
АКТУАЛЬНІ ПРОБЛЕМИ РОЗВИТКУ МЕДІА ТА ІТ-СФЕРИ

Медіафорум : аналітика, прогнози,
інформаційний менеджмент :
зб. наук. праць. – Чернівці :
Чернівецький нац. ун-т, 2024. –
Том 14. – С. 275-299

Mediaforum: Analytics, Forecasts,
Information Management:
Collection of Research Articles. – Chernivtsi:
Chernivtsi National University, 2024. –
Vol. 14. – pp. 275-299

<https://doi.org/10.31861/mediaforum.2024.14.275-299>

УДК: 004.8:330.342.24

© Володимир Фісанов¹

© Олександра Гісса-Іванович²

**ТЕХНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ ЯК СКЛАДОВА
ТА ІНСТРУМЕНТАРІЙ РОЗВИТКУ ІНДУСТРІЇ 4.0:
ЗА ЕКСПЕРТНИМИ ОЦІНКАМИ THINK TANKS**

У статті авторами піднімається декілька актуальних тем, пов'язаних з сучасним розвитком штучного інтелекту як складової та інструментарію розвитку індустрії 4.0. Підкреслено, що за версією британського словника Collins у 2023 термін «штучний інтелект» став словом року. Основні положення статті ґрунтуються на спеціальних дослідженнях у сфері ШІ, здійснених такими відомими think tanks, як Інститут Алана Тюрінга, Центр 4-ої Індустріальної революції СЕФ у Давосі за підтримки фірми «Хітачі», а також RAND Corporation. Залучені матеріали доповідей Єврокомісії та розробки з тематики ШІ останніх років в Україні, зокрема у військовому й безпековому напрямках.

275

Автори відзначають також і той факт, що у ХХІ столітті усе більше обертів набирає поки що досить незвична проблема етики штучного інтелекту: постає питання чи спроможні будемо ми у середньостроковій перспективі виробити дієву систему правил, призначених для створення середовища, так би мовити, довіреного розвитку технологій ШІ.

¹ Доктор історичних наук, професор кафедри міжнародних відносин та суспільних комунікацій Чернівецького національного університету імені Юрія Федьковича, Україна. E-mail: v.fisanov@chnu.edu.ua; [https:// orcid.org /0000-0001-6550-8903](https://orcid.org/0000-0001-6550-8903)

² Кандидат політичних наук, консультант-аналітик, фрілансер, Україна. E-mail: aleks.gissa@gmail.com; [https:// orcid.org /0000-0001-9777-7307](https://orcid.org/0000-0001-9777-7307)

Особливу увагу приділено питанням геостратегічного протиборства США і КНР у царині інноваційного розвитку ШІ. У цьому контексті штучний інтелект вимагає оригінальних управлінських рішень на глобальному рівні. Структура управління повинна бути запобіжною, гнучкою, інклюзивною, непроникною та цілеспрямованою. Усім зацікавленим сторонам – державам, інституціям у сфері ІТ, неурядовим акторам і фірмам-розробникам варто, виходячи з цих принципів, сформуванати 3 режими управління, що перетинаються. Перший – для встановлення фактів і консультування урядів щодо ризиків, пов'язаних зі штучним інтелектом; другий для запобігання повній гонці озброєнь між державами на інноваційній основі; третій для управління руйнівними силами технології, яка не схожа до всього, що уже бачив світ у минулому.

Зроблено висновок, що штучний інтелект, за умов його подальшого фреймування та етичного доведення, є лише одним з новітніх напрямків (згадаємо тут теорему Геделя) розвитку розуму людства у ХХІ столітті.

276

— **Ключові слова:** індустрія 4.0, штучний інтелект, Алан Тюрінг, Семюель Батлер, етика, наукові дослідження, українські заклади освіти, прозорість, приватність, відповідальність, соціальна взаємодія, Алфредо Валладео, think tanks.

Artificial Intelligence Technologies as a Component and Toolkit of Industry 4.0 Development: According Think Tanks Expert Evaluations

In the article, the authors raise several topical topics related to the modern development of artificial intelligence as a component and toolkit of Industry 4.0 development. It is emphasized that according to the British Collins dictionary, in 2023 the term “artificial intelligence” became the word of the year. The main provisions of the article are based on special research in the field of AI carried out by such well-known think tanks as the Alan Turing Institute, the Center for the 4th Industrial Revolution of the CEF in Davos with the support of the Hitachi company, as well as the RAND Corporation. The materials of the reports of the European Commission and developments on the subject of AI in recent years in Ukraine, in particular in the military and security areas, have been included.

The authors also note the fact that in the 21st century, a rather unusual problem of the ethics of artificial intelligence is gaining more and more momentum: the question arises whether we will be able to develop an effective

system of rules in the medium term, designed to create an environment, so to speak, of trusted development of technologies AI.

Special attention is paid to the issues of geostrategic confrontation between the United States and the People's Republic of China in the field of innovative development of AI. In this context, artificial intelligence requires original management solutions at the global level. The governance structure should be preventive, flexible, inclusive, transparent and focused. All stakeholders – states, IT institutions, non-governmental actors and development firms – should, based on these principles, form 3 overlapping governance regimes. The first is to fact-find and advise governments on the risks associated with artificial intelligence; the second to prevent a full arms race between states on an innovative basis; the third to control the destructive forces of technology unlike anything the world has seen in the past.

It is concluded that artificial intelligence, under the conditions of its further framing and ethical proof, is only one of the newest directions (recall Gödel's theorem here) of the development of the human mind in the 21st century.

277

Keywords: industry 4.0, artificial intelligence, Alan Turing, Samuel Butler, ethics, scientific research, Ukrainian educational institutions, transparency, privacy, responsibility, social interaction, Alfredo Valladão, think tanks.

Постановка проблеми. Серед технічних інновацій, притаманних розвитку 4ir, з якими сьогодні спроможна (хоча, можливо, і не дуже цього бажає) «зустрітися» пересічна людина, чільне місце займає штучний інтелект. Про його дедалі усе більш неочікувані «кроки» нашою планетою інформують різні пабліки і медіа, про ризики і досягнення у сфері ШІ дискутують під час круглих столів й на сторінках досліджень науковці та експерти.

На часі, до прикладу, розробка нейроморфних комп'ютерів що імітують функції мозку, використовуючи штучні нейрони, які програмуються і забезпечують більш ефективне розпізнавання образів за умов якщо певної інформації обмаль. Або ж реальністю стає професія промт-інженерів, які у змозі якісно запустити генерацію необхідних даних: без детально розробленого запиту штучний інтелект не зможе сформулювати чітку та зрозумілу відповідь. Усе більше набирає обертів і досить незвична у цьому контексті проблема етики штучного інтелекту: постає питання чи спроможні будемо ми у середньостроковій перспективі виробити дієву систему правил, при-

значених для створення середовища, так би мовити, довіреного розвитку технологій ІІІ.

Аналіз останніх досліджень і публікацій. Проблематика Індустрії 4.0 досить активно розробляється в Україні. Свого часу уряд у «Національній стратегії Індустрії 4.0» (2018) визначив курс на цифровізацію терміном на 5 років. Згідно з інноваційною програмою пріоритетними задачами визнано: консолідацію головних стейкхолдерів 4.0 та урядових структур; активне залучення ІТ-сектору до питань передання досвіду глобалізації та кращих бізнес-практик підприємствам і організаціям; інтеграцію до європейських та світових ланцюжків цінностей. Передбачалося також прискорити інтеграцію України до відповідних європейських програм, зокрема з розвитку промислових хайтек (H2020 SME Instrument, EaP Plus, COSME, ENN). Це й доступ до відповідних знань, покращення загальної конкурентоздатності, а також перейняття кращих новаторських практик підприємств. Проте, національні розробники стратегії 4.0 самокритично пишуть, що рівень інтеграції України в ці програми «залишається мізерним» [Національна стратегія, 2014]. Якщо вести мову насамперед про економіку, то перспективою подальших досліджень є розроблення механізмів із підвищення ефективності впровадження Індустрії 4.0 на підприємствах України, оскільки саме вони забезпечують інноваційні зрушення в економіці, проводять НДДКР, створюють додану вартість, наповнюють бюджет регіону та країни [Чайкіна А.О., 2021]. Львівські дослідники А. Завербний і К. Сало проаналізували ключові індикатори запровадження і розвитку Індустрії 4.0 в Україні у контексті євроінтеграції [Завербний А. С., Сало К. Р., 2022].

Важливі аспекти сучасного розвитку «розумного штучного інтелекту» розглядаються у колективній монографії присвяченій цій тематиці, у котрій досліджуються не просто складні питання в обчислюванні чорні скриньки, де автори намагаються не тільки надати відповіді, але й розумно пояснюють, як їх було досягнуто [Стратегія розвитку, 2023].

Однак цілі Індустрії 4.0 вже не у всьому відповідають нагальним потребам сучасного розвитку, оскільки промислова трансформація в контексті переходу до Індустрії 4.0, виявилася «суто технологічною парадигмою», де не враховувалися завдання подолання кліматичної кризи на планеті, а також зняття глибокої соціальної напруги. Десь рік тому появився «Маніфест щодо переходу України до індустрії

5.0». У ньому зокрема підкреслено таке: Індустрія 5.0 як сучасна та приваблива для інвестування в різні галузі модель виробництва, може стати тим «золотим ключиком», що відімкне двері потокам донорських та інвестиційних коштів для відновлення України. Країні потрібна нова, сучасна промисловість, тому що де-індустріалізована, слабка економіка неспроможна забезпечити стійкість в часи потрясінь, що є критичним для європейського континенту та світу в цілому в ситуації, коли під боком в Україні знаходиться агресивний сусід, що прагне «перекроїти» світ [Маніфест, 2023].

Однак у даній статті ми дещо звужуємо сферу дослідницької уваги, акцентуючи на темі розвитку штучного інтелекту як інструменту нової доби у світовій індустрії. Як відзначає експерт Центру 4-ої Індустріальної революції Світового економічного форуму, представник фірми «Хітачі» Соїчі Фуруя (Soichi Furuya), саме штучний інтелект, машинне навчання і пов'язані з ними прикладні сфери, зокрема Інтернет речей, робототехніка й розумні міста є сьогодні провідними креативними платформами для подальшого співробітництва з Центром і СЕФ [Soichi Furuya, 2019]. Нагадаємо, що Hitachi – це одна з найбільш потужних фінансово-промислових груп Японії, до складу котрої входить понад 1100 різних фірм і підприємств.

Іншу актуальну задачу у царині ШІ сформулював Стюарт Рассел, науковий радник діючого у США Дослідницького інституту машинного інтелекту (MIRI) та співавтор провідного підручника зі штучного інтелекту. Він зокрема стверджує у «Довгостроковому майбутньому штучного інтелекту», що ми повинні інтегровано підійти до цього питання. «Моя пропозиція, – каже вчений, – полягає у тому, що ми маємо припинити розробляти штучний інтелект у його простому визначенні лише щодо покращення можливостей систем прийняття рішень. [...] У цивільному будівництві ми не називаємо це «будівництвом мостів, які не падають» – ми просто називаємо це «будівництвом мостів». Звичайно, ми не хочемо, щоб вони впали. І ми повинні так само думати про штучний інтелект: системи ШІ повинні бути розроблені таким чином, щоби їх дії добре узгоджувалися з тим, чого бажають люди. Але це складна невирішена проблема, яка досі не була часткою порядку денного досліджень» [About MIRI]. Коментуючи тему, один з слухачів лекції з проблем ШІ вкрай її загострив: «Один сценарій, який продовжує мене непокоїти, – це те, що

відбудеться, якщо машина набуде самосвідомість, а потім вирішить скрити це, брехати про це, допоки для нас не стане занадто пізно».

Як відзначають експерти у сфері етики ШІ, область досліджень машинної етики (або ж робоетика) виникла нещодавно як підгалузь штучного інтелекту, що зосереджується на завданні забезпечення етичної поведінки штучних агентів (зазвичай званих АМА, Artificial Moral Agents). Вона потребує залучення учасників як з інформатики, так і з філософії. Зосереджуючись на поведінці штучних агентів, цей напрямок вирізняється від попередніх робіт у сфері етики у застосуванні до технологій – в їх використанні людьми та у рідкісних випадках поводження людей з машинами [Shulman Carl, 2009].

280

Слід при цьому особливу увагу звернути на чутливе питання для суспільства щодо «контрольованості» ШІ. Як вважають досвідчені вітчизняні експерти, для штучних інтелектуальних систем, заснованих на неймережах, які можуть мати складну структуру і виробляти непередбачувані результати, особливо важливо володіти методами пояснення ухвалених рішень. Тому українські дослідники мають зосередитися сьогодні на розробці методології та інструментів, які дадуть змогу пояснювати (особливо при підготовці фахівців вищої кваліфікації), як системи ШІ ухвляють свої рішення щоб користувачі могли бачити й розуміти процеси, які відбуваються всередині системи [Бердо Р. С., 2023].

Звернемо увагу зацікавленого читача до іншої фундаментальної праці, яка опублікована за редакцією двох дослідників Стюарта Рассела і Петера Норвіга. Останній на момент видання книги був діючим директором із досліджень Google, а раніше він очолював відділ у дослідницькому центрі NASA, де керував дослідженнями та розробками у сфері штучного інтелекту та робототехніки. А С. Рассел довгий час був професором інформатики та директором Центру інтелектуальних систем Каліфорнійського університету в Берклі. Основна мета видання, котре має й навчальні задачі, передати ті ідеї, що виникли за останні 50 років у сфері штучного інтелекту. Базовою об'єднуючою темою є ідея «розумного агента». Ідеться про відповідного реактивного «машинного агента», який функціонально може бути налаштований на сприйняття дії, а також планування у реальному часі й подальшим поясненням ролі е-навчання як розширення можливостей «дизайнера адаптованого до невідомого середовища» (designer into unknown environments) [Artificial Intelligence, 2010].

Мета статті. На основі останніх наукових досліджень, експертних доповідей провідних think tanks, а також новітніх інноваційних практик у цій сфері в Україні проаналізувати окремі напрямки розвитку штучного інтелекту як важливого інструменту Індустрії 4.0.

Виклад матеріалу. Обернемося трохи назад. Одним із перших у світі хто описав небезпеки з боку машин, машинного інтелекту був Семюель Батлер, відомий англійський письменник позаминулого століття. Він зробив це у романі «Еревхон» (Erewhon, 1872), який є сатирою на вікторіанське суспільство. Дія відбувається у вигаданій країні, котру відкрив і дослідив головний герой. Кілька розділів роману присвячені відкриттю Еревхона й фактично засновані на власному досвіді Батлера у Новій Зеландії, де, будучи юнаком, він працював вівчарем на станції Месопотамія протягом приблизно чотирьох років (1860-64). У 3-х розділах, присвячених темі «панування» машин, Батлер пише про потенційно небезпечні ідеї машинної свідомості та перспективи самовідтворюваних машин. Ще раніше у 1863 р. він написав низку статей для новозеландської The Press під загальною назвою «Дарвін серед машин», де висловив думку що машини спроможні будуть розвивати «самосвідомість» шляхом природного відбору. Насправді, це й були перші строки про штучний інтелект у світовій пресі.

Батлер зокрема пророче зауважує (160 років тому!): «Проте з кожним днем машини здобувають перевагу над нами; з кожним днем ми все більше підкоряємося їм; більше людей щодня прив'язані як раби, щоб доглядати за ними, більше людей щодня присвячують енергію всього свого життя розвитку механічного життя. Підсумок – просто це питання часу, але те, що настане час, коли машини отримають справжнє панування над світом і його мешканцями (курсив наш – В.Ф., О.Г.), – це те, про що жодна людина справді філософського розуму не може сумніватися навіть на мить» [Butler Samuel, 1863]. Про «жахи індустріалізму» писав свого часу Олдос Гакслі. Простежуючи походження соціальних «течій» регресу і милосердя – в рамках загальноприйнятої матеріалістичної схеми утилітаризму – до Прорітництва вісімнадцятого століття і «до жахів індустріалізму» (the horrors of industrialism) в дев'ятнадцятому столітті, Гакслі підсумовує: «Технологічний прогрес просто надав нам більш ефективні засоби для руху назад» [Milz Manfred, 2023].

Якщо говорити відверто, то тема «розумних машин» ще раз і вже остаточно виринула під час Другої світової війни, коли групі Алана Тюрінга вдалося зламати нацистські коди внаслідок чого були нанесені нищівні удари по німецькому воєнно-морському флоту. У 1995 р. британський письменник Роберт Гарріс описав це інтелектуальне протиборство в романі «Еніґма» (книга отримала розголос від Найробі до Торонто й від Лондона до Веллінгтона; була екранізована в 2001 р.). Головний герой – молодий математик Том Джеріко (його прообразом був А.Тюрінг) працює у центральному офісі британської криптології в Блетчлі Парк, доводячи себе до фізичного та психічного виснаження. Врешті-решт зусилля увінчуються успіхом – створюється «меню» для роботи криптоаналітичних «бомб».

282

Зауважимо, що саме тоді з'являється проривна ідея про можливості створення у майбутньому машинного інтелекту людського рівня. Однак і таке небачене досі досягнення не є кінцевим пунктом призначення. Наступна зупинка, трохи далі – це машинний інтелект надлюдського рівня. На станції Humanville потяг може не зупинитися або навіть сповільнюватися. Інший математик І. Дж. Гуд, який працював головним статистиком у групі Алана Тюрінга, фактично був одним з перших, хто проголосив основні аспекти цього сценарію. У 1965 р. учений писав: «Нехай ультрарозумна машина буде визначена як машина, котра може значно перевершити всі інтелектуальні дії будь-якої розумної людини. Оскільки проектування машин є одним з цих інтелектуальних видів діяльності, ультрарозумна машина могла б проектувати ще кращі машини. Тоді, безсумнівно, відбувся б «вибух інтелекту», і інтелект людини залишився б далеко позаду. Таким чином, перша надрозумна машина є останнім винаходом, який людина повинна коли-небудь зробити, за умови, що машина достатньо слухняна, щоб підказувати нам, як тримати її під контролем». Однак, як зазначає визнаний експерт у сфері ШІ Нік Бостром (Nick Bostrom), інші піонери машинного інтелекту, не зважаючи на свою віру в неминучість штучного інтелекту людського рівня, усе-таки не змогли зрозуміти головний наслідок – що «машини згодом стануть надрозумними» [Bostrom N., 2014, p.5].

Варто звернутися до актуальної тематики e-learning, яка сьогодні активно дискутується експертами та медіа. Як вважає Джеффри Хінтон, один із трьох хрещених батьків нинішньої парадигми штучного інтелекту у царині глибокого навчання, не слід очікувати позитив-

них результатів, коли «розумні» штучні агенти будуть спроможні контролювати речі більшого інтелекту. Щоб успішно здійснити перехід до світу з розумнішим за людину штучним інтелектом, нам потрібна «здатність керувати або вирівнювати системи штучного інтелекту так, щоб вони бажали того, що ми хочемо, з метою підтримки позитивного майбутнього. Для цього нам знадобиться, як мінімум, глибоке розуміння того, як ці системи працюють». На жаль, продовжує свою думку Хінтон, зараз ми не маємо реального розуміння того, як вони працюють або чому вони роблять те, що вони роблять. Це результат того факту, що поточне машинне навчання виглядає певним чином не як створення штучного інтелекту, а більше як його «виращування». Відповідаючи на уточнююче запитання, що мається на увазі коли ми кажемо: «ми точно не знаємо, як працює ШІ», учений зауважив: «Ми розробили алгоритм навчання. Це трохи схоже на розробку принципу еволюції. Проте, коли цей алгоритм навчання потім взаємодіє з даними, він створює складні нейронні мережі, які добре справляються з діяльністю, але ми насправді не розуміємо, як саме вони це роблять» [Written statement, 2023].

283

Дійсно, чимало експертів з аналітико-прогностичних центрів, які досліджують довгострокові наслідки впровадження штучного інтелекту у контексті розвитку Індустрії 4.0, вважають, що розширені системи ШІ на основі глибокого навчання схожі сьогодні на чорні ящики. Ми можемо бачити, що ми вкладаємо і що виходить, але коли ми дивимося всередину коробки, усе, що ми бачимо, це трильйони дробових чисел, котрі якимось чином досягають результатів, котрі ми спостерігаємо від системи. Сучасні системи штучного інтелекту поки не становлять загрози, але існує значна ймовірність, що у найближчому майбутньому вони будуть здатними виконувати дедалі складніші багатоетапні завдання без участі людини.

Тут слід вказати на ту обставину що практика і перспективи впровадження штучного інтелекту не були такими безхмарними у минулому столітті. Під час першої «зими штучного інтелекту» кінця 1970-х і 1980-х років нездатність дослідників виконати амбітні обіцянки призвела до краху фінансування та інтересу до ШІ. Щоправда, тоді чимало дослідників здебільшого погоджувалися, що штучний інтелект з часом, ймовірно, почне перевершувати людей у більшості когнітивних завдань.

Проте, у роки Холодної війни практичні напрацювання штучного інтелекту використовувалися насамперед у військовій сфері. Цьому слугувала, до прикладу, удосконалена модель «Вихор-II» (Whirlwind II), яка стала серцем системи SAGE (Semi-Automatic Ground Environment), призначеної для виявлення та відстеження літаків ймовірного супротивника. «Вихор» тоді, аналізуючи величезні потоки даних, що надходять від радарів, миттєво обчислював траєкторії польотів. Згодом ця інформація передавалася операторам, які ухвалювали рішення про перехоплення цілей. При цьому одним із ключових завдань було навчання операторів системи SAGE. Whirlwind моделював різні сценарії порушення повітряного простору США, дозволяючи операторам тренуватися в умовах максимально наближених до реальних бойових. У даному випадку штучний інтелект допомагав виявляти слабкі місця у системі повітряної оборони та відпрацьовувати дії у надзвичайних ситуаціях [Вихрь времени].

284

За умов ведення сучасної війни такі питання набувають особливої ваги. Сьогодні ведуться активні пошуки асиметричних, ефективних рішень у воєнно-технічній сфері та їх безпосереднього втілення у польових/бойових умовах. Одним з таких факторів є застосування БПЛА. Слід розуміти: мета поєднання БПЛА і штучного інтелекту полягає у тому, щоби зробити ефективне використання великих наборів даних (наприклад, аерофотознімків) якомога більш автоматизованим і безпроблемним. БПЛА можуть розкрити весь свій потенціал тільки тоді, коли збір і аналіз даних досягають найвищого рівня автоматизації. Проте слід враховувати, що поєднання БПЛА і ШІ має сенс тільки у тому випадку, якщо це економить гроші і час, що особливо важливо при військовому застосуванні [Щодо проекту стратегії, 2022].

Варто підкреслити, що не слід абсолютизувати можливості впровадження технологій ШІ у військову справу. Особливо багато проблем і справедливих питань виникає до різних рівнів забезпечення автономних системам озброєнь, зокрема до їх автоматизації у бойових умовах. Як відзначає український дослідник та експерт В.І.Слюсар, хоча необхідно визнати, що можливості ШІ перевершують людину при вирішенні багатьох конкретних завдань класифікації, сегментації та розпізнавання візуальних чи акустичних образів, однак у цій сфері все ще існують значні обмеження, коли справа доходить до реального застосування ШІ за межами контрольованого серед-

овища чи лабораторії. Відповідна проблема відома як моделювання реального розриву. Протівниками автономних систем справедливо робиться акцент на поки що нездатності ШІ до узагальнення та перерозподілу продуктивності між різними завданнями в залежності від їх контексту, оскільки ШІ здатний добре виконувати конкретне завдання у дуже специфічних умовах. Це означає, що один і той же алгоритм не буде однаково ефективно працювати під час вирішення інших завдань. Причому у рамках AI-технологій не вирішено також проблему розуміння сенсів [Слюсар В.І., 2021].

Створення інтелектуальних мобільних систем є однією з головних концепцій організації різномірних національних і міжнародних сил, відомих під назвами «Мозаїчні війни» та «Багатодоменні операції», які зараз активно реалізує Агентство оборонних дослідницьких проєктів DARPA (США). Концепція «мозаїчні війни», подібно до керамічної плитки в мозаїці, об'єднує бойові платформи разом, щоб зробити пакет сил. Мета полягатиме в тому, щоб скерувати на ворога стільки різномірних платформ зброї та датників, що його сили будуть переважані. Головну ідею даної концепції можна висловити таким чином: узяти складність і перетворити її на асиметричну перевагу. Завдання організації та планування розподілених мобільних систем вписується в нещодавні спільні концепції «Багатодоменних операцій». Протистояння проти рівних конкурентів означає необхідність захищати сили від загроз, які можуть надходити з будь-якого домену – наземного, повітряного, космічного, морського та/або кіберпростору. Виникає необхідність організовувати багатокomпонентні робототехнічні комплекси в інтелектуальні автономні системи (рої), що діють без центрального та зовнішнього контролю [Казимир В. В., 2022].

285

Отже, штучний інтелект є однією з семи технологічних галузей, які союзники в НАТО виокремили як пріоритетні у зв'язку з їх значимістю для оборони та безпеки. До них належать квантові технології, дані та обчислення, автономні системи, біотехнології та покращення людського потенціалу, гіперзвукові технології та космос. З усіх цих технологій подвійного призначення штучний інтелект, як відомо, є найпоширенішим, особливо у поєднанні з такими як великі дані, автономні системи чи біотехнології [Ланде Д. В., 2022].

Важливість регулювання розвитку ШІ (особливо за умов зростаючого воєнно-політичного протистояння держав й стимульованої

ним глобальної турбулентності та крихкості усїєї системи МВ) під контролем людини й у межах розроблених рамок і правил, зокрема у сфері етики, стає доконаним імперативом часу.

Складні проблеми регулювання ШІ піднімають у своєму експертному дослідженні учені та практики найавторитетнішого у США think tank – корпорації РЕНД. У спеціальній доповіді (2024) зазначається таке: «Спочатку роздуми про етичні наслідки штучного інтелекту призвели до розробки керівних принципів високого рівня. [...] Згодом ці етичні принципи почали перетворюватися на спроби регулювати штучний інтелект, наприклад, запропоновану ЄК законодавчу базу щодо штучного інтелекту». Більш того, сьогодні точаться дискусії щодо стандартизації штучного інтелекту. Виникає також необхідність «налагодження» певних технічних інструментів, які мають неабияке значення для програмістів і розробників у контексті подальшого розгортання працюючих моделей ШІ [Gunashekar Salil and others, 2024].

286

Якщо ж вести мову про актуальні європейські підходи, то у двох відповідних Зверненнях від 25 квітня та 7 грудня 2018 р. Єврокомісія виклала своє бачення штучного інтелекту, яке підтримує «етичний, безпечний і передовий ШІ, створений у Європі». В основі знаходиться підготовка до соціально-економічних змін і забезпечення належної етичної та правової бази для зміцнення європейських цінностей. Підтримуючи реалізацію цього бачення, Комісія створила незалежну Групу експертів високого рівня зі штучного інтелекту (AI HLEG), яку уповноважили розробити, по-перше, керівні принципи щодо етики ШІ та, по-друге, рекомендації щодо політики та інвестицій. Для цього системи штучного інтелекту мають бути орієнтовані на людину, маючи за мету покращення добробуту та свободи людини й запобігаючи можливим ризикам. Щоби суспільство було спроможне «розробляти, розгортати та використовувати» певні техносистеми та інструменти, як підкреслено у документі ЄК, необхідною умовою має стати «надійність» (Trustworthiness) штучного інтелекту. Отже, надійний штучний інтелект повинен мати 3 базові компоненти упродовж усього життєвого циклу системи: (1) бути законним і відповідати всім застосовним законам і нормам; (2) бути етичним, забезпечуючи дотримання етичних принципів і цінностей; (3) стати надійним як з технічної, так і з соціальної точок зору, оскільки, на-

віть з благими намірами, системи ШІ можуть завдати ненавмисної шкоди [Ethics guidelines, 2019].

Експерти RAND, зі свого боку, пропонують навіть для «процвітаючого суспільства» ШІ (AI4People) створити міжнародний консорціум дослідників. Останній має визначити ті принципи, які є ключовими для забезпечення позитивного впливу штучного інтелекту на суспільство. Принцип «благодійності» зокрема охоплює сприяння добробуту, збереження гідності та екологічну стійкість. «Нешкідливість» передбачає вимоги конфіденційності та безпеки. «Автономія» – це баланс між делегуванням прийняття рішень ШІ та збереженням повноважень приймати рішення за людиною. «Справедливість» означає необхідність збереження процвітання та солідарності, а «зрозумілість» підтримує прозорість і розуміння людьми самого процесу прийняття рішень ШІ [Gunashekar Salil and others, 2024]. Тобто узгоджені принципи дотримання прозорості також можуть сприяти етичному використанню штучного інтелекту, даючи змогу дослідникам, регуляторам і громадськості критично оцінювати системи й забезпечувати їх відповідність етичним нормам та стандартам.

287

Звернемося тут до важливих рішень ще одного експертного зібрання, яке відбувалося на глобальному рівні на початку листопада 2023 р. поблизу Лондона, у відомому Блетчлі-парку. Там відбувся Перший міжнародний саміт з питань безпеки штучного інтелекту. Зокрема були представлені 28 країн, серед них США, КНР, Італія, Канада, Японія, Ізраїль, Туреччина, Швейцарія, Саудівська Аравія, Нігерія, Південна Корея, Сінгапур, Україна. До речі, сусідні з Україною – Польща і Румунія не брали участь у конференції. У підсумковій Декларації країни-учасники підкреслили свою готовність сприяти не тільки економічному зростанню та розвитку ШІ як інструменту реалізації політики Індустрії 4.0, але й «зміцненню суспільної довіри та довіри до систем штучного інтелекту для повної реалізації їх потенціалу». Разом з тим, у документі відзначено, що пріоритетом залишається «захист прав людини, прозорість і пояснюваність, справедливість, підзвітність», а також регулювання та безпека. При цьому варто приділити увагу ризикам у таких сферах як кібербезпека та біотехнології, а також у питаннях дезінформації. Сьогодні існує потенціал для «серйозної, навіть катастрофічної, шкоди, навмисної чи ненавмисної», що впливає з найважливіших можливостей цих моделей штучного інтелекту.

Особливі ж ризики для безпеки виникають на «кордоні» штучного інтелекту, який розуміється, з одного боку як високопродуктивні моделі ШІ з широким спектром завдань, а, з іншого, як відповідний специфічний вузький штучний інтелект, котрий може демонструвати можливості, що викликають шкоди. Тому у глобальному контексті важливою задачею залишається досягнення «пропорційного управління та регуляторного підходу», який враховує ризики, пов'язані з ШІ. Це може включати створення, де доцільно, класифікації та категоризації ризику на основі національних обставин і застосовної правової бази. «Ми також відзначаємо актуальність співпраці, де це доречно, щодо таких підходів як спільні принципи та кодекси поведінки» [The-bletchley-declaration, 2023].

288

У цьому контексті варто нагадати той факт, що раніше, на Генеральній конференції ЮНЕСКО 21 листопада 2021 р., 193 країни, включаючи Україну, ухвалили глобальні етичні стандарти для штучного інтелекту, у яких виокремлено чотири основні напрями регулювання поведінки ШІ: захист даних; соціальна оцінка та масове спостереження; контроль; захист навколишнього середовища.

Взагалі, якщо поглиблювати тему машинної етики (або роботоетики) відзначимо, що в якості підгалузі вона виникла нещодавно і зосереджується на завданні забезпечення етичної поведінки штучних агентів, зазвичай, званих АМА (Artificial Moral Agents). Інакше кажучи, ідеться про особливий напрямок в етиці у застосуванні до технологій. Ось яким чином описують цю проблему експерти з техноетики, розробляючи низхідну та висхідну стратегії формування штучної моралі агентів. Отже, слід надати їм «достатньо розуму», щоби оцінити «наслідки своїх дій на живих істотах» і тому вміти робити правильний вибір. Тобто відповідні підходи зверху вниз передбачають перетворення явних теорій моралі в алгоритми. Підходи ж знизу вгору включають спроби навчити або розвивати агентів, чия поведінка спроможна наслідувати «моральну поведінку що схвалюється людиною» [Wallach Wendell and others, 2008].

Важливою змінною тут є якість «соціальної епістемології» галузі штучного інтелекту і якщо, можна так сказати, підвищення його «внутрішньої відповідальності». Про це пише зокрема згадуваний Нік Бостром: уявіть собі проект, який інвестує мільйони доларів і роки праці для розробки прототипу ШІ, і що після подолання багатьох технічних проблем система нарешті починає демонструвати

реальний прогрес. Однак при цьому виявлено, що інша, більш досконала, модель його «соціальної поведінки» може стати набагато безпечнішою. Як наслідок, чи зможе вже зроблений проєкт ШІ вбити себе як «зганьблений самурай», відмовляючись від свого небезпечного дизайну та прогресу, якого було досягнуто, або ж він реагує як «стурбований восьминіг, що випускає хмару вмотивованого скептицизму в надії уникнути нападу»? На думку експерта, проєкт, який надійно обрав би самурайський варіант у такій дилемі, був би кращим девелоперським рішенням [Bostrom N., 2014, p.5].

Проте, людство очікують ще більш глобальні ризики, якщо вчасно не звернути увагу на напружене цивілізаційне протистояння. У цьому контексті важливо акцентувати на питаннях геостратегічного «змагання» у використанні штучного інтелекту, яке відбувається у перші десятиліття XXI ст., особливо між США і Китаєм. Тут звернемо увагу на програмну статтю «Парадокс сили штучного інтелекту», опубліковану у 2023 р. в «Foreign Affairs», який є органом Ради з зовнішньої політики (США). Її автори, Ян Бреммер і Мустафа Сулейман, наголошують на тому, що сучасні дискусії про управління штучним інтелектом залишаються в пастці небезпечної хибної дилеми – що робити: використовувати його для розширення національної влади чи придушувати його, щоби уникнути ризиків. Навіть ті, хто точно діагностує проблему, намагаються її вирішити, вбудовуючи ШІ в існуючі або історичні рамки управління. Проте, на думку експертів, штучним інтелектом «не можна керувати, як будь-якою попередньою технологією, і він уже змінює традиційні уявлення про геополітичну силу». Штучний інтелект вимагає сьогодні оригінальних управлінських рішень на глобальному рівні. Причому структура управління повинна бути запобіжною, гнучкою, інклюзивною, непроникною та цілеспрямованою.

Спираючись на ці принципи, політики повинні створити принаймні 3 режими управління, що перетинаються:

- один для встановлення фактів і консультування урядів щодо ризиків, пов'язаних зі штучним інтелектом;
- другий для запобігання повній гонці озброєнь між державами;
- третій для управління руйнівними силами технології, яка не схожа до всього, що уже бачив світ у минулому [Bremmer I., Suleyman M., 2023].

Тут також варто враховувати швидкість прогресу у сфері ШІ. Згадаємо, що відомий закон Мура передбачав подвоєння обчислювальної потужності кожні 2 роки. Завдяки новій хвилі штучного інтелекту цей темп прогресу виглядає дивним. Коли OpenAI запустив свою першу велику мовну модель, відому як GPT-1, у 2018 р., вона мала 117 мільйонів параметрів – показник масштабу та складності системи. Через 5 років модель четвертого покоління компанії GPT-4, як вважають, має понад трильйон параметрів [International roadmap, 2016].

Різниця між Китаєм і Заходом усе-таки буде відчутна в організаційному аспекті. У КНР майже виключно усі сфери пов'язані з ШІ буде контролювати держава, а на Заході, особливо у США, в найближчій перспективі траєкторія штучного інтелекту значною мірою визначатиметься рішеннями кількох приватних компаній, незалежно від того, що робитимуть політики в Брюсселі чи Вашингтоні. Інакше кажучи, технологи, а не політики чи бюрократи, здійснюватимуть владу над силою, яка може глибоко змінити як міць національних держав, так і те як вони ставляться одна до одної. Це робить проблему управління штучним інтелектом несхожою на будь-яку іншу з якою уряди стикалися раніше [Bremmer I., Suleyman M., 2023]. Регулятивний алгоритм балансування делікатніший і, зрозуміло, вимагатиме від політиків значно більшої гнучкості та передбачуваності.

Одне суттєве зауваження щодо ролі Китаю, який бажає сьогодні вибудувати ієрархічну систему залежних відносин з багатьма державами, особливо, по периметру власних кордонів. Однак дослідники РЕНД вказують на факти що засвідчують про відсутність підтримки синоцентризму у АТР. У спеціальній доповіді експерти відзначили, що успіх і привабливість системи данини значною мірою ґрунтувалися на китайсько-центричному уявленні про те, що Китай служив центром людської культури та цивілізації. У минулі століття слабші сусіди визнавали культурну перевагу Китаю і прагнули стати складовою його цивілізації. Деякі держави, насамперед у корейських та в'єтнамських прикордонних районах, продемонстрували своє прийняття цієї ідеології, наслідуючи китайську культуру.

Однак сьогодні Китай зіткнувся зі складним міжнародним суспільством, де розвиваються могутні суперницькі цивілізації, включаючи Японію, Сполучені Штати, Індію та Європу. Навіть попередні клієнти імперського Китаю в Кореї та В'єтнамі розвинули унікальну культуру, мови та ідентичність. Практично жодна країна не розглядає

Китай як центр цивілізації. Таким чином, «ідеологічного підґрунтя традиційної системи данини просто немає». Країни, які приймають китайський патронат, наприклад, Камбоджа, здається, мотивовані більше «бажанням щедрості та економічної вигоди, ніж будь-яким ототожненням із культурою та світоглядом Китаю» [Mazarr Michael J. and others, 2024].

Таким чином, виміри геополітичної гри XXI ст. не можливо більше розглядати поза контекстом розвитку Індустрії 4.0 та її важливого інструменту штучного інтелекту, який набуває неабиякої ваги у сучасному світі.

Висновки. Отже, говорячи про різні аспекти становлення Індустрії 4.0 (а тепер вже і 5*ir*), навряд чи якомусь серйозному оглядачу або ж експерту вдасться обійти тему штучного інтелекту – цього потужного та інноваційного каталізатора змін у світі. У документах щодо розвитку Індустрії 4.0 та 5.0, котрі останніми роками оприлюднені в Україні, саме з новітніми технологіями та їх гуманітарною складовою й супроводом пов'язані надії та плани відновлення економіки країни після завершення сучасної війни. Ідеться насамперед про активне залучення ІТ-сектору і ШІ до питань передання досвіду глобалізації та кращих бізнес-практик підприємствам і організаціям; інтеграцію до європейських та світових ланцюжків цінностей. Не менш активно у глобальному та національному контекстах обговорюється проблематика можливостей «етичної поведінки» штучного інтелекту.

291

Провідні фабрики думок у світі задаються засадничим питанням чи можливе формування «розумного машинного агента», який, не виходячи з під контроль людини, відповідав би тим чи іншим моральним імперативам й «діяв», керуючись конкретними гуманними цілями на благо інтересів суспільства й пересічного громадянина.

Саме таку актуальну тематику розробляють знані think tanks, зокрема такі як Інститут Алана Тюрінга, Центр 4-ої Індустріальної революції СЕФ у Давосі за підтримки фірми «Хітачі», RAND Corporation. Як відзначають американські експерти Я. Бреммер і М. Сулейман у нещодавній статті в «Foreign Affairs», гонка інтелектуальних та виробничих моделей штучного інтелекту між США і КНР набирає усе більших обертів в їх геостратегічному протиборстві. Тому штучний інтелект вимагає оригінальних управлінських рішень на глобальному рівні. Структура управління повинна бути запобіжною, гнучкою,

інклюзивною, непроникною та цілеспрямованою. Сьогодні усім зацікавленим сторонам – державам, інституціям у сфері ІТ, неурядовим акторам і фірмам-розробникам варто, виходячи з цих принципів, сформувати 3 режими управління, що перетинаються. Перший – для встановлення фактів і консультування урядів щодо ризиків, пов’язаних зі штучним інтелектом; другий для запобігання повній гонці озброєнь між державами на інноваційній основі; третій для управління руйнівними силами технології, яка не схожа до всього, що уже бачив світ у минулому.

Дехто з аналітиків вважає, що у майбутньому ШІ спроможний буде вибудувати власну картину світу й навіть створити свою віртуальну копію всесвіту. А радикальні трансгуманісти сходяться в поглядах із футуристом Р. Курцвайлом, який стверджує, що у кінцевому рахунку ми зіллємося з нашими технологіями: до середини 2040-х років небіологічна частина нашого інтелекту буде діяти в мільярди разів потужніше, ніж його біологічна частка.

292

Проблематика штучного інтелекту «увірвалася» й у царину політичного. Зокрема подальше упровадження ШІ значно порушує наші уявлення про те, що у політиці зазвичай продовжує домінувати модель досягнення задалегідь визначених рішень. І як припускає сучасний марокканський дослідник проблематики штучного інтелекту у гуманітарних науках Алфредо Валладео (Alfredo Valladão), можливо, на противагу жорсткій «причинно-наслідковій» схемі, саме концепція людської історії як рекурентної нейронної мережі, керованої штучним інтелектом, стане внеском у пошуках нових обріїв для сучасної політичної науки [Valladão, Alfredo G. A., 2018].

Висловимо тут авторську суб’єктивну думку. Мабуть, штучний інтелект, за умов його подальшого фреймування та етичного доведення, є лише одним з новітніх напрямків (згадаємо тут тезу Геделя) розвитку розуму людства у XXI столітті. Окремі моделі надлюдського інтелекту можуть використовуватися ШІ, але усе-таки замінити людський інтелект/ витончені виміри поведінки людини повністю вони вряд чи будуть спроможні. Теорема Геделя про неповноту змусила учених замислитися про обмеженість знань людини щодо власних ментальних можливостей. Адже людське мислення не ґрунтується виключно на формально-обчислювальних засадах, але й включає ще далеко недостатньо знану, так би мовити «невираховну» сферу, проявом якої є інтуїція та раптові осяяння. Чи будуть вони

взагалі досяжні для штучного інтелекту? Наша відповідь поки що: вряд чи.

Норберт Вінер колись зауважив, що битва за знання є важливішою за перемогу. Дотримуючись об'єктивності, слід визнати: тема штучного інтелекту у контексті розвитку Індустрії 4.0 і 5.0, породжує у багатьох експертів більше запитань, ніж дає адекватні й верифіковані відповіді на них, особливо в етико-гуманітарній сфері. Проте, у запропонованій розвідці автори й намагалися задати собі і Читачеві декілька непростих запитань, відповідаючи на які інші дослідники зможуть надати з часом більш адекватні й притомні відповіді. Чекаємо на них.

Джерела та література:

1. Бердо, Р. С., Расюн, В. Л., Величко, В. А. 2023. Штучний інтелект та його вплив на етичні аспекти наукових досліджень в українських закладах освіти. Академічні візії, (22). <https://academy-vision.org/index.php/av/article/view/469> (accessed May 21, 2024). 293

2. Вихрє времени. История уникального компьютера Whirlwind. <https://habr.com/ru/companies/serverspace/articles/819429/> (accessed May 12, 2024). —

3. Завербний А. С., Сало К. Р. 2022. Проблеми та перспективи розвитку індустрії 4.0 в Україні за умов євроінтегрування. <https://science.lpnu.ua/sites/default/files/journal-paper/2022/dec/29557/220972maket-378-386.pdf> (accessed May 12, 2024).

4. Казимир В. В. 2022. До підрозділу 7.1. “Штучний інтелект у сфері безпеки та в оборонно-промисловому комплексі України”. Щодо проекту стратегії розвитку штучного інтелекту в Україні на 2022-2030 рр. Artificial Intelligence. № 1. https://www.slyusar.kiev.ua/AI_2022-1-1_ua.pdf (accessed April 12, 2024).

5. Ланде Д. В. 2022. До підрозділу 7.1. “Штучний інтелект у сфері безпеки та в оборонно-промисловому комплексі України”. Щодо проекту стратегії розвитку штучного інтелекту в Україні на 2022-2030 рр. Artificial Intelligence. № 1. https://www.slyusar.kiev.ua/AI_2022-1-1_ua.pdf (accessed April 12, 2024).

6. Маніфест щодо переходу України до індустрії 5.0. 2023. 3 липня. <https://www.industry4ukraine.net/category/publications/> (accessed May 12, 2024).

7. Національна стратегія «Індустрії 4.0 в Україні»: проєкт. 2018. <https://industry4-0-ukraine.com.ua/2019/01/02/ukrainska-strategiya-industrii-4-0-7-napriankiv-rozvitku/> (accessed May 12, 2024).

8. Слюсар В.І. 2021. Концепція виртуалізації поля боя 2050 года. Озброєння та військова техніка. 3(31). С.111-112.

9. Стратегія розвитку штучного інтелекту в Україні. 2023. За редакцією А. І. Шевченка. Київ. НАН України. 305 с. https://jai.in.ua/archive/2023/ai_mono.pdf (accessed May 15, 2024).

10. Чайкіна А.О. 2021. Індустрія 4.0: особливості цифрової трансформації України. Вчені записки ТНУ імені В. І. Вернадського. Серія: Економіка і управління. https://www.econ.vernadskyyournals.in.ua/journals/2021/32_71_3/7.pdf (accessed May 21, 2024).

11. Щодо проєкту стратегії розвитку штучного інтелекту в Україні на 2022-2030 рр. 2022. Artificial Intelligence. № 1. https://www.slyusar.kiev.ua/AI_2022-1-1_ua.pdf (accessed May 22, 2024).

294
— 12. About MIRI. <https://intelligence.org/about/> (accessed May 12, 2024).

13. Artificial Intelligence. 2010. A Modern Approach. Third Edition. Ed. by Stuart J. Russell and Peter Norvig. Pearson Education, Inc., New Jersey. 1132 p. <https://dl.ebooksworld.ir/books/Artificial.Intelligence.A.Modern.Approach.4th.Edition.Peter.Norvig.%20Stuart.Rus.sell.Pearson.9780134610993.EBooksWorld.ir.pdf> (accessed May 27, 2024).

14. Butler, Samuel [Cellarius, pseud.]. 1863. Darwin among the machines. Christchurch Press, June 13. <http://www.nzetc.org/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html>. (accessed April 12, 2024).

15. Bostrom N. 2014. Superintelligence. Paths, Dangers, Strategies. Oxford, Oxford University Press. XVI, 328 p.

16. Bremmer I., Suleyman M. 2023. The AI Power Paradox Can States Learn to Govern Artificial Intelligence-Before It's Too Late? https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox?check_logged_in=1&utm_medium=promo_email&utm_source=lo_flows&utm_campaign=registered_user_welcome&utm_term=email_1&utm_content=20230829 (accessed May 27, 2024).

17. Ethics guidelines for trustworthy AI. 2019. High-Level Expert Group on Artificial Intelligence. Brussels. European Commission. 8 April. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (accessed May 17, 2024).

18. Gunashekar, Salil, Henri van Soest, Michelle Qu, Chryssa Politi, Maria Chiara Aquilino, and Gregory Smith. 2024. Examining the landscape of tools for trustworthy AI in the UK and the US: Current trends, future possibilities, and potential avenues for collaboration. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RRA3194-1.html. (accessed June 2, 2024).

19. International roadmap for devices and systems. 2016. More Moore white paper. https://ir.ds.ieee.org/images/files/pdf/2016_MM.pdf (accessed May 27, 2024).

20. Mazarr, Michael J., Alexis Dale-Huang, John Deak, Gregory Weider Fauerbach, Stacie Goddard, Timothy R. Heath, and Joshua Shifrinson. 2024. The Fates of Nations: Varieties of Success and Failure for Great Powers in Long-Term Rivalries. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RRA2611-2.html. (accessed May 27, 2024).

21. 20 a. Milz, Manfred. 2023. "Maine de Biran in Huxley's Brave New World: Transcending the Utilitarian through a Spiritual Self". In *Towards a New Anthropology of the Embodied Mind: Maine de Biran's Physio-Spiritualism from 1800 to the 21st Century*. Amsterdam, Brill. <https://brill.com/display/book/9789004683778/BP000015.xml> (accessed June 7, 2024).

295

22. Shulman, Carl, Henrik Jonsson, and Nick Tarleton. 2009. "Machine Ethics and Superintelligence." In *AP-CAP 2009: The Fifth Asia-Pacific Computing and Philosophy Conference*, October 1st-2nd, University of Tokyo, Japan. <https://intelligence.org/files/MachineEthicsSuperintelligence.pdf>

23. Soichi Furuya. 2019. Consideration of Governance of Innovative Technology at World Economic Forum's Centre for the Fourth Industrial Revolution. Global innovation report. Hitachi Review. Vol. 68. No. 4. P. 454-455. https://www.hitachi.com/rev/archive/2019/r2019_04/pdf/P044-049_R4-GIR.pdf (accessed Oct. 27, 2024).

24. The-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november. 2023. <https://www.gov.uk/government/topical-events/ai-safety-summit-2023> (accessed May 12, 2024).

25. Valladão, Alfredo G. A. 2018. Artificial Intelligence and Political Science. Rabat: OCP Policy Center. https://www.policycenter.ma/sites/default/files/OCPPC-PP1807_0.pdf (accessed May 27, 2024).

26. Written statement of MIRI CEO Malo Bourgon to the AI Insight Forum. 2023. December 6. <https://intelligence.org/2023/12/06/written-statement-of-miri-ceo-malo-bourgon-to-the-ai-insight-forum/> (accessed May 12, 2024).

27. Wallach, Wendell, Colin Allen, and Iva Smit. 2008. "Machine Morality: Bottom-Up and Top-Down Approaches for Modelling Human Moral Faculties." In *Ethics and Artificial Agents*. Special issue, *AI & Society* 22 (4): 565–582. https://www.academia.edu/30154468/Machine_Morality_Bottom-up_and_Top-down_Approaches (accessed May 27, 2024).

References:

1. Berdo, R. S., Rasyun, V. L., Velychko, V. A. 2023. Shtuchnyy intelekt ta yoho vplyv na etychni aspekty naukovykh doslidzhen' v ukrayin's'kykh zakladakh osvity. *Akademichni vizyyi*, (22). <https://academy-vision.org/index.php/av/article/view/469> (accessed May 21, 2024).

296

— 2. Vykhr' vremeny. Ystoryya unykal'noho komp'yutera Whirlwind. <https://habr.com/ru/companies/serverspace/articles/819429/> (accessed May 12, 2024).

3. Zaverbnyy A. S., Salo K. R. 2022. Problemy ta perspektyvy rozvytku industriyi 4.0 v Ukraini za umov yevrointehruvannya. <https://science.lpnu.ua/sites/default/files/journal-paper/2022/dec/29557/220972maket-378-386.pdf> (accessed May 12, 2024).

4. Kazymyr V. V. 2022. Do pidrozdilu 7.1. «Shtuchnyy intelekt u sferi bezpeky ta v oboronno-promyslovomu kompleksi Ukrainy». Shchodo proyektu stratehiyi rozvytku shtuchnoho intelektu v Ukraini na 2022-2030 rr. *Artificial Intelligence*. № 1. https://www.slyusar.kiev.ua/AI_2022-1-1_ua.pdf (accessed April 12, 2024).

5. Lande D. V. 2022. Do pidrozdilu 7.1. «Shtuchnyy intelekt u sferi bezpeky ta v oboronno-promyslovomu kompleksi Ukrainy». Shchodo proyektu stratehiyi rozvytku shtuchnoho intelektu v Ukraini na 2022-2030 rr. *Artificial Intelligence*. № 1. https://www.slyusar.kiev.ua/AI_2022-1-1_ua.pdf (accessed April 12, 2024).

6. Manifest shchodo perekhodu Ukrainy do industriyi 5.0. 2023. 3 lypnya. <https://www.industry4ukraine.net/category/publications/> (accessed May 12, 2024).

7. Natsional'na stratehiya «Industriyi 4.0 v Ukraini»: projekt. 2018. <https://industry4-0-ukraine.com.ua/2019/01/02/ukrainska-strategiya-industrii-4-0-7-napriankiv-rozvitku/> (accessed May 12, 2024).

8. Slyusar V.I. 2021. Kontseptsyya vyrtualyzatsyy polya boya 2050 hoda. *Ozbroyennya ta viys'kova tekhnika*. 3(31). S.111-112.

9. Stratehiya rozvytku shtuchnoho intelektu v Ukraini. 2023. Za zah. redaktsiyeyu A. I. Shevchenka. Kyiv. NAN Ukrainy. 305 c. https://jai.in.ua/archive/2023/ai_mono.pdf (accessed May 15, 2024).

10. Chaykina A.O. 2021. Industriya 4.0: osoblyvosti tsyfrovoyi transformatsiyi Ukrainy. *Vcheni zapysky TNU imeni V. I. Vernads'koho*. Seriya: Ekonomika i upravlinnya. https://www.econ.vernadskyjournals.in.ua/journals/2021/32_71_3/7.pdf (accessed May 21, 2024).

11. Shchodo projektu stratehiyi rozvytku shtuchnoho intelektu v Ukraini na 2022-2030 rr. 2022. *Artificial Intelligence*. № 1. https://www.slyusar.kiev.ua/AI_2022-1-1_ua.pdf (accessed May 22, 2024).

12. About MIRI. <https://intelligence.org/about/> (accessed May 12, 2024).

297

13. *Artificial Intelligence. 2010. A Modern Approach*. Third Edition. Ed. by Stuart J. Russell and Peter Norvig. Pearson Education, Inc., New Jersey. 1132 p. <https://dl.ebooksworld.ir/books/Artificial.Intelligence.A.Modern.Approach.4th.Edition.Peter.Norvig.%20Stuart.Rus.sell.Pearson.9780134610993.EBooksWorld.ir.pdf> (accessed May 27, 2024).

14. Butler, Samuel [Cellarius, pseud.]. 1863. Darwin among the machines. Christchurch Press, June 13. <http://www.nzetc.org/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html>. (accessed April 12, 2024).

15. Bostrom N. 2014. *Superintelligence. Paths, Dangers, Strategies*. Oxford, Oxford University Press. XVI, 328 p.

16. Bremmer I., Suleyman M. 2023. The AI Power Paradox Can States Learn to Govern Artificial Intelligence-Before It's Too Late? https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox?check_logged_in=1&utm_medium=promo_email&utm_source=lo_f_lows&utm_campaign=registered_user_welcome&utm_term=email_1&utm_content=20230829 (accessed May 27, 2024).

17. Ethics guidelines for trustworthy AI. 2019. High-Level Expert Group on Artificial Intelligence. Brussels. European Commission. 8 April. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (accessed May 17, 2024).

18. Gunashekar, Salil, Henri van Soest, Michelle Qu, Chryssa Politi, Maria Chiara Aquilino, and Gregory Smith. 2024. Examining the landscape of tools for trustworthy AI in the UK and the US: Current trends, future possibilities, and potential avenues for collaboration. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RRA3194-1.html. (accessed June 2, 2024).

19. International roadmap for devices and systems. 2016. More Moore white paper. https://ir.ds.ieee.org/images/files/pdf/2016_MM.pdf (accessed May 27, 2024).

20. Mazarr, Michael J., Alexis Dale-Huang, John Deak, Gregory Weider Fauerbach, Stacie Goddard, Timothy R. Heath, and Joshua Shifrinson. 2024. The Fates of Nations: Varieties of Success and Failure for Great Powers in Long-Term Rivalries. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RRA2611-2.html. (accessed May 27, 2024).

298 — 21. 20 a. Milz, Manfred. 2023. “Maine de Biran in Huxley’s Brave New World: Transcending the Utilitarian through a Spiritual Self”. In *Towards a New Anthropology of the Embodied Mind: Maine de Biran’s Physio-Spiritualism from 1800 to the 21st Century*. Amsterdam, Brill. <https://brill.com/display/book/9789004683778/BP000015.xml> (accessed June 7, 2024).

22. Shulman, Carl, Henrik Jonsson, and Nick Tarleton. 2009. “Machine Ethics and Superintelligence.” In *AP-CAP 2009: The Fifth Asia-Pacific Computing and Philosophy Conference*, October 1st-2nd, University of Tokyo, Japan. <https://intelligence.org/files/MachineEthicsSuperintelligence.pdf>

23. Soichi Furuya. 2019. Consideration of Governance of Innovative Technology at World Economic Forum’s Centre for the Fourth Industrial Revolution. Global innovation report. Hitachi Review. Vol. 68. No. 4. P. 454-455. https://www.hitachi.com/rev/archive/2019/r2019_04/pdf/P044-049_R4-GIR.pdf (accessed Oct. 27, 2024).

24. The-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november. 2023. <https://www.gov.uk/government/topical-events/ai-safety-summit-2023> (accessed May 12, 2024).

25. Valladão, Alfredo G. A. 2018. Artificial Intelligence and Political Science. Rabat: OCP Policy Center. https://www.policycenter.ma/sites/default/files/OCPPC-PP1807_0.pdf (accessed May 27, 2024).

26. Written statement of MIRI CEO Malo Bourgon to the AI Insight Forum. 2023. December 6. <https://intelligence.org/2023/12/06/written-statement-of-miri-ceo-malo-bourgon-to-the-ai-insight-forum> (accessed May 12, 2024).

27. Wallach, Wendell, Colin Allen, and Iva Smit. 2008. "Machine Morality: Bottom-Up and Top-Down Approaches for Modelling Human Moral Faculties." In *Ethics and Artificial Agents*. Special issue, *AI & Society* 22 (4): 565–582. https://www.academia.edu/30154468/Artificial_Morality_Top_down_Bottom_up_and_Hybrid_Approaches (accessed May 27, 2024).