

ФУТУРОЛОГІЧНІ ПРОГНОЗИ НІКА БОСТРОМА ТА РЕЯ КУРЦВЕЙЛА: ЧОГО СПОДІВАТИСЬ ЛЮДСТВУ

¹ Вінницький національний технічний університет

Анотація

В роботі наведено основні футурологічні прогнози Ніка Бострома та Рея Курцвейла. Проаналізовано відмінності і подібності у поглядах двох футурологів.

Ключові слова: футурологічні прогнози; штучний інтелект; Рей Курцвейл; Нік Бостром.

Abstract

The paper presents the basic futurological projections of Nick Bostrom and Ray Kurzweil. The differences and similarities in the views of the two futurologists are analyzed.

Keywords: futurological forecasts; Artificial Intelligence; Ray Kurzweil; Nick Bostrom.

Вступ

Інтенсивний технологічний розвиток є однією з ключових особливостей початку XXI століття. Серед різноманіття технологій сучасності визначальне місце посідає цифровізація як виробництва, так і соціальних взаємодій. Ще кілька десятиліть тому назад комп'ютер був завбільшки з невелику кімнату, сьогодні звичайний смартфон має значно потужніший процесор, обсяг пам'яті й цілу низку різноманітних функцій про які тільки могли мріяти розробники в середині XX ст. Людство швидким темпом породжує численні технічні пристрої, виробляє нові матеріали, змінює на свій розсуд середовище свого існування, а разом навколишній світ. Приріст наукового знання, винаходів інженерів просто неймовірний, але за цими успіхами є й інший бік медалі, є свої ризики та загрози. Усвідомлення небезпеки нестримного науково-технічного розвитку, знаходження мінімізації негативних впливів на людину та природу техногенного впливу є першочерговим завданням як філософів, футурологів, так самих розробників нових технологій та пристроїв. Зокрема, це стосується й розробки штучного інтелекту.

Результати дослідження

Сьогодні спостерігається зникнення цілого ряду професій, що породжує великий рівень безробіття і зменшення рівня життя деяких груп населення, широка автоматизація у більшості сфер життя, дуже серйозно впливає на практику та навички людини. Через високий рівень автоматизації у пілотів менше практики управління літаком, що в результаті впливає на безпеку польотів, хоча технології здавалося б мали її підвищити. Втрата навиків через впровадження технологій може вплинути на всіх, хто працює в офісах і на промислових підприємствах [1].

Дані про людину, які доступні в мережі, зменшують рівень приватного життя особи та захисту конфіденційної інформації, що призводить до великих хакерських атак. Розвиток технологій значно модернізує сучасні види зброї та збільшує негативні наслідки, що в деякій мірі можуть бути фатальними для людства.

У XIX столітті, особливо чітко це почало виявлятися у другій його половині, наука набула статусу виробничої сили. Знання яке здобували науковці отримало статус не просто цінності як істини, а як надзвичайно практичного й ефективного засобу в повсякденному житті людини через реалізацію в технічних розробках і технологіях виробництва. В цей час філософи, вчені, письменники (Ж. Верн, Г. Веллс) починають перейматись наслідками розвитку науки та технологій, що згодом призвело до виникнення футурології. Сам термін «футурологія» з'явився 1943 році з легкої руки німецького

соціолога Осипа Флехтгема, чимале сприяння в популяризації футурології зіграла робота Роберта Юнга «Майбутнє вже почалось», яка побачила світ у 1952 році та « Сума технологій» (1963) Станіслава Лема, який спробував спрогнозувати як технології будуть змінювати людину, суспільство й природу. Особливе місце в розвитку футурології належить Римському клубу (заснований в 1968 році з ініціативи видатного італійського економіста і підприємця Ауреліо Печчеї), який формувався як незалежна від урядів, корпорацій, релігійних чи політичних організацій експертна група для оцінки довгострокових наслідків існуючих проблем та майбутнього людства. До основних цілей Римського клубу належать:

- виявлення найбільш суттєвих проблем, які визначатимуть майбутнє людства на основі комплексного та перспективного аналізу;
- аналіз та оцінка альтернативних сценаріїв майбутнього;
- розробка практичних пропозицій щодо виявлених проблем;
- стимулювання широкої суспільної дискусії.

У першій доповіді Римського клубу було сформульовано концепцію «меж зростання», яка не втратила своєї актуальності й через пів століття, хоча в дещо іншому аспекті, передусім це стосується розвитку інформаційних технологій та зокрема, штучного інтелекту.

Прогнозування розвитку складних, багатокомпонентних систем, які здатні до самоорганізації надзвичайно важка справа. Враховуючи особливість такого об'єкту, як людське суспільство, не завжди можливо передбачити стовідсоткову ймовірність тенденції чи проблеми, що відбудеться в майбутньому. Наприклад, Джон Уоткінс у своїй статті «Що може відбутися протягом найближчого століття» зробив і кілька прогнозів, які не збулися. Він, зокрема, припустив, що з англійського алфавіту зникнуть літери С, Х і Q, у великих містах не буде машин, а в природі – комарів [2]. Зокрема, йому досить точно вдалося описати мобільні телефони, телебачення, цифрову фотографію, літаки-розвідники, танки та інше.

Обрі де Грей, пророкує майбутнє, в якому ми зможемо змінити свої тіла на клітинному і молекулярному рівні з метою усунення пошкоджень або навіть запобігти старінню [3]. Мічіо Кайку розмірковує як наука вплине на повсякденне життя у XXI ст., Френсіс Фукуяма пише про кінець історії і останню людину та цілий ряд інших мислителів намагаються зазирнути в майбутнє, побачити хоча б основні його обриси й можливо запобігти тим загрозам, які на нас чекають.

До числа найвідоміших сучасних футурологів належать Рей Курцвейл і Нік Бостром. Їхні прогнози дуже популярні серед інтелектуалів, але далеко не всі в усьому з ними погоджуються, проте однозначно можна стверджувати, що вони активізують процеси рефлексії щодо майбутнього людства.

Курцвейл не був першим, хто вирішив, що машини врешті решт перевершать людський розум, але він сміливо визначив дату технологічної сингулярності. У 2005 році Курцвейл проголосив, що «синтетичний інтелект» з'явиться в 2045 році й не тільки перевершить можливості людського мислення, а й буде в мільярд разів розумніший, ніж все людство сьогодні. При цьому Курцвейл не переживає, що деякі злісні машини забажають знищити людство, як в «Термінаторі». Навпаки, в майбутньому людина і машина будуть тісно пов'язані для досягнення неймовірних етапів прогресу і інновацій. Курцвейл вважає, що люди стануть більш залучені в штучний інтелект, ніж ми можемо уявити, а в 2030 році багато з людських органів будуть замінені крихітними роботами, які будуть працювати довше, ніж наші, і куди більш ефективно[3].

Прогнози Рея Курцвейла, щодо штучного інтелекту можна звести до таких тез:

- У мене є стійка дата - 2029 рік - для цього передбачення. І сюди входить не тільки поняття логічного інтелекту. Сюди входить емоційний інтелект, який може бути веселим, жартівливим, милим, люблячим, розуміючим. Це найскладніше, що ми можемо зробити. Це те, що розділяє комп'ютери і людей зараз. Але я вірю, що до 2029 року ця прірва зникне;

- Я вважаю, що ми досить близькі до необхідного рівня програмного забезпечення. Частково нам допомагає глибоке розуміння того, як працює людський мозок, і воно стає дедалі більше. Зараз ми можемо зазирнути всередину живого мозку і побачити як формуються думки в режимі реального часу. Ми можемо побачити, як мозок думає, і як думки формують розум;

- Розуміння методів роботи людського мозку допоможе нам створити схожі на нього машини. Іншим важливим винаходом стане інтерфейс, який з'єднує наші мізки з комп'ютерами. Мені здається, що в найближчі десятиліття технології такого роду отримають широке розповсюдження;

• Еволюція йде непрямим шляхом: для того, щоб створити наступну стадію або епоху використовуються інформаційно-процесуальні методи з попередньої. Я концептуалізував історію еволюції - як біологічної, так і технологічної, - і представив її у вигляді шести епох. Сингулярність почнеться в п'ятій, а в шостій пошириться з Землі на весь Всесвіт;

Ми вже пройшли через п'ять парадигм (електромеханічні калькулятори, релейні комп'ютери, вакуумні трубки, дискретні транзистори і інтегральні схеми), і чим далі, тим кращим ставало співвідношення ціна-якість і обчислювальні можливості. Кожен раз, коли ми досягали меж однієї парадигми, її місце займала інша. На горизонті вже видніються обриси шостої, де обчислення будуть здійснюватися на молекулярному рівні [4].

Шведський філософ Нік Бостром, директор оксфордського Інституту майбутнього людства, прославився своїми роботами про екзистенціальні ризики, антропний принцип, етику удосконалення людини і загрози, які становить штучний інтелект. Його перу належить «Доказ симуляції», в якому він припускає, що наша реальність може бути імітацією на комп'ютері, створеному якоюсь просунутою цивілізацією.

Погляди Ніка Бострома, що стосуються штучного інтелекту можна викласти в таких тезах:

• штучний інтелект вже сьогодні перевершує людський в багатьох областях. Так, протягом багатьох років різні види штучного інтелекту перемагають чемпіонів всіляких ігрових турнірів, будь то шахи або покер. Такі досягнення можуть і не здаватися особливо вражаючими, але лише тому, що наші вимоги до дивовижного швидко адаптуються до прогресу;

• для нас важливо створити штучний інтелект, у якого вистачить розуму вчитися на своїх помилках. Він буде здатний нескінченно удосконалювати себе. Перша версія зможе створити другу, яка буде кращою, а друга, будучи розумнішою оригіналу, створить ще більш просунуту третю і так далі. У певних умовах такий процес самовдосконалення може повторюватися до тих пір, поки не буде досягнутий інтелектуальний вибух - момент, коли інтелектуальний рівень системи підскочить за короткий час з відносно скромного рівня до рівня суперінтелекту;

• штучний інтелект може бути менш людяний, ніж прибулець. Немає нічого дивного, що будь-якого розумного прибульця можуть спонукати до дії такі речі, як голод, температура, травми, хвороби, загроза життю або бажання завести потомство. Штучний інтелект, по суті, ніщо з перерахованого цікавити не буде. Можна уявити існування штучного інтелекту, чиєю єдиною кінцевою метою буде перерахувати всі піщинки на острові Боракай або знайти десяткове подання числа π .

• людина - найбільш ненадійна система. Сьогодні хакери часто звертаються до принципів соціальної інженерії з метою отримання доступу до чужого комп'ютера. А якщо хакером-маніпулятором виявиться суперінтелект, то можна припустити, що він з легкістю знайде собі спільника або ж просто стане використовувати нас проти нашої волі як свої руки і ноги."

• людина сама по собі представляє корисний ресурс (зручно згруповані атоми), а її виживання і процвітання залежить від інших ресурсів. Результатом розвитку штучного інтелекту, що потребує цих ресурсів, може дуже легко стати зникнення людства. Так настане час технологічно високорозвиненого суспільства, що включає в себе безліч складних структур, багато з яких будуть розумнішими за все, що є сьогодні на планеті. Це буде час економічних і технологічних чудес. Але нікого, хто б міг цим скористатися, не залишиться. На Землі запанує Діснейленд, в якому більше не буде дітей;

• якщо інтелектуальний вибух загрожує нам зникненням, то ми повинні зрозуміти, чи можемо ми контролювати процес детонації. Сьогодні було б більш розумним прискорити роботу щодо вирішення проблеми контролю, ніж зупиняти проведення досліджень в області штучного інтелекту. Але поки вирішенням проблеми контролю займаються кілька людей, в той час як над створенням штучного інтелекту трудяться десятки, якщо не сотні тисяч;

• найрозумнішим вчинком дитини бомбою, що з цокає, в руках було б акуратно покласти бомбу на підлогу, швидко вибігти з кімнати і покликати дорослого. Але що якщо вся наша кімната повна дітей і кожна дитина має вільний доступ до пускового механізму. Шанси, що ми всі покладемо небезпечну іграшку на підлогу, вкрай малі. Який-небудь маленький ідіот обов'язково натисне на кнопку, аби побачити, що станеться [5].

Аналізуючи позиції Курцвейла та Бострома може зазначити, що вони передбачають виникнення суперінтелекту, не враховуючи сферу, де він буде винайдений. Також однакова позиція в тому, що через деякий час суперінтелект буде набагато розумніший людського або всіх людей разом.

Основою для штучного інтелекту буде копіювання принципу роботи та механізму з'єднань людського розуму.

Відрізняються позиції щодо сумісного життя штучного інтелекту та суспільства. Курцвейл відноситься оптимістично, бачить майбутнє в дружньому співіснуванні людей з штучним інтелектом, що покращить наше життя і підніме на наступний етап розвитку. Бостром, виявляється більш реалістичним, бачить в штучному інтелекті загрозу знищенню людського виду так, як суперінтелект зможе еволюціонувати до найрозумнішої істоти на планеті та мати ціль знищити все живе.

Ознайомившись з прогнозами Ніка Бострома та Рея Курцвейла можна зробити висновок, що штучний інтелект буде винайдений, враховуючи темпи розвитку та зацікавленість вчених у даній сфері й інвестиції великих технологічних компаній. З кожним наступним роком штучний інтелект буде вдосконалюватися, розширювати свої функціональні можливості, здатність до самонавчання, проте залишається відкритим питання щодо меж зростання штучного інтелекту. Сьогодні ми спостерігаємо як штучний інтелект проникає в усе нові й нові сфери нашого життя, замінює людину в багатьох виробничих позиціях, спрощує виконання багатьох операцій. Але при цьому виникають дві ключові загрози.

По-перше, чи не набуде штучний інтелект статусу суб'єкта (не просто ми йому цей статус припишемо), чи не буде він самостійно визначати цілі своєї діяльності? Якщо це станеться, то яким чином ми можемо запобігти, щоб ці цілі не йшли в розріз з цілями людства? Чи можемо реально забезпечити виконання законів робототехніки Айзека Азімова?

По-друге, чи не станеться використання штучного інтелекту (у варіанті суперінтелекту) в інтересах вузького кола людей чи окремих особистостей? Які нові вимоги постають в сфері етики науки, в сфері інженерії, створення нових продуктів, які будуть активно взаємодіяти з зовнішнім середовищем і людиною, напрацьовувати новий досвід, якого не було в них при створенні?

Футурологічні прогнози як Ніка Бострома так і Рея Курцвейла хибують лінійною екстраполяцією наявних тенденцій чи то стосовно штучного інтелекту, чи то загалом майбутнього людства в цілому. В їхніх прогнозах бракує здорового скептицизму, вони немов забувають знамениту річ у собі Імануїла Канта. Ми ж лише знаємо про світ (той чи інший об'єкт) лише в нашому сприйнятті й ніколи не маємо повноти знання про нього, тому наші припущення завжди є лише певним наближенням до того як є насправді чи як воно буде. Проблема прогнозування дуже складна, це зумовлено тим, «що ми живемо не Середньостані, а в Крайньостані. Наші прогнозисти більш-менш дають собі раду з чимось звичайним. Але надзвичайне їм не по зубам, воно завдає їм нищівної поразки» [6, с. 151]

Висновки

Особливо варто звернути уваги на сферу штучного інтелекту, оскільки вона швидко розвивається і на даний момент є не врегульована законодавством. Також потрібно слідкувати за використанням наукових знань, праць і дослідів у відповідних сферах, для покращення життя, а не пришвидшення його знищення високотехнологічною зброєю. Людина завжди залишається основною причиною своїх проблем, тому ми можемо знищити себе швидше, ніж штучний інтелект розвинеться.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Розвиток технологій: позитивні та негативні сторони [Електронний ресурс] : Анна Оніщук / Bezoman. – Режим доступу до ресурсу: <https://bezoman.pp.ua/uk/questions/69/rozvitok-tekhnologiy-pozitivni-ta-negativni-storoni>.
2. ТОП-10 столітніх пророкувань, які збулися з лякаючою точністю [Електронний ресурс]: Сьогодні. – Режим доступу до ресурсу: <https://www.segodnya.ua/ua/lifestyle/fun/top-10-stoletnih-predskazaniy-kotorye-sbylis-s-pugayushchey-tochnostyu-661996.html>.
3. 10 самых выдающихся футурологов современности [Електронний ресурс]: Илья Хель / Hi-news.ru. – Режим доступу до ресурсу: https://hi-news.ru/technology/chtivo-10-samyx-vydayushhixsya-futurologov-sovremennosti.html#obri_di_grej
4. Футуролог Рэй Курцвейл – О победе искусственного интеллекта и шести эпохах эволюции [Електронний ресурс]: Артем Слободчиков/Apparat. – Режим доступу до ресурсу: <https://apparat.cc/world/kurtzveil-on-ai/>.
5. 7 цитат філософа Ніка Бострома – о том, почему искусственный интеллект может погубить человечество [Електронний ресурс]: Владимир Мальженков /Apparat. – Режим доступу до ресурсу: <https://apparat.cc/world/ai-danger/>.

6. Талєб Насїм Ніколас. Чорний лебїдь. – К.: Наш формат, 2017, - 392.

Теклюк Анатолій Іванович — кандидат філософських наук, доцент кафедри філософії та гуманітарних наук, директор Центру соціальних комунікацій і лідерства ВНТУ, Вінницький національний технічний університет, Вінниця, e-mail: gcvr216@gmail.com.

Кирган Ольга Анатоліївна – студентка групи ЛОТ-19, факультет комп'ютерних систем та автоматики, Вінницький національний технічний університет, Вінниця.

Алексєєв Максим Андрійович — студент групи МНТ-19, факультет інфокомунікацій, радіоелектроніки та наносистем, Вінницький національний технічний університет, Вінниця.

Tekliuk Anatolii Ivanovich - Candidate of Philosophy, Associate Professor, Department of Philosophy and Humanities, Director of the a center for social communication and leadership VNTU, Vinnytsia National Technical University, Vinnytsia, e-mail gcvr216@gmail.com.

Kirgan Olga Anatoliivna - student of LOT-19 group, Faculty of Computer Systems and Automation, Vinnitsa National Technical University, Vinnitsa.

Alieksieiev Maksym Andreevich — student of MNT-19 group, Faculty of Information Communications, Radio Electronics and Nanosystems, Vinnitsa National Technical University, Vinnitsa.