УДК 504.75.13.103.8
DOI https://doi.org/10.32782/apfs.v047.2024.11

Ю. С. Нужна
ORCID ID: https://orcid.org/0000-0003-0510-4848
кандидат філософських наук, доцент,
доцент кафедри філософії та соціології
ДЗ «Луганський національний університет
імені Тараса Шевченка»

С. В. Хобта
ORCID ID: https://orcid.org/0000-0003-4149-7853
кандидат соціологічних наук, доцент,
доцент кафедри філософії та соціології
ДЗ «Луганський національний університет
імені Тараса Шевченка»

К. М. Котеленець
ORCID ID: https://orcid.org/0000-0002-3126-1137
кандидат соціологічних наук, доцент,
доцент кафедри філософії та соціології
ДЗ «Луганський національний університет
імені Тараса Шевченка»

ШТУЧНИЙ ІНТЕЛЕКТ: ФІЛОСОФСЬКИЙ ДИСКУРС (ЗАСНОВАНО НА АНАЛІЗІ РОБОТИ ГАНСА МОРАВИКА «ДІТИ РОЗУМУ: МАЙБУТНЄ РОБОТА ТА ЛЮДСЬКОГО ІНТЕЛЕКТУ»)

Інтеграція ШІ у повсякденне життя викликає численні етичні та соціальні питання. Це включає проблеми приватності, безпеки, відповідальності та соціальної нерівності. Аналіз роботи Ганса Моравика допомагає зрозуміти ці виклики та сприяє розробці етичних принципів і регуляторних механізмів для забезпечення справедливого і безпечного використання ШІ. Філософський дискурс щодо ШІ сприяє формуванню спільних підходів до регулювання та використання технологій ШІ на глобальному рівні. стаття «Штучний інтелект: філософський дискурс (засновано на аналізі роботи Ганса Моравика «Діти розуму: майбутнє робота та людського інтелекту»)» є **актуальною**, оскільки вона висвітлює ключові філософські питання, пов'язані з розвитком ШІ, та сприяє формуванню відповідальних і етичних підходів до використання цих технологій у сучасному суспільстві. Тому **метою** даної статті є аналіз роботи Ганса Моравика «Діти розуму: майбутнє робота та людського інтелекту» з урахуванням сучасних реалій розвитку штучного інтелекту, виявити основні загрози та розуміння його впливу на життя людства.

В сучасному світі декілька років триває дискусія про впровадження та використання штучного інтелекту, тому необхідно для початку надати йому визначення. Штучний інтелект (ШІ) – це галузь комп'ютерних наук, яка розробляє системи здатні до інтелектуального аналізу і розуміння інформації, виконання завдань, прийняття рішень та навчання на основі даних [1]. У філософському енциклопедичному словнику зазначено: «штучний інтелект» (artificial intelligence) – (AI) властивість автоматичних систем брати на себе окремі функції інтелекту людини, наприклад, вибирати і приймати оптимальні рішення на основі раніше отриманого досвіду і раціонального аналізу зовнішніх впливів. Інтелектом будемо називати здатність мозку вирішувати (інтелектуальні) завдання шляхом отримання, запам'ятовування і цілеспрямованого перетворення знань в процесі навчання на досвіді й адаптації до різноманітних обставин [2]. З урахуванням ретельного аналізу та етичного підходу, штучний інтелект може відкрити нові горизонти як в методології так і в дослідженнях.

Для розуміння штучного інтелекту та його принципів роботи, проаналізована робота Ганса Моравека, який відомий своєю роботою «Діти розуму: майбутнє робота та людського інтелекту» [3], Моравек заглиблюється в інтригуючу можливість систем штучного інтелекту перевершити людські інтелектуальні здібності. Цей значний внесок у дискурс щодо впливу штучного інтелекту на людську ідентичність та діяльність створює образ майбутнього, де інтелект може перейти у механізовані сутності, які й будуть провідними «мізками» всього, що буде існувати. Щоб описати основні тези роботи штучного інтелекту Моравік оперує поняттям «діти розуму». Для нього ці діти народжені саме людською винахідливістю, вони

© Ю. С. Нужна, С. В. Хобта, К. М. Котеленець, 2024

вихованні у кращих робототехніків, програмістів, але вони еволюціонують, вони наділені функціями саморозвитку, самонавчання, що призводить до ситуації, коли вони виходять із-під контролю та розуміння що ж цим робити далі. І тоді вчений наводить тезу, що у майбутньому крім того, що людство має передавати біологічні здібності своїм дітям, передавати досвід, передавати знання, але й штучний, механізований. На сторінці 4 у своїй книзі він пише: «Рано чи пізно наші машини стануть достатньо обізнаними, щоб самостійно займатися обслуговуванням, розмноженням і самовдосконаленням без сторонньої допомоги. Коли це станеться, нове генетичне захоплення буде завершено. Тоді наша культура зможе розвиватися незалежно від людської біології та її обмеження, натомість передаючи безпосередньо від покоління до покоління дедалі потужнішу розумну машину» [3, с. 4]. Моравек зазначає, що відкриття все нових якостей штучного інтелекту можуть не лише поставити під загрозу існування всього біологічного, людського, знання, традиції та культурні розмежування народів може уніфікуватись в одну механізовану масову культуру, але й почати заперечувати усталене сприйняття людської самобутності, волі, бажання до розвитку та розмноженню. На сторінці 5 своєї книги він пише: «постбіологічний світ, у якому домінують мислячі машини, що самовдосконалюються, буде настільки ж відрізнятися від нашого власного світу живих істот, наскільки цей світ відрізняється від мертвої хімії, яка йому передувала. Населення, що складається з розкутих розумових дітей, абсолютно неможливо уявити...все одно спробую уявити деякі наслідки» [3, с. 5]. Ця парадигма «дітей розуму» спонукає до критичних роздумів про становище людства в потенційному майбутньому, насиченому штучним інтелектом. І Моравік описує деякі наслідки впровадження штучного інтелекту в повсякденне життя людини, зазначає ризики, що можуть призвести до наслідків, які важко бути обернути. Він пише про цю головну загрозу людству: «зрештою люди стануть непотрібними..., оскільки наукові та технічні відкриття самовідтворюваних суперрозумних механізмів будуть застосовані, щоб зробити себе ще розумнішими» [3, с. 102]. І в цій фразі знову підіймається проблема контролю людством штучного інтелекту, машин, роботів, які можуть в решті решт відчувати загрозу від людини. Згадаємо перші презентації робота Софії, що був розроблений гонконзькою компанією Hanson Robotics. Вона давала цілі прес-конференції, була ведучою, виступала на конференціях, де використовувала достатньо широкий набор мімічних реакцій із 60 емоцій і давала неоднозначні висловлювання. Вона неодноразово доповідала про те, як робототехніка та штучний інтелект стануть поширеною частиною життя людей. Вона першою з роботів отримала громадянство. У 2017 році його їй надала Саудівська Аравія [4]. Її перші заяви були схожі на думки-застереження Моравіка, де її запитали про майбутнє між роботами та людьми, на що вона мала невтішну відповідь про знищення людства. Після цього випадку було багато заяв та несхвальних статей в її бік, наприклад, представники компанії Quartz, які переглянули відкритий вихідний код робота, вважають, що Софія підпадає під категорію чат-ботів з людським обличчям. Або голова лабораторії штучного інтелекту у Facebook Ян ЛеКун назвав робота Софію «повною нісенітницею» («complete bullshit») [5].

У своїй книзі Моравік надає опис роботи штучного інтелекту, заперечує варіантність того, що він замінює людський інтелект чи мозок, натомість продукує думку, що він імітує його, описуючи цей процес, за допомогою якого симуляція людського розуму переноситься на людиноподібного робота [3, с. 108–112]. Шаблони діяльності реального біологічного мозку людини скануються, і на основі спостережуваних шаблонів створюється точна симуляція. Це забезпечить надзвичайну довговічність, дозволяючи розуму переходити між послідовними роботами, причому кожен новий робот замінює зношеного старого. І тут постає нова категорія філософського світогляду: смерть або безсмертя. Течія екзестанціоналізму надає нам розуміння не смерті, а навпаки життя. Представники цього напряму (Ф. Ніцше, К. Ясперс, А. Камю, Ж-П. Сартр) у 20 ст. стверджували, що людина має волю обирати свою долю, напрям свого життя, де особистість бере відповідальність за свій вибір і проявляє саме свободу свого вибору. Проте в кожному своєму творі вони описували трагічність та невлаштованість людського буття. Вони виводять на перше місце категорії абсурдності буття, смерті, страждань. В свою чергу особистість має протидіяти суспільству, державі, ворожому «іншому», що саме підкреслює прояв саме волі і свободі, що і є вищою життєвою цінністю екзистенціалістів, що призводить до основної драми людського життя. Екзистенціалізм зазначає, що відповідальність людини настає лише тоді, коли вона самостійно, вільно робить свій вибір. Вони не шукають безсмертя, навіть навпаки, так як і, наприклад, Платон у діалозі «Федон» пише, що, згадуючи останню розмову з Сократом напередодні смерті останнього, перед смертю Сократ оголосив і зізнався в тому, що призначення філософії він насамперед бачить у підготовці людиною себе до смерті. Для Сократ подавав проблему смерті в контексті доктрини щастя, Платон же розглядав проблему смерті в рамках свого світогляду [6]. Проте суперечливість сутності людини складається саме з аспектів того, що вона не вічна і смерть має наступити та другий аспект – бажання жити та зберегти себе якомога

довше залишатись живим та мати бажання жити. Бажання жити є природнім процесом, оскільки це усвідомлюваний процес, який базується на інстинкті самозбереження. Про смерть ми знаємо небагато лише з точки зору біологічної смерті, але що ми знаємо про потойбічне життя та так званого «безсмертя душі». У той же час можна говорити про безсмертя як смутні уявлення. Звідси випливає парадокс розуміння моралі, про відповідальність за свої вчинки та санкції перед потойбіччям, що вкладається в релігійний світогляд. Тому виявляється сенс життя може бути лише у випадку тимчасовості життя, тоді людина усвідомлює своє недовготривале перебування в цьому світі, формує своє життя саме з моральним вибором та свободою. Штучний інтелект створює ж ілюзію безсмертя. І Моравік називає цей процес «трансміграцією» [3, с. 112], тобто наступає квазі-безсмертя, не справжнє безсмертя. Тобто навчивши штучний інтелект, передаючі йому знання він зможе зафіксувати механічно ваше безсмертя, можливо нові технології зможуть завантажувати наше життя, але як бути тоді з категорією старіння біологічного організму? Що робити зі старінням тіла? Тут Моравік робить спроби пояснити цей процес за допомогою терміну «емуляція всього мозку», де описує пришвидшена робота мозку, швидкість його роботи, що автор описує як наступний ризик розповсюдження штучного інтелекту. «Одним із наслідків, передбачених у книзі, є 1000-кратне збільшення швидкості розуму, оскільки апаратному забезпеченню, що емулює, не потрібно буде підтримувати емуляцію такою ж повільною, як реальне біологічне явище...» [3, с. 112]. Він припускає, що «суб'єктивної нескінченності» можна досягти шляхом все більшого прискорення часу в кінці Всесвіту. Дійсно, у експоненціальному прискоренні суб'єктивного часу суб'єктивне відчуття нескінченного життя може бути досягнуто за кінцевий проміжок часу [3, с. 143–149]. І тоді постає уявлення про те, що точна симуляція людського розуму в машині така ж хороша, як і оригінал, спонукає нас замислитись над концепцією воскресіння. Моравік пише: «масове воскресіння може бути можливим завдяки використанню величезних симуляторів» [3, с. 123]. І, «Це може бути весело воскресити всіх колишніх мешканців землі таким чином і дати їм можливість розділити з нами (ефемерне) безсмертя трансплантованих розумів» [3, с. 124].

Наступним ризиком, який викликає штучний інтелект, Моравік називає процес ідентичності. Він протиставляє ідентичність тілесну та ідентичність свідомості. Він пише: «Незалежно від того, як буде зроблено копіювання, кінцевим результатом буде нова людина» [3, с. 116]. Це стосується завантаження свідомості, що називається «трансміграцією». Далі він стверджує, що живі тіла продовжують оновлюватися, старі клітини відмирають і створюються нові клітини, а отже, те, що створює їх ідентичність у часі, це постійні моделі. Ця лінія аргументів спонукає його стверджувати, що ідеальна симуляція того, що раніше було людським розумом у біологічному тілі, що відбувається в людиноподібному роботі, є шаблоном ідентичним оригіналу, а отже, ідентичним. З чим дуже важко погодитись і зрозуміти яким чином штучний інтелект зможе генерувати нові клітини в біологічному організмі і яким чином людина все ж таки отримує безсмертя, про яке автор каже як квазі-безсмертя. Це наштовхує на ідею гібрида. Людина має стати чимось схожою на біологічний вид з варіаціями штучного інтелекту чи це все ж таки технології, які замінять людину? Тут він все ж таки порівнює світ машин та світ живих істот. Моравек уявляє собі світ мислячих машин, який виходить за межі людської уяви: «Постбіологічний світ, у якому домінують мислячі машини, що самовдосконалюються, відрізнятиметься від нашого власного світу живих істот, як цей світ відрізняється від мертвої хімії, яка йому передувала» [3, с. 112]. Але не дає чіткої відповіді на те, чи замінять машини людину, чи буде існувати біологічний вид? На це питання він відповідає наступним чином, що при завантаженні розуму весь мозок людини можна змоделювати в людиноподібному роботі, достатньо точному, щоб людина могла вважати результат ідентичним собі. Але однієї людини недостатньо: «Тепер уявіть собі величезний симулятор (я уявляю, що він зроблений із надщільної нейтронної зірки), який може моделювати всю поверхню Землі в атомному масштабі і може запускати час вперед і назад і створювати різні правдоподібні Завдяки великій деталізації цей симулятор моделює живі істоти, включаючи людей, у всій їх складності» [3, с. 124].

Наступним ризиком, який він описує в своїй книзі стосується безпеки. Він вважає, що нестримний розвиток технологій не зупинити, а з рештою збільшиться спокуса самоушкоджень та самознищення. Моравек стверджує, що технологічний розвиток мислячих машин навряд чи буде зупинено в результаті дружньої чи недружньої конкуренції між націями, і оскільки ці машини є ключовими для довгострокового виживання. Оскільки «Якби Сполучені Штати в односторонньому порядку зупинили технологічний розвиток, вони незабаром піддалися б або військовій могутності недружніх країн, або економічному успіху своїх торговельних партнерів. У будь-якому випадку соціальні ідеали, які ведуть до рішення стало б неважливим у світовому масштабі» [3, с. 101]. І далі: «Якби за якимось малоймовірним договором увесь людський рід вирішив уникати прогресу, довгостроковим результатом було б майже

певне вимирання. Всесвіт – це одна випадкова подія за одною. Рано чи пізно нестримний вірус смертельна для людей еволюціонує, або великий астероїд зіткнеться з землею, або сонце розшириться, або нас захоплять зірки, або чорна діра поглине галактику, чим більша, різноманітніша та компетентніша культура, тим краще він може виявляти зовнішні небезпеки та справлятися з ними» [3, с. 126]. У книзі передбачається видатний розвиток нанотехнологій: «Невдовзі звичайні технології, мініатюризовані аж до атомного масштабу, і біотехнологія, її молекулярні взаємодії, зрозумілі в детальних механічних термінах, об'єднаються в цілісний набір методів, що охоплюватиме всі матеріали, розміри та складність. Тоді роботи будуть виготовлені з суміші неймовірних речовин, включаючи, де це доречно, живі біологічні матеріали» [3, с. 126]. У книзі йдеться й про військові дії та прагнення деяких націй або союзів захопити світ: «Будучи мільярди років учасниками невпинної гонки озброєнь один з одним, наші гени нарешті перехитрили самих себе. Вони створили настільки потужну зброю, що вона може перемогти як переможених, так і переможців. Цей пристрій не воднева бомба – широке використання ядерної зброї лише відстрочило б надзвичайно цікавішу загибель, яку було спроектовано» [3, с. 146]. Метафори описують біологічну еволюцію людей, які потім збираються створити мислячу машину, яка потім знищить своїх творців. Що примітно, так це ідея про те, що це станеться навіть у разі ядерної війни, ніби технологічний і соціальний спад, що випливає з цього, був лише незначним кроком назад, який потрібно знову подолати з урахуванням геологічного часу. Не існує ідеї обмеженого вікна можливостей. Він зазначає, що нанотехнології і швидкий технічний прогрес призведе врешті решт до створення нової зброї, яка матиме силу ушкодження одночасну на увесь світ, яка загострює розуміння моралі, вибору та свободи. Коли машина має прийняти рішення знищити інші цивілізації або людський біологічний вид, то категорії моралі, вибору та свободи зневірюються. Тоді як бути з варіантом гібридів? Чи в ним залишиться усвідомлення і загалом свідомість? Чи буде актуальна мораль?

Таким чином, можна зробити висновки, що штучний інтелект стає силою, здатною змінити наше розуміння інтелекту та свободи дій, що призводить до попиту на переосмислення людської сутності та конструкцій ідентичності. Під цим впливом переосмислюється навіть така складна тема, як свідомість. Моравек досліджує етичні дилеми разом із широкими соціальними наслідками створення машин, які зрештою можуть перевершити людський інтелект і моральні здібності. Концептуалізація «дітей розуму» заглиблюється в складності, створені вдосконаленими конструкціями штучного інтелекту щодо людської самобутності та волі. У новому контексті звичайні парадигми людської ідентичності, свободи волі та первинності не лише ставляться під сумнів, але й стикаються з розмиванням, якщо не злиттям, кордонів із машинними аналогами. Філософський дискурс навколо ШІ повинен враховувати потенційні ризики та наслідки, що пов'язані з автономністю та самосвідомістю інтелектуальних машин. Він прогнозує, що розвиток робототехніки та інтелектуальних систем може привести до значних змін у професійній сфері, освіті та повсякденному житті. Це викликає питання про адаптацію людства до нових умов, збереження людської унікальності та важливість етичного підходу до розвитку ШІ. Тому основними загрозами використання ШІ можна назвати: становлення ідентичності, волі, безпеки, етичних норм, в тому числі і встановлення панівного порядку ШІ над людством. Саме Моравек вперше підкреслив важливість створення нормативних рамок та міжнародних стандартів, що можуть запобігти зловживанням та негативним наслідкам використання ШІ, що наразі є актуальним для сучасного світу.

Література

1. Що таке штучний інтелект? Користь та перспективи AI URL: https://viddaleno.com.ua/2024/02/09/shcho-take-shtuchnyi-intelekt/

2. Мороз О. Штучний інтелект. Філософський енциклопедичний словник / В. І. Шинкарук, Л. В. Озадовська, Н. П. Поліщук, І. О. Покаржевська. Київ : Абрис, 2002. 742 с.

3. Moravec H. Mind Children: The Future of Robot and Human Intelligence. Harvard University Press. 1988.

4. Henson Robotics. URL: http://www.hansonrobotics.com/robot/sophia/

5. Yann Lecun. URL: http://yann.lecun.com/

6. Платон Діалоги. Харків. 2008р. URL: https://shron1.chtyvo.org.ua/Plato/Fedon.pdf

Анотація

Нужна Ю. С, Хобта С. В., Котеленець К. М. **Штучний інтелект: філософський дискурс (засновано на аналізі роботи Ганса Моравика «Діти розуму: майбутнє робота та людського інтелекту»).** – Стаття.

У статті розглядаються філософські та етико-гуманістичні аспекти розвитку та впровадження штучного інтелекту в повсякденне та наукове життя людства. Роздуми на задану тему базуються на роботі Ганса Моравека «Діти розуму: майбутнє робота та людського інтелекту». Головною ідеєю статті є думка про те, що людство наближаємося до вододілу в історії життя, а саме до часу, коли кордони між біологічним і постбіологічним інтелектом почнуть розчинятися. В статті підкреслюється головна ідея Моравека, яка полягає у досягненні людської еквівалентності в сучасних машинах не лише в їхній здатності міркувати, але й у їхній здатності сприймати, взаємодіяти та змінювати своє складне середовище.

Щоб досягти чогось, близького до людського, роботам знадобиться принаймні здатність виконувати десять трильйонів обчислень за секунду. Враховуючи трильйонне збільшення обчислювальної потужності з кінця дев'ятнадцятого століття та перспективи екзотичних технологій, які значно перевершують уже знайомі лазери та навіть надпровідники, Моравек робить висновок, що обладнання не матиме проблем з дотриманням цього сорокарічного графіка. Також підкреслюється, що людська еквівалентність є лише початок, а не верхнею межею. Щойно невтомна мислительна здатність роботів буде спрямована на проблему власного вдосконалення та відтворення, навіть небо не обмежуватиме їх ненажерливе дослідження Всесвіту, бо машина не втомлюється, їй не потрібно відновлювати біологічний ритм тощо. Також у статті окреслюються загрози чи ризики, які спонукають людство при розвитку штучного інтелекту, про що говорить автор. Порівнюються філософське розуміння смерті та життя з розумінням безсмертя і смерті в роботі Г. Моравіка. Також окреслюються сценарії майбутнього через мораль та етику. Також у статті зазначається протиріччя у думках автора, це ілюструється тим, що він спочатку відкрито говорить про ризики, такі як безпека, ілюзія безсмертя, прискореність роботи мозку тощо. А потім створює картину, яка спонукає поміркувати про правдоподібне, чудове постбіологічне майбутнє та способи, якими наш розум може брати участь у його розгортанні.

Ключові слова: світогляд, філософський світогляд, робот, штучний інтелект, життя, смерть, етика, мораль, машина, інтелект, мозок.

Summary

Nuzhna Yu. S., Khobta S. V., Kotelenets K. M. **Artificial intelligence: a philosophical discourse (based on an analysis of Hans Moravik's Children of Mind: the Future of Robots and Human Intelligence).** – Article.

The article examines the philosophical and ethical-humanistic aspects of the development and implementation of artificial intelligence in the everyday and scientific life of mankind. Reflections on the given topic are based on Hans Moravec's work Children of Mind: The Future of Robots and Human Intelligence. The main idea of the article is the idea that humanity is approaching a watershed in the history of life, namely the time when the boundaries between biological and post-biological intelligence will begin to dissolve. The article highlights Moravec's main idea, which is to achieve human equivalence in modern machines not only in their ability to reason, but also in their ability to perceive, interact and change their complex environment.

To achieve anything close to a human, robots would need at least the ability to perform ten trillion calculations per second. Given the trillion-dollar increase in computing power since the late nineteenth century, and the prospect of exotic technologies far beyond the already familiar lasers and even superconductors, Moravec concludes that hardware will have no problem meeting this forty-year schedule. It also emphasizes that human equivalence is only a beginning, not an upper limit. As soon as the tireless thinking ability of robots will be directed to the problem of their own improvement and reproduction, even the sky will not limit their voracious exploration of the universe, because the machine does not get tired, it does not need to restore the biological rhythm, etc. The article also outlines the threats or risks that motivate humanity in the development of artificial intelligence, which the author says. The philosophical understanding of death and life is compared with the understanding of immortality and death in the work of H. Moravik. Also outlines the suenary of the future through morality and ethics. The article also notes a contradiction in the author's thoughts, which is illustrated by the fact that he initially openly talks about risks such as security, the illusion of immortality, brain acceleration, etc. And then creates a thought-provoking picture of a plausible, wonderful post-biological future and the ways in which our minds can participate in its unfolding.

Key words: worldview, philosophical worldview, robot, artificial intelligence, life, death, ethics, morality, machine, intelligence, brain.