

УДК 159.922:004.8

СУЧАСНА ФІЛОСОФІЯ СВІДОМОСТІ ПРО ПЕРСПЕКТИВИ СТВОРЕННЯ ШТУЧНОГО ІНТЕЛЕКТУ

Олександра Стебельська

*Національний університет «Львівська політехніка», кафедра філософії
вул. С. Бандери 12, 79013, м. Львів, Україна*

Любомир Федорів

*Національний університет «Львівська політехніка», кафедра філософії
вул. С. Бандери 12, 79013, м. Львів, Україна*

Ця стаття присвячена аналізу перспектив створення штучного інтелекту. Автори концентрують увагу на тих труднощах, з якими стикаються дослідники, намагаючись реалізувати інтелектуальні, творчі здібності людини, її етичні та естетичні принципи, мовлення та звичайну буденну поведінку. У результаті критичного аналізу стверджується, що інтелектуальні можливості людини володіють такими унікальними властивостями, які не дають змогу на даний час відтворити в машинному середовищі щось подібне за складністю та гнучкістю функціонування. Проте це не повинно зупиняти науковців у спробах, оскільки їх прагнення не лише спричиняють прогрес у сфері штучного інтелекту, але й дають змогу краще зрозуміти нашу власну природу.

Ключові слова: свідомість, штучний інтелект, сингулярність, свобода, творчість.

Постановка проблеми. Проблема свідомості є однією з провідних проблем в історії філософії. Із давніх-давен людина ставила питання про те, ким вона є, яким є її призначення, що таке пізнання і якими є його межі; в центрі її зацікавлень поставали граничні поняття та відношення «буття-небуття», «життя-смерть», «душа-тіло», «добро-зло», «права-брехня». Вирішення всієї проблематики неможливе поза розумінням природи людської свідомості та її значення для людського буття у світі. Тому природа свідомості поставала в центрі праць багатьох мислителів. Намагаючись осмислити суть людської свідомості, мислителів цікавила можливість створення певного штучного аналогу, що стимулювало розвиток сфери штучного інтелекту (ШІ). Це були амбітні прагнення створити собі подібного, випробувати свої можливості, ввійти у сферу абсолютно невідомого та забороненого. Проте прагнення відтворити себе в машині були зумовлені не лише науковим зацікавленням; одночасно вони призводили до кращого розуміння саме людської природи і загострювали питання суті людської свідомості.

Починаючи із середини ХХ ст. розвиток сфери ШІ набув надзвичайної сили, потужності та динаміки. Він носить експоненціальний характер, і тому багато науковців замислюються не просто над значущістю тих досягнень, що вже є, але й над тим, чого людина здатна досягти загалом. Такі футурологи, як Н. Бостром та Р. Курцвейль, відкрито говорять про те, що в недалекому майбутньому людина створить настільки складні технології, що настане час сингулярності – перехідний момент, коли штучний інтелект досягне і навіть перевершить людські можливості (Boyles 2018; Bostrom 2003; Kurzweil 1999). Відповідно, перед людством постає багато питань: чи взагалі ми здатні створити *настільки* розумні машини? Чи будуть ці машини безпечними? Чи будуть вони володіти емоціями та почуттями?

Наскільки ці машини будуть гнучкими в поведінці, мовленні, мисленні? Чи володітимуть вони ціннісним виміром? Чи зможуть вони любити? Чи зможуть вони навчатись? Чи не призведе створення таких машин до зникнення людства? Ці питання стоїть у центрі уваги всіх науковців, сфера зацікавлень яких обертається навколо проблеми свідомості та штучного інтелекту. І досі науковці не можуть дати нам однозначних відповідей.

Метою статті є аналіз розвитку сфери штучного інтелекту, дослідження можливостей, перспектив та наслідків його реалізації, а також впливів, які здатні здійснити технології не лише на наше особисте, але й на суспільне життя. Важливим є усвідомлення того, що розвиток сфери ШІ – це не просто змагання наших здібностей, амбіцій та прагнень, а спроба зрозуміти, ким ми є.

Розроблення проблематики в працях дослідників. Серед відомих сучасних дослідників у сфері філософії свідомості та штучного інтелекту слід відзначити Д. Чалмерса, Дж. Серла, Д. Деннета, Р. Доукінза, Р. Пенроуза, Н. Бострома, Р. Курцвейля, Дж. Мура, В. Валлаха, К. Аллена, І. Сміт, С. Брінгсйорда, П. Белло, М. Андерсона, С. Л. Андерсон, Л. Мюльхаузера, Л. Хелма, Д. МакДермотта, М.І. Ганеш, Х. Дрейфуса, Р.Дж. Бойлза, І. Мунтяна, Д. Ховарда, М. Боден тощо. У наукових доробках вони загострили проблеми свободи, творчості, уяви, кваліа, спонтанності та гнучкості людської поведінки, природи етичних принципів та орієнтирів. У центрі їх зацікавлень постали характерні риси людської свідомості та можливість їх реалізації в технічному середовищі.

Виклад основного матеріалу. Перші дослідження людського інтелекту розпочались ще в добу античності і пов'язані з досягненнями Аристотеля у сфері логіки. Проте серйозно вищезгадана сфера почала розвиватись лише в середині ХХ ст. Достатньо згадати дві події. 1950 року Алан Т'юринг видав статтю «Обчислювальні машини і розум», де запропонував замінити питання *Чи здатна машина думати?* на питання *Чи може машина діяти таким чином, щоб її дії було неможливо відрізнити від людських?* (важливо звернути увагу на постановку питання: пріоритетним було не визначення розумності машини, а оцінка її поведінки та (не)можливості відрізнити її від поведінки звичайної людини). Ще одним ключовим моментом в історії штучного інтелекту можна вважати Дартмутський семінар 1956 року, де вперше і було сформульовано термін «штучний інтелект», а також Н'юелл та Саймон презентували програму Logic Theorist (LT), яка була здатна доводити елементарні теореми. На семінарі були присутні 10 науковців: Дж. Маккарті, М. Мінський, К. Шеннон, А. Самюель, Г. Саймон, А. Н'юелл, Т. Мур, Р. Соломонофф, О. Селфрідж, Н. Рочестер (Bringsjord 2018). Із того часу дослідження проводились із перемінним успіхом, були періоди як застою, так і активізації експериментів, змінювались підходи до розуміння принципу роботи самого штучного інтелекту (символьний, нейросітки, модульний). Нині комп'ютери досягли надзвичайно високого рівня потужності та складності. Їх застосовують у сільському господарстві, промисловості, військовій сфері, побуті, сфері розваг. Важливо відзначити, що сфери філософії свідомості та штучного інтелекту надзвичайно тісно переплітаються, стимулюючи розвиток одна одної. Дослідження у сфері філософії свідомості стимулюють розвідки у сфері штучного інтелекту. у свою чергу, дослідження у сфері ШІ сприяють кращому розумінню того, хто ми є, як ми мислимо, на що ми здатні.

Сучасні дослідники досить акуратно висловлюються стосовно перспектив створення розумних машин. Це пов'язано з рядом причин. По-перше, ми не можемо однозначно та безапеляційно щось стверджувати, особливо коли це стосується можливості створення штучного інтелекту. Алан Т'юринг прогнозував, що до початку ХХІ ст. усі машини проходять тест Т'юринга (ТТ), але, як ми бачимо, цього не сталося повною мірою. Врешті-решт, час розставити все на свої місця. По-друге, якими б шокуючими не були

сучасні дослідження у сфері ШІ (серед останніх проривів можна згадати здатність машини перетворювати мозкові сигнали у членороздільне мовлення), все ж таки ШІ ще не досягнув такого рівня, щоб зрівнятися із людським. Він обмежений певною конкретною сферою, йому бракує гнучкості та спонтанності, йому невідомі людські почуття та труднощі етичного вибору (про це йтиметься нижче).

Ситуація ускладнюється тими протиріччями та нестиківками, що пронизують сферу штучного інтелекту на глибинному рівні. У першу чергу, нечітким є застосування самого терміна «штучний інтелект». У книзі «Штучний інтелект: сучасний підхід» С. Рассел та П. Норвіг класифікують різноманітні визначення за двома основними категоріями: одні формулювання стосуються мисленнєвих процесів та способів розмірковування, інші – поведінки. В одних визначеннях успіх зумовлюється достовірним відтворенням здібностей людини, в інших – кінцеві досягнення розглядаються через призму раціональності (Рассел, Норвіг 2006, 35). Неоднозначним є застосування і самого терміна «інтелект» (Суворов 2018). Якщо під цим терміном ми маємо на увазі лише калькуляційні здібності людини, то ШІ чудово виконує обчислювальні завдання. Якщо ж ми розглядаємо термін «інтелект» дещо ширше, невід'ємно від людських емоцій, почуттів, інтуїції, творчості та уяви, етичних орієнтирів, як певну духовну силу, то його реалізація стає надзвичайно складним завданням і є рівноцінною реалізації самої людської свідомості. З одного боку, науковці виявляють обережність і, на перший погляд, не ставлять собі за мету змодельовати людську свідомість, усвідомлюючи складність цієї задачі. Вони займаються реалізацією конкретних програм, які могли би чудово рахувати, доводити теореми, розв'язувати математичні головоломки, бути чудовими гравцями в різних іграх або виконувати певні конкретні завдання. З другого боку, дії науковців свідчать про надзвичайно амбітні плани та прагнення. Свідомо чи ні, дослідники прагнуть реалізувати увесь потенціал людської свідомості. Як інакше трактувати спроби змодельовати не лише поведінку, але й внутрішній світ людини, її почуття, емоції, етичні орієнтири, творчі здібності?! Якщо вчені (свідомо чи ні) рухаються в напрямку моделювання людської свідомості, то постає питання, якою є її природа? Такі сучасні мислителі, як Дж. Серл, Д. Деннет, Д. Чалмерс, Р. Доукінз, П. та П. Черчленди, загострюють *body-mind problem* та висловлюють власні погляди на можливості реалізації свідомості на штучному носії (Chalmers 2010; Chalmers 1996; Searle 1992; Юлина 2004; Деннет 2004; Черчленд П. і П. Черчленд; Доукінз). Виникає питання: якщо ми достеменно не знаємо природи свідомості, то що саме ми збираємось реалізувати?!

Звертаючись до конкретних досліджень в сфері ШІ, необхідно визнати, що науковці значно просунулись у своїх спробах втілення людських можливостей та здібностей. Достатньо згадати різноманітні пристрої, девайси, програми та системи. Комп'ютер вже давно обіграє людей у різноманітних іграх, зокрема шашки, шахи та го. А нещодавно робот зі штучним інтелектом AlphaStar переміг професійного гравця в складній стратегії StarCraft II (Хель, 2019). Експертні системи здатні бути повноцінними помічниками людині. Необхідно визнати, що відсоток помилок, здійснений експертними системами, постійно скорочується. Дослідники створили андроїда Софі, яка стала громадянкою Саудівської Аравії. Софі здатна вести розмови, посміхатись, проявляти емоції. Надзвичайної популярності здобули різноманітні розумні системи, такі як розумний будинок, який здатен полегшити наше життя, виконуючи буденну та механічну роботу. Розвинулась і сфера розпізнавання образів.

Разом із тим ми не можемо заперечити того факту, що якими б шокуючими подеколи не були відкриття, вони, як правило, обмежені певними конкретними сферами, в яких машина «обходить» людину. Автомобілі-автопілоти діють згідно із чіткими

правилами, прописаними для конкретних обставин та ситуацій; системи безпеки та розпізнавання облич орієнтовані лише на виконання певного виду роботи; ігрові системи здатні перемогти людину лише в певному виді гри. Проте природа людської свідомості є набагато універсальнішою: людина може бути (чи стати) ким завгодно, в неї безліч можливостей та перспектив розвитку, вона здатна ефективно діяти не лише за наявності чітко прописаних правил, але й за відсутності інформації загалом, її діяльність не обмежується однією сферою. Коли ми починаємо занурюватись у проблему глибше, то виявляється, що сфера мовлення, поведінки, емоцій та почуттів далека від повноцінної реалізації в технічному середовищі. Машинам бракує гнучкості та спонтанності дій. Серед основних проблем, з якими стикаються науковці, намагаючись створити розумні машини, можна відзначити такі (McCarthy 2006, Dreyfus 1972):

1) створюючи інтелектуальні машини, науковці не можуть наділити їх розумінням, у тому числі розумінням самих себе (або самосвідомістю). Комп'ютери здатні обробляти велику кількість інформації, проте не здатні її осмислити. Цю проблему піднімав Дж. Серл, провівши мисленневий експеримент «китайської кімнати». Машина маніпулює символами, проте не розуміє те, що робить. Якщо спитати комп'ютер про його думку чи його ставлення до інформації, він не зможе задовольнити нашу цікавість. Потужні ігрові системи можуть обігрувати людей в шахи чи різні стратегії, проте вони не здатні захотіти грати, продовжити змагання чи просто зрозуміти, що вони виграли. Сенсовним виміром здатна оперувати нині лише людина. Також машина позбавлена самосвідомості. Звісно, вона може фіксувати наявність певних операцій, виявити їх послідовність та взаємозв'язок, проте вона не володіє знанням першої особи, суб'єктивністю, а значить стосовно неї неможливо вести мову про свободу та відповідальність;

2) другою перепоною на шляху створення ШІ є *проблема невизначеності*:

– ми живемо у плінному світі, який постійно змінюється. Відповідно, ми маємо справу з мінливою, неточною та суперечливою інформацією. Будь-яка нова інформація може змінити хід розмірковувань, сприйняття ситуації або оцінку обставин. Наша свідомість безперервно знаходиться в ситуації оновлення та перезавантаження. Кодування та збереження інформації в комп'ютерному середовищі у формі фреймів спричинило нерозуміння процесів її оновлення. Виникає питання: яким чином необхідно налагодити зв'язок між штучною системою та реальністю? Коли та яким чином система буде знати, що потрібно змінити, доповнити та уточнити те знання, яке в ній вже закладене? Створюючи розумні машини, ми повинні наділити їх здібністю виявляти помилки, нестиковки, протиріччя в інформаційному потоці та виправляти їх. Тільки в такому випадку інтелектуальна машина буде мати змогу виробити відповідність між тим, що вона знає, і тими фактами, що її оточують;

– мовленнєва сфера людини є надзвичайно контекстною та багатозначною. Недарма пізній Л. Вітгенштайн сприймав мовлення за гру, гру контекстів та значень. Одна і та ж сама фраза чи слово можуть сприйматись по-різному в залежності від того, коли, ким, за яких умов і якою інтонацією вони були сказані. Окрім того, засвоєння мови передбачає застосування певних правил, але далеко не завжди їх чітке усвідомлення. Знання та застосування мови часто носять інтуїтивний характер. Відповідно, надзвичайно складно навчити машину говорити настільки ж природно, як це робить людина. Це потребує безпосередньої залученості машини в людське середовище, безперервного тісного спілкування, не обмеженого формальними ситуаціями;

– будь-яка штучна система володіє чітко сформульованими твердженнями. На основі цих тверджень вона аналізує, робить висновки та діє. Людська поведінка є набагато

складнішою, оскільки передбачає наявність певних прагнень, бажань, переконань, мотивів, ірраціональних компонентів;

– надзвичайно складним завданням для сучасних машин є розпізнавання образів. Людська свідомість бачить будь-який образ цілісно, тому будь-які його зміни або спотворення («шуми») не будуть настільки різко впливати на його сприйняття. Якщо ми почуємо видозмінену мелодію, ми все рівно її впізнаємо. У той час як машини фіксують чіткі риси, тому будь-яке їх викривлення призведе до помилки ідентифікації;

3) невід'ємною складовою частиною людини є її тілесність, через призму якої вона здатна сприймати світ, в якому живе. Тілесність допомагає людині відчувати та переживати світ, вступати з ним у тісний контакт. Завдяки тілесності людина вплетена в різноманітні відношення та взаємини, а значить, постійно перебуває в нерозривній внутрішній єдності зі світом. Жодна з існуючих машин поки що не має такого тісного та глибокого зв'язку з реальністю;

4) складною проблемою в процесі реалізації ШІ є ціннісний вимір людини. Будь-які дії та ситуації людина оцінює не лише з позиції їх користі та практичності, але і з позиції їх значущості та важливості для неї та інших. Необхідність створення моральних машин зумовлена: а) страхом людства за своє майбутнє, адже розумні та могутні машини можуть знищити людський вид; б) прагненням поглибити наші власні знання про природу людини. Для реалізації моральних машин необхідно, щоб вони володіли свідомістю, інтенціональністю та свободою волі (Moog, 2006). Досі жодна машина не досягла такого рівня складності. Причиною цього є не лише технічні перешкоди, недосконалість методологічного апарату, але й природа самих ціннісних норм, застосування та функціонування яких часто залежить від ситуації та контексту (Dreyfus, 1972; Muehlhauser, Helm, 2012). Часто людський досвід взагалі важко однозначно описати в термінах «добре» чи «погано», оскільки це пов'язано зі складністю, неоднозначністю, проблематичністю самої людської природи. Людська поведінка настільки гнучка, що це створює проблеми під час спроби її машинного моделювання. Відповідно, постає питання, якими саме принципами повинна володіти машина? Чи можливо технічно прописати всі можливі варіанти застосування тих самих етичних моделей? Чи може машина володіти відповідальністю?

5) проблема відповідальності чітко взаємопов'язана з проблемою свободи. Дискусії стосовно того, чи людина володіє свободою волі, точаться досі (компатибілістські та імкомпатибілістські напрями). Така невизначеність стосовно існування свободи волі спричиняє не менш гострі дискусії стосовно можливості автономної поведінки штучного інтелекту;

6) ще однією причиною тих труднощів, з якими стикаються дослідники, є природа квалій. Більше того, мислителі не можуть дійти навіть згоди стосовно їх існування (Nagel, 1974; Searle, 1992; Chalmers, 1996; Деннет, 2004). Як реалізувати те, природа чого є досі загадкою? Проблемою постає і метод реалізації суб'єктивності бачення на штучному носії;

7) однією з провідних особливостей людини є творчий характер її діяльності, здатність створювати щось таке, чого до неї в природі не існувало і існувати не могло. Незважаючи на те, що сучасні комп'ютери здатні писати вірші, музику, малювати, рівень їх креативності неможливо зрівняти з людським. Справа в тому, що творчий процес полягає не лише в певному продукуванні інформації, але й у здатності *осмислити* цінність та значущість винайденого, на що здатна лише людина. Окрім того, творча людина завжди повинна розуміти наслідки застосування своїх винаходів та нести за них відповідальність. Комп'ютер не володіє такими ціннісними критеріями, вони задаються людиною. Комп'ютер зосереджується на кількісних, а не на якісних параметрах. У такому разі «творчість» комп'ютерів на даному етапі розвитку штучного інтелекту зводиться до механічного перекомбінування

певного матеріалу. Але чи є творчість перекомбінуванням? Мабуть, відповідь буде негативною. Саме комбінування елементів творчого процесу ніколи би не спричинило виникнення чогось абсолютно нового. Це завжди було би повторенням та іншим варіантом того, що вже існує. Тим не менше, сама людина впродовж історії доводить своїми винаходами, що, використовуючи природний матеріал, вона створює речі, які є проривом у природному бутті. До людини нічого подібного у природі не існувало (Стебельська, 2018).

Варто зазначити, що і самі корені та глибини творчого процесу донині залишаються науковцям невідомими. Традиційно вважається, що найбільш втягнутою у творчий процес є права півкуля мозку, яка пов'язана з образним мисленням. Проте ці прояви творчої діяльності є проявами вторинними, тому що в мозку ми спостерігаємо лише альфа- та бета-ритми, які не пояснюють якості творчості, її смислової компоненти й того, чому в нас виникають одні образи, а не інші; вони не пояснюють ірраціонального, дещо містичного характеру творчості, її несподіваних осяянь тощо; загадкою для них є і взаємодія творчого акту з емоційною сферою (Стебельська, 2015). Складається враження, що ми спостерігаємо лише певні зовнішні його прояви, в той час як суть лишається повністю закритою. Якби ми могли пояснити механізми функціонування творчості та уяви за допомогою нейропсихології, алгоритмічних операцій чи певних психологічних законів, ми давно були би творчими та геніальними без винятку. Проте творчий процес, попри всі наукові та психологічні дослідження, досі залишається загадкою. Ми чітко усвідомлюємо незавершеність наших уявлень про цей феномен. А значить, знову з гостротою та безапеляційністю постає питання: як відтворити те, що досі не розкриває нам своїх таємниць? Необхідно визнати, що всі дослідження у сфері штучного інтелекту мають цей недолік, а відтак фрагментарні та хаотичні за своєю суттю;

8) люди часто діють і роблять висновки, спираючись на здоровий глузд. Твердження здорового глузду складаються з певних знань, бажань, прагнень, на основі яких ми можемо робити подальші висновки та прогнозувати розвиток ситуації. Вони можуть бути помилковими, але їх можливо відкорегувати. Вони сприймаються як те, що саме собою зрозуміле, проте для інтелектуальної машини вони не володіють очевидністю, та їх доведення та обґрунтування потребуватиме складних операцій.

Таким чином, питання створення ШІ залишається відкритим. Ми не можемо однозначно стверджувати, що цього ніколи не станеться, але певні принципові труднощі в розв'язанні цієї задачі є. У теоремі Пенроуза дана проблема прозвучала гостро та безапеляційно: «Якою би потужністю не володів пристрій, що має архітектуру кінечного автомата (комп'ютера, в сучасному розумінні), людське мислення має деякі можливості, недоступні такому пристрою» (Панов, 2013). Варто зазначити, що основні проблеми реалізації штучних моральних машин лежать не так у технічній, як у філософсько-методологічній площині.

Незважаючи на труднощі реалізації інтелектуальних машин, вже зараз є помітним їх вплив на всі сфери нашого життя. У політичній та соціальній структурах виникає небезпека маніпуляції нашою свідомістю, нелегального збору та використання інформації. На даний час Інтернет-ресурси «знають» про нас більше, ніж ми самі. Серйозно обговорюється проблема втрати приватності. Ми живемо в суспільстві тотального контролю, де кожен наш клік в Інтернеті фіксується, а кожна фотографія, допис чи повідомлення зберігаються на віддалених серверах. Навіть останні дослідження, пов'язані зі здатністю нейросіток переводити мозкові сигнали у членороздільне мовлення, фіксують ситуацію втрати внутрішньої приватності, коли в майбутньому ми будемо мати змогу читати думки інших людей. Дослідження у сфері штучного інтелекту здатні активно впливати і на наше влас-

не становлення. Інтернет-речі, розумні будинки, наднова побутова техніка, роботи-помічники впливають двояко на формування та становлення нас як людей. З одного боку, всі ці блага цивілізації дійсно роблять наше життя комфортнішим та зручнішим, звільняють нас від рутинної роботи, дають змогу витратити час на саморозвиток. Вони свідчать про потужний розвиток науки, в процесі якого людина реалізує закладений у ній потенціал. З іншого боку, те, що мало би свідчити про наше становлення та зростання, працює на зниження нас як людей. Адже тепер немає необхідності боротись із певними труднощами, перемагати та дисциплінувати себе, долати певні негаразди. Раніше для того, щоб приділити час собі, ми повинні були виконати ряд рутинних, не завжди приємних робіт; вільний час поцінувався і витрачався з користю. Зараз спостерігається тенденція, що чим більше вільного часу ми маємо, тим більше ми його марнуємо. Р. Курцвейль говорить про те, що в майбутньому ми матимемо змогу вживляти людині різноманітні чіпи або ж ще до народження людини тасувати гени і вибирати, ким вона стане. З одного боку, це шокує майбутнє відкриває перед кожним із нас величезні можливості. З іншого, виникає проблема вільного вибору власного майбутнього. Від нас вже нічого не буде залежати, адже наші батьки самі виберуть, якими рисами, особливостями та потенціалом ми володітимемо. За умови абсолютного планування людині не прийдеється докладати практично ніяких зусиль для того, щоб стати тим, ким вона хоче стати. Вона припинить боротись за саму себе, все буде даватись їй в готовому вигляді. Тому і виникає небезпека зниження значущості самого людського існування.

Безумовно, виникає питання, а чи не буде момент сингулярності одночасно моментом кінця людського виду? На це питання надзвичайно важко відповісти, оскільки воно носить спекулятивний характер. Філософи розмірковують над тими наслідками, до яких може призвести створення потужного штучного інтелекту. Одні займають стриману позицію, вважаючи, що на даному етапі рано говорити про створення ШІ такої складності. Більше того, не варто боятись ШІ, оскільки він може бути надійним помічником для нас у повсякденному житті. Р. Доукінз взагалі припускає, що якщо ШІ захопить планету, життя на ній може стати кращим. Якщо хтось і складає небезпеку, так це сама людина. Інші, навпаки, вважають, що штучний інтелект може поглинути людство, що призведе до кінця людського виду. Дискусія триває досі. У будь-якому разі, наше майбутнє в наших руках. Дослідження у сфері ШІ будуть продовжуватись, і те, в якому світі ми житимемо, залежить від нас. Важливо зрозуміти, що навіть за умов невдачі у створенні ШІ вищезгадані розвідки можуть бути корисними для розуміння того, ким є ми. Можливо, ШІ – це дзеркало, в яке ми дивимось. І в такій ситуації як ніколи актуальною стає потреба критичного та самостійного мислення, здатного виявити всі потенційні можливості та загрози для нас.

Висновки. Починаючи із середини ХХ ст. відбувається потужний прорив у сфері створення інтелектуальних машин. Спочатку дослідження носили суто математичний характер і були пов'язані з доведенням різноманітних теорем. Проте вже тоді було помітно, що в науковців амбітніші плани; вони прагнули створити інтелект, який був би подібний людському. Перші досягнення були настільки шокуючими, що А. Т'юрінг висловив думку про те, що через 40 років (приблизно 2000 р.) машина зможе пройти тест Т'юрінга (ТТ) і її дії складно буде відрізнити від людських. Сучасні футурологи теж постійно прогнозують вражаюче майбутнє, але досі науковцям так і не вдалось створити машину, рівноцінну за своїми можливостями людині. Це пов'язано з багатьма причинами, розглянутими вище. Але основні проблеми лежать в філософсько-методологічній площині. У першу чергу, незрозуміло, *як створити розумну машину?* Окрім того, дослідження у сфері штучного інтелекту володіють одним спільним недоліком: науковці пробують змоделювати та прописати

те, природу чого самі однозначно і чітко не можуть пояснити: свідомість, мораль, уява, творчість, емоції та почуття, свобода тощо. Людська свідомість постає настільки складним, неоднозначним, багатоплановим феноменом, що не піддається на даний час технологізації. Навіть сам термін «інтелект» не має чіткого визначення. Проте вищезгадані перешкоди у створенні інтелектуальних машин обертаються до нас позитивним боком: чим більше ми намагаємось створити ШІ, тим більше ми поглиблюємо розуміння власної природи.

Список використаної літератури

1. Витгенштейн Л. Философские исследования. *Философские работы*. Москва : «Гнозис», 1994. С. 165–410.
2. Деннет Д. Виды психики: на пути к пониманию сознания. Москва : Идея-Пресс, 2004. С. 184.
3. Докинз Р. Эгоистичный ген. Москва : АСТ:CORPUS, 2013. С. 512.
4. Панов А. Технологическая сингулярность, теорема Пенроуза об искусственном интеллекте и квантовая природа сознания. *Метафизика*. 2013. № 3(9). С. 141–188.
5. Стебельська О. (2018). Свідомість. Штучний інтелект. Уява. *Гуманізм. Людина. Уява: Матеріали Міжнародних людинознавчих філософських читань, 2018*. Дрогобич. С. 3–9.
6. Стебельська О.І. Позиції менталізму та еволюціонізму у вивченні свідомості в історії сучасної філософії : дисертація канд. філос. наук : 09.00.05 ; Львів. нац. ун-т ім. Івана Франка. Львів. 2015. 208 с.
7. Суворов О. Интеллект. 2018. URL : <https://iphlib.ru/greenstone3/library/collection/newphilenc/document/HASH523b26975f21f2447bc452>.
8. Хель И. Очередная победа Deep Mind: после шахмат и го искусственный интеллект покорил StarCraft. 2019. URL : <https://hi-news.ru/robots/ocherednaya-pobeda-deep-mind-posle-shahmat-i-go-iskusstvennyj-intellekt-pokoril-starcraft.html>.
9. Черчленд П., Черчленд П. Искусственный интеллект: может ли машина мыслить. 1990. URL : <http://www.raai.org/library/books/sirl/ai.htm>.
10. Юлина Н. Головоломки проблемы сознания: концепция Дэниела Деннета. Москва : Канон+, 2004. 544 с.
11. Boden M. The Creative Mind: Myths and Mechanisms. Retrieved. 2004. URL : <https://pdfs.semanticscholar.org/52f1/53075b22469fa82ecb35099b8810e95c31f6.pdf>.
12. Bostrom N. Ethical Issues in Advanced Artificial Intelligence. 2003. URL : <https://nickbostrom.com/ethics/ai.html>.
13. Boyles M. A Case for Machine Ethics in Modeling. Human Level Intelligent Agents. 2018. URL : https://www.kritike.org/journal/issue_22/boyles_june2018.pdf.
14. Bringsjord S. Artificial Intelligence (Stanford Encyclopedia of Philosophy). 2018. URL : <https://plato.stanford.edu/entries/artificial-intelligence/#HistAI>.
15. Chalmers D. The Conscious Mind: In Search of a fundamental theory. New York : Oxford University Press, 1996. 415 p.
16. Chalmers D. (2010). The Singularity: A Philosophical Analysis. 2010. URL : <http://consc.net/papers/singularity.pdf>.
17. Dreyfus H. What Computers Can't Do. New York : MIT Press, 1972.
18. Kurzweil R. The age of spiritual machines. When computers exceed human intelligence. 1999. URL : https://docs.google.com/file/d/0BwjX_dbOIwbSNjlkYmZkNjEtYjU0NS00NjYwLWE2ZmEtZmI4ODQ2ZjQ1MDhl/edit.
19. McCarthy J. The philosophy of AI and the AI of philosophy. 2006. URL : <http://jmc.stanford.edu/articles/aiphil2/aiphil2.pdf>.
20. Moor J. The Nature, Importance, and Difficulty of Machine Ethics. Intelligent Systems, IEEE. 2006. 21(4). P. 18–21.

21. Muehlhauser L., Helm L. Intelligence Explosion and Machine Ethics. 2012. URL : <https://intelligence.org/files/IE-ME.pdf>.
22. Nagel T. What Is It Like to Be a Bat? *The Philosophical Review*. 1974. 83(4). P. 435–450.
23. Рассел С., Норвиг П. (2006). Искусственный интеллект: современный подход. Москва : Издательский дом «Вильямс», 2006. 1408 с.
24. Penrose R. *Shadows of the Mind*. Oxford University Press. 1994.
25. Searle J. *The rediscovery of the mind*. London. MIT Press Ltd, 1992. 270 p.
26. Searle J. (1990). Is the Brain a Digital Computer? 1990. URL : <https://philosophy.as.uky.edu/sites/default/files/Is%20the%20Brain%20a%20Digital%20Computer%20-%20John%20R.%20Searle.pdf>.

MODERN PHILOSOPHY OF CONSCIOUSNESS ABOUT THE PROSPECTS OF CREATING THE ARTIFICIAL INTELLIGENCE

Oleksandra Stebelska

*Lviv Polytechnic National University, Philosophy Department
Bandera str. 12, 79013, Lviv, Ukraine*

Liubomyr Fedoriv

*Lviv Polytechnic National University, Philosophy Department
Bandera str. 12, 79013, Lviv, Ukraine*

In this article the prospects for creating artificial intelligence are analyzed. The authors focus on the difficulties faced by researchers, trying to realize the intellectual, creative abilities of the person, his/her ethical and aesthetic principles, language and ordinary behavior. On the basis of a critical assessment it is argued that the intellectual capabilities of a person possess such unique properties that currently do not allow to reproduce in the machine environment, something similar in complexity and flexibility of functioning. However, this should not stop the scientists from attempts, because their aspirations lead not only to progress in the AI, but also allow them better understanding of their own nature.

Key words: consciousness, artificial intelligence, singularity, freedom, creativity.