retrieved from https://hai.stanford.edu/research/ai-index-2022

Asilomar AI Principles. (2017). Retrieved from https://futureoflife.org/open-letter/ai-principles/

Bostrom, N. (2019). The Vulnerable World Hypothesis. *Global Policy*, 10(4), 455-476.

Dafoe, A. (2018). AI Governance: A Research Agenda. *Center for the Governance of AI*. University of Oxford. Retrieved from https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf

Ding, J. (2018). *Deciphering China's AI Dream. The Context, Components, Capabilities, and Consequences of China`s Strategy to Lead the World in AI*. Retrieved from https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation. (2019). Retrieved from https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf

Future Proof. The Opportunity to Transform the UK`s Resilience to Extreme Risks. (2021). *The Centre for Long-term Resilience*. Retrieved from https://www.longtermresilience.org/futureproof

Griffin, W. (2022, February 15). America Must Win the Race for A.I. Ethics. *Fortune*. Retrieved from https://fortune.com/2022/02/15/america-must-win-the-race-for-a-i-ethics-tech-artificial-intelligence-politics-biden-dod-will-griffin/

Mccormic, D. (2021, February 19). America Needs Smart Immigration Reforms to Win the Race for Global Talent. *Fortune*. Retrieved from https://fortune.com/2021/02/19/immigration-reform-biden-us-global-talent-race/?queryly=related_article

Oxford Government AI Readiness Index. (2022). Retrieved from https://www.oxfordinsights.com/government-ai-readiness-index-2022

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). The Ethics of Algoritms: Key Problems and Solutions. *AI&SOCIETY*, 37, 215-230.

REFERENCES

Artificial Intelligence in Ukraine. (2021). Ministry of Digital Transformation of Ukraine. Retrieved from https://niss.gov.ua/sites/default/files/2021-10/ai_ukraine_v2.pdf

Artificial Intelligence Index Report 2022. (2022). *HAI Stanford University Human-Centered Artificial Intillegence*. Retrieved from https://hai.stanford.edu/research/ai-index-2022

Asilomar AI Principles. (2017). Retrieved from https://futureoflife.org/open-letter/ai-principles/

Bostrom, N. (2019). The Vulnerable World Hypothesis. *Global Policy*, 10(4), 455-476.

Bostrom, N. (2020). *Super Intelligence. Strategies and Dangers of the Development of Intelligent Machines*. Trans. from English. K.: Our Format.

Dafoe, A. (2018). AI Governance: A Research Agenda. *Center for the Governance of AI*. University of Oxford. Retrieved from https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf

Ding, J. (2018). *Deciphering China's AI Dream. The Context, Components, Capabilities, and Consequences of China`s Strategy to Lead the World in AI*. Retrieved from https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation. (2019). Retrieved from https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf

Future Proof. The Opportunity to Transform the UK`s Resilience to Extreme Risks. (2021). *The Centre for Long-term Resilience*. Retrieved from https://www.longtermresilience.org/futureproof

Griffin, W. (2022, February 15). America Must Win the Race for A.I. Ethics. *Fortune*. Retrieved from https://fortune.com/2022/02/15/america-must-win-the-race-for-a-i-ethics-tech-artificial-intelligence-politics-biden-dod-will-griffin/

Kostenko, O. V. (2022). Analysis of National Strategies for the Development of Artificial Intelligence. *Information and Law*, 2(41), 58-69. https:/doi.org/10.37750/2616-6798.2022.2(41).270365

Kronivets, T. M., & Tymoshenko, E. A. (2022). Legal and Value features of the Phenomenon of Artificial Intelligence as an Element of Legal Reality. *Scientific Bulletin of Kherson State University. Series Legal Sciences*, 2, 20-22. https://doi.org/10.32999/ksu2307-8049/2022-2-4

Kryvytskyi, Yu. V. (2021). Artificial Intelligence as a Tool of Legal Reform: Potential, Trends and Perspectives. *Scientific Bulletin of the National Academy of Internal Affairs*, 2(119), 90-101. https://doi.org/10.33270/01211192.90

Lysenko, S. (2020). Future Trends in Information Security According to Artificial Intelligence Technologies. *Public administration*, 5(25), 151-163. https://doi.org/10.32689/2617-2224-2020-3(23)-151-163

Mccormic, D. (2021, February 19). America Needs Smart Immigration Reforms to Win the Race for Global Talent. *Fortune*. Retrieved from https://fortune.com/2021/02/19/immigration-reform-biden-us-global-talent-race/?queryly=related_article

Onischuk, I. I. (2020). Legal Regulation of Piece Intelligence Technologies: Theoretical, Applied and Ethical Ambush. *Scientific Notes to the Institute of Legislation of the Verkhovna Radi of Ukraine*, 3, 50-57. https://doi.org/10.32886/instzak.2020.03.06

Oxford Government AI Readiness Index. (2022). Retrieved from https://www.oxfordinsights.com/government-ai-readiness-index-2022

Regarding the Project of the Strategy for the Development of Artificial Intelligence in Ukraine for 2022-2030. (2022). *Artificial Intelligence*, 1, 75-157.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). The Ethics of Algoritms: Key Problems and Solutions. *AI&SOCIETY*, 37, 215-230.

Yonas, G. (2001). *The Principle of Responsibility. In Search of Ethics for a Technological Civilization*. Trans. from German. K.: Libra.