

УДК 004:140.8

DOI: 10.26565/2226-0994-2018-58-3

Віталій Карпенко

СОВРЕМЕННЫЕ ПОДХОДЫ К СОЗДАНИЮ САМООПРЕДЕЛЯЮЩЕГОСЯ МАШИННОГО ИНТЕЛЛЕКТА: ФИЛОСОФСКИЙ АСПЕКТ

В статье эксплицируется философский аспект основных подходов к созданию самоопределяющегося машинного интеллекта. Принцип предосторожности рассматривается как адекватный методологический принцип для этой сферы исследований. Согласно этому принципу научное сообщество и вообще общество должны идти друг другу навстречу в смысле как познания опасений общества относительно неопределенности перспектив научно-технического роста, так и популяризации научных знаний, преодоления мифов, связанных с наукой и популярных среди неспециалистов. Делается вывод о том, что обществу следует создавать резервы для ответа на допустимые, а не только уверенно прогнозируемые угрозы.

Ключевые слова: самоопределяющийся машинный интеллект, искусственный интеллект, принцип предосторожности, дегуманизация.

Актуальным синтетическим направлением компьютеризации, привлекающим менее всего внимания научного сообщества в любом научном аспекте, являются усилия в направлении создания самоопределяющегося машинного интеллекта (искусственного интеллекта). Однако, как мы покажем далее, вопрос самоопределяющегося машинного интеллекта (обычно лишь подразумеваемый как часть – не важно истинной или ложной – информации, которой оперирует человечество) заслуживает отдельного рассмотрения.

Проблемное поле машинного интеллекта предстает в современной философии как чрезвычайно комплексное и многомерное. Одним из самых дискуссионных вопросов в этом контексте является создание и гипотетическое будущее самоопределяющегося машинного интеллекта, то есть машинного интеллекта как основы функций сознания (метафора) в их целостности. Именно в такой интерпретации раскрывал понятие технологической сингулярности профессор математики В. С. Виндж. Этот вопрос обсуждается и представителями частных наук, и в университетских философских кругах (Н. М. Амосов, А. Е. Баранович, И. В. Бестужев-Лада, Д. Э. Гаспарян, К. И. Карпенко, А. Н. Колмогоров, А. Я. Мороз, П. Норвиг, К. А. Павлов, Р. Пенроуз, С. Рассел, М. Ратнер, Д. Ратнер, А. С. Степаненко, В. Толстых и другие).

Для дальнейшего рассмотрения необходимо припомнить трансгуманистические идеи, критикуемые нами [см.: Карпенко, 2017, с. 95–98] за деантропологизацию, и подчеркнуть, что они всё же проявляют заботу об определенном благе, хотя и в своем специфическом понимании. Имеем в виду как благо для человека (например, медицинское продление жизни, устойчивое медицинское улучшение настроения и памяти), так и для тех (трансляудей, постлюдей), кто рассматривается в ряду прямой преемственности по отношению к человеку.

Разумеется, учитывая игнорирование побочных эффектов многих из предложенных «благ» и навязываемое идеологическое размежевание людей и постлюдей, то есть проведение новой линии разлома (контекстуализируя термин С. Ф. Хантингтона), это обобщенное «благо» получает окрас деструктивности. Однако соотношение трансгуманизма с антропоцентризмом как познавательной установкой и гуманизмом как типом философского мировоззрения можно сравнить с соотношением биоцентризма с антропоцентризмом и гуманизмом: биоцентристы отрицают антропоцентризм как устаревший, но не отстраняются от гуманизма, используют последнее понятие в широком смысле, что делает возможным выход за рамки, наложенные антропоцентризмом, а также

© Карпенко В. Е., 2018.



This is an open-access article distributed under the terms of the Creative Commons Attribution License 4.0.

провозглашение Нового (обновленного) гуманизма (см., напр.: А. Печчен), в основу которого положен не антропоцентризм, а благо человека в биоцентрическом понимании. Подобным образом трансгуманисты, отрицая антропоцентризм, выступают за «гуманное» отношение к транслядам, постлюдям как потомкам *homo sapiens*.

Тем не менее в ходе взаимообогащения понятий, обозначающих различные направления компьютеризации и трансгуманизма, возникает нечто качественно новое: понятие самоопределяющегося машинного интеллекта (под разными названиями) в его современном виде. В его экспликациях не просто обосновывается возможность моделирования функций сознания (метафора) в их целостности, но делаются гораздо более смелые прогнозы в этом контексте: происходит введение в науку (как дискурсивно равноправного) сочетания идей создания самоопределяющегося машинного интеллекта и допустимости вытеснения людей его носителями, не рассматривая последнее как угрозу, которая требует поиска путей противодействия.

Временно абстрагируясь от вопроса возможности практической реализации таких прогнозов, подчеркнем то, что уже само их включение в философский дискурс является не просто отрицанием антропоцентризма и заменой его на машиноинтеллектцентризм, но еще и полным отрицанием гуманизма с его отстаиванием самоценности человека (пусть даже трансчеловека, постчеловека), его достоинства. По нашему мнению, философия должна замечать и давать адекватную оценку подобным процессам еще на ранних стадиях.

Вопрос возможности создания самоопределяющегося машинного интеллекта оказывается не чисто техническим уже с позиций соотношения ответов на него с философским учением о сознании. Сквозь призму проблемы дегуманизации этот вопрос получает широкий мировоззренческий резонанс и находит отклик в философии науки и техники, социальной философии и пр. В целом проблематика нашей статьи, вовлекающая разнообразное философское и специально-научное знание, обоснованно предполагает философский уровень обобщения. Соответственно, целью статьи является экспликация философского аспекта основных подходов к созданию самоопределяющегося машинного интеллекта.

Подход, в рамках которого утверждается, что «... человек является действительно сложной материальной системой, но системой конечной сложности и весьма ограниченного совершенства и поэтому доступной имитации» [Колмогоров, 2004, с. 311], можно считать достаточно влиятельным в современном философско-теоретическом дискурсе. Оптимизм у исследователей касательно создания самоопределяющегося машинного интеллекта как основы функций сознания (метафора) в их целостности вызывает длинный ряд существующих наработок. Например, как прототип свободы воли человека, многие десятилетия насчитывает идея компьютера со случайным элементом. Такие машины вовлекают электронные процессы, подобные в этом смысле игре в кости.

К созданию более совершенных образцов машинного интеллекта всё больше привлекают достижения нано- и биотехнологий. Уже созданы ДНК-компьютеры, способные решать математические задачи с большей скоростью, чем компьютеры на традиционной основе, речь идет также о так называемом «квантовом компьютере».

Открытой является проблема конкретного носителя самоопределяющегося машинного интеллекта [Мороз, 2008, с. 105]. Это может быть как компьютер в близком к современному пониманию (или же совокупность объединенных в систему компьютеров), так и культивированная мозговая ткань (последнее порождает определенные этические проблемы). Насколько второй вариант может быть в форме всё еще машинного интеллекта, или уже киборга? В воображении исследователей стираются границы между живым и неживым (методологическая четкость которых и так была предметом дискуссий).

Тем не менее даже в пределах материалистической парадигмы исследователям следует учитывать целый ряд существенных затруднений. В частности, Р. Пенроуз утверждает, что явление сознания (а наличие сознания – хотя и в метафорическом

смысле – является необходимым условием самоопределяющегося машинного интеллекта) не может быть адекватно описано в рамках современной физической теории. Более того, он аргументировано доказывает: оно не может быть объяснено с позиций современного научного мировоззрения в целом. Однако исследователь отмечает, что он далек от отстаивания позиции, согласно которой понимание сознания невозможно в рамках научного подхода. Просто, по его мнению, современная наука еще не достигла уровня, необходимого для решения такой задачи [Пенроуз, 2005, с. 10].

Однако К. А. Павлов пытается вывести проблему моделирования интеллекта на качественно новый уровень уже в наше время, утверждая среди прочего, что «... потенциальные богатства виртуального компьютерного мира могут быть превращены в идеальную симуляцию самосознающего и самоконтролирующегося существа, если взять на вооружение иную логику организации внутреннего мира этих электронных существ...» [Павлов, 2005, с. 80–81], чем та, которая сейчас царит в компьютерной технике. Фактически речь идет о смещении акцента с логики тождества (аристотелева логика), которая доминирует в естественных науках, на логику аналогий, которая лежит в основе гуманитарных наук. В частности, мыслитель указывает на более высокие выразительные возможности «неточных» наук.

Здесь мы, по сути, выходим на необходимость фундаментальных изменений в системе методов, парадигматических ориентаций указанной области. Но насколько современная наука способна к таким трансформациям? Проблема синтеза методологий естественных и гуманитарных наук (а в сфере машинного интеллекта без такого синтеза, понимаем, не обойтись) является едва ли не наиболее дискуссионной. Пока что такой подход порождает больше вопросов, чем ответов, и ожидает дальнейших исследований. В случае их успеха на повестке дня оказываются фундаментальные изменения не только в области машинного интеллекта, но и во всей общенаучной картине мира.

В то же время современные концептуализации самоопределяющегося машинного интеллекта являются ярким примером зависимости технологических форм от субъекта созидания, а не только от объективных законов природы. Подчеркнем наличие прямых и обратных связей между человеческой субъективностью и технаукой. Исследователи во многом пытаются воссоздать именно интеллект *homo sapiens*. В частности, для современного этапа характерно «... изучение и моделирование рациональных структур в связи с эмоциями, верованиями, чувствами, практическими навыками и неаналитическими методами обработки образной информации...» [Петрунин, 2010, с. 159].

Но насколько конструктивным является стремление создать так называемый «некоторый полный аналог интеллекта человека» [Швирков, 2006, с. 151–152]? Можно предположить, что функции интеллекта человека будут воспроизведены во многом на основе других принципов действия. В этом контексте приведем лишь один яркий пример: первые попытки создания прототипов современных летательных аппаратов заключались в неудачном воспроизведении механизма полета птиц.

Некоторые исследователи даже рассматривают возможность отойти от метафоричности, интерпретируя самоопределяющийся машинный интеллект, что отражается в соотношении понятий «сильного» и «слабого» машинного интеллекта. Так, С. Рассел и П. Норвиг, авторы фундаментальной работы «Искусственный интеллект: современный подход», формулируют упомянутую проблему следующим образом [см.: Рассел, Норвиг, 2006, с. 1248]. Утверждение, согласно которому машины, возможно, способны действовать так, как будто действительно являются интеллектуальными, называют гипотезой слабого искусственного интеллекта (машинного интеллекта. – В. К.). Утверждение, что машины действительно мыслят (а не просто имитируют мыслительные процессы), называют гипотезой сильного искусственного интеллекта (машинного интеллекта. – В. К.).

То есть, как мы видим, если следовать гипотезе сильного машинного интеллекта, принципиальной разницы между текущими достижениями в области машинного

интеллекта и самоопределяющимся машинным интеллектом они не усматривают. И тот, и другой якобы обладают мышлением в буквальном смысле, просто первый технически недостаточно совершенен, чтобы проявить свою способность мыслить сполна. Однако упомянутые ученые также констатируют: большинство исследователей искусственного интеллекта (машинного интеллекта) принимают гипотезу слабого машинного интеллекта как данность и не задумываются над тем, что может дискутироваться также гипотеза сильного машинного интеллекта; коль скоро разработанная ими программа успешно функционирует, их не волнует, назовут ли ее работу имитацией интеллекта или настоящим интеллектом.

С другой стороны, такие исследователи, как, например, А. С. Степаненко, придерживаются мнения о принципиальной невозможности конструирования самоопределяющегося машинного интеллекта [Степаненко, 2007, с. 17; см. также: Гаспарян, 2017, с. 92]. В частности, А. С. Степаненко исходит из тезиса о невычислимости сознания: речь идет о неудачах сциентистско-рационалистического редукционизма в гносеологической интерпретации интеллекта человека как субъекта социально-коммуникативной деятельности. Сформулируем в этом контексте вопрос: можно ли воплотить качественное своеобразие человеческого интеллекта на цифровой количественной основе (путем цифровых количественных программных преобразований)? При этом рост вычислительной мощности компьютеров, объемов памяти и т. п. тоже будут обобщенно пониматься здесь как количественные изменения.

Однако в контексте потенциальных невиданных высот научно-технического прогресса А. П. Назаретян приводит следующие концептуальные положения постнеклассической науки. Всякое знание признается культурно производным и исторически ограниченным, а абсолютные «истины» уступают место взаимодополнительным «моделям». Соответственно, любой окончательный запрет на технические решения, будучи порождением определенной модели, оказывается таким же относительным, как сама модель. Носитель информационной модели, творчески «играя» ее компонентами, способен формировать метамодели, в рамках которых неуправляемые константы превращаются в управляемые переменные. «Таким образом, технические задачи, принципиально неразрешимые в одной модели, оказываются решенными в другой модели, и, коль скоро мы перестали догматически возводить функциональные модели в ранг “объективных истин”, не видно иных оснований считать возможности целенаправленного управления потенциально исчерпаемыми» [Назаретян, 2001, с. 212–213].

Тем не менее нельзя переоценивать эти положения для конкретных случаев. Попробуем подробно пояснить последний тезис. Предположим, моделирование функций сознания (метафора) в их целостности с техническими средствами как носителем следует истолковывать в качестве общей цели комплекса исследований ряда ученых. Но вот с формулировкой общей проблемы соответствующей отрасли не всё так просто. Упомянутое моделирование на современном этапе нельзя истолковать и как общую проблему в виде «каким способом (способами) можно смоделировать функции сознания (метафора) в их целостности с техническими средствами как носителем», поскольку неизвестно, можно ли вообще смоделировать упомянутые функции сознания (метафора) в их целостности с техническими средствами как носителем (припомним приведенный тезис о невычислимости сознания у А. С. Степаненко). При этом, по всей видимости, утвердительный теоретический ответ на данный вопрос, предлагаемый рядом исследователей в рамках постнеклассической науки, лишь в очередной раз порождает контраргументы оппонентов (в обратном случае происходит ровно наоборот). То есть проблема выражается в последнем вопросе (можно ли вообще смоделировать), а не в предыдущем (каким способом / способами; ведь если вообще невозможно, то «какими способами» – неважно). И лишь достижение (либо недостижение) общей цели оказывается удовлетворительным критерием адекватности теоретических построений.

Практика в этой области исследований, выходящих на крайне высокий уровень абстракции, оказывается основным критерием истины.

С учетом наших замечаний о постановке проблемы, достижение такой сверхсложной (исходя даже из современных неполных знаний о сознании) цели требует решения целого ряда последовательных задач, каждая из которых будет иметь в своей основе конкретную исследовательскую проблему, о большинстве из которых в настоящее время мы даже не догадываемся (вспомним приведенные выше воззрения Р. Пенроуза). То есть «метамодель» (А. П. Назаретян), которая «обещает» достижение цели (удачное моделирование), должна последовательно порождать соответствующие исследовательские проблемы. Но на определенном этапе, так и не доведя нас до «логического завершения», она может оказаться одной из тех самых парадигм, которые, по Т. С. Куну, прекратили порождать исследовательские проблемы вообще и вместо этого превратились в средства инженерных дисциплин (например, геометрическая оптика) [Кун, 1977, с. 113]. Что касается возможностей конструирования других метамodelей, то приведенный пример Т. С. Куна о геометрической оптике позволяет сделать вывод, что в некоторых случаях подобный путь оказывается непродуктивным – то есть реформативное геометрической оптики хоть с позиций «моделей», хоть с позиций «метамodelей» так и не смогло восстановить порождение исследовательских проблем в геометрической оптике.

Таким образом, в рамках уже полученных наукой результатов можно сделать *вывод*, что сейчас идея самоопределяющегося машинного интеллекта содержит больше от концептуализации суператтрактора, в стремлении к которому могут совершенствоваться современные системы машинного интеллекта, методология области и даже общенаучная методология, чем от конкретной цели исследований, концептуализации аттрактора. Однако следует отметить, что отрицание важности реагирования на эту угрозу по причине ее гипотетичности может предоставить прецедент для нереагирования (по аналогии) на любую другую несколько «менее гипотетическую». Причем в контексте беспрецедентного научно-технологического роста в различных сферах неизвестно, где здесь провести качественную границу. В культуре утвердился образец (модель) игнорирования. Ситуация с феноменом самоопределяющегося машинного интеллекта может рассматриваться как пробный камень, в определенном смысле даже типичный пример отношения ученых и общества в целом к возможности внезапных научных прорывов в самых неожиданных областях. Как деструктивные крайности здесь выступают дегуманизация и игнорирование. М. Хайдеггер отмечал: «Чем ближе мы подходим к опасности, тем ярче начинают светиться пути к спасительному, тем более вопрошающими мы становимся» [Хайдеггер, 1993, с. 238]. И такую «близость» не следует толковать буквально, мы как раз и призваны оказаться «ближе» в ходе философской рефлексии.

Адекватным методологическим принципом для этой сферы нам представляется принцип предосторожности. «Важность принципа определяется характером подлежащих предупреждению потенциальных угроз. В действительности, их знание можно квалифицировать и как “знание”, и как “незнание”. <...> Он [принцип предосторожности. – В. К.] привносит в науку новый фактор, связанный с вмешательством общества в сферу научной деятельности» [К обществу знания, 2005, с. 144–145]. Согласно этому принципу научное сообщество и общество в целом должны идти друг другу навстречу: как в форме познания опасностей общества относительно неопределенности перспектив научно-технического роста, так и путем популяризации научных знаний, преодоления мифов, связанных с наукой и распространенных среди неспециалистов. Следует создавать резервы для ответа на допустимые, а не только уверенно прогнозируемые угрозы. Эти резервы создаются и в ходе дальнейших научных исследований, и в ходе гуманизации науки, общества, гуманизации ментальности как глубинного, устойчивого фактора духовной сферы. Последние являются резервами готовности общества адекватно реагировать на кризисные ситуации.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- Гаспарян Д. Э. Таинство естественной семантики: трансцендентальное измерение смысла и проблема искусственного интеллекта. *Вопросы философии*. 2017. № 4. С. 81–94.
- К обществам знания. Всемирный доклад ЮНЕСКО / предс. совета Всемирного доклада Ф. Ривьер. Париж: Юнеско, 2005. 240 с.
- Карпенко В. Е. Человек и культура в условиях техноинтеллектуализации антропосферы. Х.: ХНУ имени В. Н. Каразина, 2017. 328 с.
- Колмогоров А. Автоматы и жизнь. *Информационное общество: сб.* М.: АСТ, 2004. С. 285–312.
- Кун Т. Структура научных революций / пер. с англ. И. З. Налетова. М.: Прогресс, 1977. 301 с.
- Мороз А. Я. Исторические судьбы Homo sapiens в контексте развития искусственного интеллекта, эволюции сингулярных технологий. *Наука и образование: современные трансформации*. К.: Парапан, 2008. С. 89–112.
- Назаретян А. П. Цивилизационные кризисы в контексте Универсальной истории: Синергетика, психология и футурология. М.: ПЕР СЭ, 2001. 239 с.
- Павлов К. А. Существует ли неискусственный интеллект? *Вопросы философии*. 2005. № 4. С. 76–85.
- Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики: монография [Электронный ресурс] / пер. с англ. под общ. ред. В.О. Малышенко. М.: УРСС, 2005. 453 с. URL: http://hotmix.narod.ru/books_rus/anomal/newmind.html.
- Петрунин Ю. Ю. Искусственный интеллект. *Новая философская энциклопедия*. М., 2010. Т. 2. С. 159–160.
- Рассел С., Норвиг П. Искусственный интеллект: современный подход / пер. с англ. К. А. Птицына. М.: Вильямс, 2006. 1408 с.
- Степаненко А. С. Социокультурные и технологические предпосылки искусственного интеллекта: дис. ... доктора филос. наук: 09.00.08. Ростов-на-Дону, 2007. 203 с.
- Хайдеггер М. Время и бытие: Статьи и выступления / пер. с нем. В. В. Бибихина. М.: Республика, 1993. 447 с. (Мыслители XX в.)
- Швирков О. І. Проблема штучного інтелекту і людиновимірність штучних інтелектуальних систем: дис. ... канд. філос. наук: 09.00.09. Житомир, 2006. 174 с.

Карпенко Виталий Евгеньевич

кандидат философских наук, доцент кафедры философии
Национальный технический университет «Харьковский политехнический институт»
ул. Кирпичева, 2, Харьков, 61002
E-mail: mailvyek@gmail.com
ORCID: <https://orcid.org/0000-0001-9327-4147>

Статья поступила в редакцию: 24.04.2018

Утверждена к печати: 14.06.2018

СУЧАСНІ ПІДХОДИ ДО СТВОРЕННЯ МАШИННОГО ІНТЕЛЕКТУ, ЩО САМОВИЗНАЧАЄТЬСЯ: ФІЛОСОФСЬКИЙ АСПЕКТ

Карпенко Віталій Євгенович

кандидат філософських наук, доцент кафедри філософії
Національний технічний університет «Харківський політехнічний інститут»
вул. Кирпичова, 2, Харків, 61002
E-mail: mailvyek@gmail.com
ORCID: <https://orcid.org/0000-0001-9327-4147>

Проблема осмислення машинного інтелекту постає для сучасної філософії як комплексне та багатовимірне завдання. Одним з найбільш дискусійних питань в цьому контексті є створення і гіпотетичне майбутнє машинного інтелекту, що самовизначається. Особливої актуальності у цьому зв'язку набувають також різні трансгуманістичні ідеї. Ці ідеї висловлюють турботу про певне благо, хоча й у своєму специфічному розумінні. Мається на увазі благо як для сучасної людини, так і для тих, кого розглядають у якості інших форм людини (транслюди, постлюди тощо). В ході взаємозбагачення понять, що позначають різні напрямки комп'ютеризації і трансгуманізму, виникає дещо якісно нове: поняття машинного інтелекту, що самовизначається, в його сучасному вигляді. Однак повний опис і адекватне розуміння такого типу інтелекту є надскладним завданням. В цьому випадку потрібно створення загальної «метамодель», яка «обіцяє» досягнення мети (тобто вдале моделювання). Така «метамодель» повинна послідовно породжувати відповідні дослідницькі проблеми для більш локальних досліджень. Але на певному етапі вона може виявитися однією з тих самих «парадигм», що, як писав Т. Кун, припинили породжувати дослідницькі проблеми й замість цього перетворилися на засоби інженерних дисциплін. Отже, потрібно відзначити, що заперечення важливості реагування на можливі загрози на рівні загальної «метамодель» може надати прецедент для нереагування і на будь-яку іншу «менш гіпотетичну» небезпеку. Адекватним методологічним принципом для вирішення цієї проблеми, на нашу думку, є принцип обережності. Згідно з цим принципом наукове співтовариство і суспільство загалом повинні йти назустріч одне одному. Мова йде як про пізнання реального підґрунтя побоювань суспільства щодо невизначеності перспектив науково-технічного зростання, так і про популяризацію наукових знань, подолання міфів, пов'язаних з наукою та популярних серед неспеціалістів, тощо. Слід створювати резерви для відповідей на ймовірні, а не тільки на впевнено прогнозовані загрози. Ці резерви мають створюватися і в ході подальших наукових досліджень, і в ході гуманізації науки та суспільства, в ході гуманізації людської ментальності як глибинного, стійкого чинника духовної сфери. Отже, саме ці процеси забезпечують вже зараз і забезпечуватимуть у майбутньому готовність суспільства адекватно реагувати на кризові ситуації.

Ключові слова: машинний інтелект, що самовизначається, штучний інтелект, принцип перестороги, дегуманізація.

Стаття надійшла до редакції: 24.04.2018

Схвалено до друку: 14.06.2018

MODERN APPROACHES TO THE CREATION OF SELF-DETERMINING MACHINE INTELLIGENCE: THE PHILOSOPHICAL ASPECT

Karpenko Vitaliy Ye.

PhD in Philosophy, Associate Professor of the Department of Philosophy

National Technical University "Kharkiv Polytechnic Institute"

2, Kyrpychova str., 61002, Kharkiv, Ukraine

E-mail: mailvyek@gmail.com

ORCID: <https://orcid.org/0000-0001-9327-4147>

ABSTRACT

The article explores the philosophical aspect of the basic approaches to the creation of self-determining machine intelligence. The problem of understanding machine intelligence appears for modern philosophy as a complex and multidimensional task. One of the most controversial issues in this context is hypothetical future of self-determining machine intelligence. Very topical at the same time are various transhumanist ideas. These ideas express concern for a certain good, although in their specific understanding. It refers to the benefit for both the modern man and those who are considered as other forms of human being (transhumans, posthumans, etc.). In the course of mutual enrichment of concepts that denote different directions of computerization and transhumanism, something qualitatively new arises: the concept of self-determining machine intelligence in its modern form. However, a complete description and an adequate understanding of this type of intelligence is a super complicated task. In this case, it is required to create a common "metamodel", which "promises" the achievement of the goal (ie, successful modeling). Such a "metamodel" should consistently generate relevant research problems for more specific studies. But at a certain stage it can turn out to be one of those "paradigms" that,

as T. S. Kuhn wrote, stopped generating research problems and instead turned into the “tools” of engineering disciplines. Therefore, it should be noted that denying the importance of responding to possible threats at the level of a common “metamodel” can provide a precedent for non-response for any other “less hypothetical” danger. An adequate methodological principle for solving this problem is the precautionary principle. According to this principle, the scientific community and society must move towards one common goal. This means that it is necessary to eliminate fears of society about the dangerous prospects of scientific and technological growth, popularize scientific knowledge, overcome myths that non-specialists associate with science. It is necessary to create reserves to answer the permissible, and not only confidently predicted threats. These reserves are created not only in the course of scientific research, but also through the humanization of science and society, the humanization of the human mentality itself. Only on this basis the ability of society to respond adequately to crisis situations can be made.

Keywords: self-determining machine intelligence, artificial intelligence, precautionary principle, dehumanization.

REFERENCES

- Gasparian, D. E. (2017). The mystery of natural semantics: the transcendental dimension of meaning and the problem of artificial intelligence. *Voprosy Filosofii – Questions of Philosophy*, 4, 81–94. (In Russian).
- Heidegger, M. (1993). *Time and being: articles and speeches*. (V. V. Bibikhin, Trans.). Moscow: Republic. (In Russian).
- Karpenko, V. Ye. (2017). *Man and culture in the conditions of techno-intellectualization of the anthroposphere*. Kharkiv: V. N. Karazin Kharkiv National University, 2017. (In Russian).
- Kolmogorov, A. (2004). Automata and life. In A. Laktionov (Ed.), *Information Society* (pp. 285–312). Moscow: AST. (In Russian).
- Kuhn, T. S. (1977). *The structure of scientific revolutions*. (I. Z. Naletova, Trans.). Moscow: Progress. (Original work published 1962). (In Russian).
- Moroz, A. Ya. (2008). Historical destinies of Homo sapiens in the context of the development of artificial intelligence, the evolution of singular technologies. In *Science and education: modern transformations* (pp. 89–112). Kyiv: Parapan. (In Russian).
- Nazaretyan, A. P. (2011). *Civilization crises in the context of universal history. Synergetics, psychology and futurology*. Moscow: PER SE. (In Russian).
- Pavlov, K. A. (2005). Does artificial intelligence exist? *Voprosy Filosofii – Questions of Philosophy*, 4, 76–85. (In Russian).
- Penrose, R. (2005). *The Emperor’s new mind: concerning computers, minds, and the laws of physics*. (A. Dambis, Yu. Danilov, S. Kokarev, V. Malysenko, I. Olshevskij & L. Yakovenko, Trans., N. Malysenko, Ed.). Moscow: URSS. (Original work published 1989). Retrieved from http://hotmix.narod.ru/books_eng/anomal/newmind.html. (In Russian).
- Petrinin, Yu. Yu. (2010). Artificial Intelligence. In V. S. Styopin (Ed.), *New philosophical encyclopedia* (Vol. 2., pp. 159–160). Moscow: Mysl. (In Russian).
- Russell, S., & Norvig, P. (2006). *Artificial intelligence: a modern approach*. (A. Ptitsyn, Trans.). Moscow: Vilyams. (Original work published 1995). (In Russian).
- Stepanenko, A. S. (2007). *Socio-cultural and technological preconditions of artificial intelligence*. (Doctoral thesis). Rostov-on-Don: Southern Federal University. (In Russian).
- Shvirkov, O. I. (2006). *The problem of artificial intelligence and the human dimension of artificial intelligence systems*. (PhD thesis). Zhytomyr: Zhytomyr Ivan Franko State University. (In Ukrainian).
- UNESCO. (2005). *Towards knowledge societies: UNESCO World Report*. Paris: UNESCO.