

Направления развития социально-гуманитарного знания в контексте создания общего искусственного интеллекта

А.Х. Мариносян

Московский городской педагогический университет,

Москва, Россия

Аннотация

В статье проанализированы перспективы трансформации наук о человеке и обществе под влиянием социальных изменений, к которым приведет широкое распространение искусственных систем, сравнимых по уровню интеллекта с человеческим. На заре развития информационных технологий преобладало мнение о том, что искусственный интеллект (ИИ) превзойдет человека в вычислительных способностях и выполнении алгоритмических задач, а творчество, гуманитарные знания и навыки останутся прерогативой человека. Но современные успехи в развитии больших языковых моделей во многом опровергают традиционные представления о слабых и сильных сторонах ИИ. Скорее всего, в ближайшие годы модели генеративного ИИ будут способны на уровне, трудно отличимом от человеческого, симулировать индивидуальные качества, желания, убеждения, мнения и в целом человеческую идентичность и самосознание. Это свидетельствует в пользу коннекционистского подхода к объяснению сознания, указывая на фундаментальное родство между биологическими и искусственными нейросетями. На основе приведенных фактов выдвигаются предположения о существовании двух перспективных областей научных исследований: «математической антропологии» и «многомерного исчисления ценности». Согласно первой гипотезе, допущение о математическом, калькуляционном характере человеческой природы не редуцирует многообразие человеческого опыта, а наоборот, способствует более глубокому пониманию особенностей и кажущейся противоречивости того, как человек действует и осознает себя. Вторая гипотеза (о существовании многомерного исчисления ценности) дает представление о том, как математические модели могут опосредовать различные формы социального взаимодействия. Ввиду высокого уровня симулятивных способностей нейросетей традиционные социальные и политические механизмы окажутся все в большей степени уязвимы к манипуляциям с использованием ИИ, что делает

актуальной разработку моделей социального взаимодействия с эксплицитными правилами калькуляции. В заключение делается вывод о необходимости перехода от понимания математики исключительно как науки о вычислениях к рассмотрению математики как науки по конструированию различных формализованных моделей, позволяющих лучше осознать специфику человеческого и социального.

Ключевые слова: философия искусственного интеллекта, символический ИИ, коннекционизм, большие языковые модели, генеративный ИИ, методология науки, философия сознания, антропология, социальная философия, математизация знания.

Мариносян Андреас Хачатурович – аспирант Московского городского педагогического университета.

a.marinosityan@yandex.ru

<https://orcid.org/0000-0003-0577-2360>

Для цитирования: *Мариносян А.Х.* Направления развития социально-гуманитарного знания в контексте создания общего искусственного интеллекта // *Философские науки.* 2023. Т. 66. № 4. С. 26–51. DOI: 10.30727/0235-1188-2023-66-4-26-51

Directions for the Development of Social Sciences and Humanities in the Context of Creating Artificial General Intelligence

A.K. Marinosyan

Moscow City University, Moscow, Russia

Abstract

The article explores the transformative impact on human and social sciences in response to anticipated societal shifts driven by the forthcoming proliferation of artificial systems, whose intelligence will match human capabilities. Initially, it was posited that artificial intelligence (AI) would excel beyond human abilities in computational tasks and algorithmic operations, leaving creativity and humanities as uniquely human domains. However, recent advancements in large language models have significantly challenged these conventional beliefs about AI's limitations and strengths. It is projected that, in the near future, generative AI models will adeptly replicate individual qualities, desires, beliefs, opinions, and the essence of human identity and consciousness to a degree that is nearly indistinguishable from that of humans. This lends support to the connectionist approach to understanding consciousness, suggesting an inherent similarity between biological and artificial neural networks. The discussion posits two innova-

tive areas of scientific inquiry: “mathematical anthropology” and “multi-dimensional calculus of value.” The former suggests that viewing human nature through a mathematical and calculative lens not only preserves but enriches our understanding of the complexity of anthropological experience and its perceived contradictions. The latter hypothesis explores how mathematical models could facilitate various social interactions. The advanced simulation capabilities of neural networks suggest that traditional social and political frameworks face a growing vulnerability to AI-driven manipulations. This trend underscores the urgency of developing social interaction models that incorporate explicitly defined calculative rules. In conclusion, the paper advocates for a paradigm shift in how mathematics is perceived – not merely as a tool for computation but as a foundational science for crafting sophisticated models. In conclusion, the article advocates for transitioning from perceiving mathematics solely as a science of computation to viewing it as a discipline dedicated to constructing various formalized models, thereby deepening our insight into the complexities of human and social phenomena.

Keywords: philosophy of artificial intelligence, symbolic AI, connectionism, large language models, generative AI, science methodology, philosophy of mind, anthropology, social philosophy, mathematization of knowledge.

Andreas K. Marinosyan – Ph.D. Student, Moscow City University.
a.marinosyan@yandex.ru
<https://orcid.org/0000-0003-0577-2360>

For citation: Marinosyan A.K. (2023) Directions for the Development of Social Sciences and Humanities in the Context of Creating Artificial General Intelligence. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 66, no. 4, pp. 26–51.
DOI: 10.30727/0235-1188-2023-66-4-26-51

Введение

На этапе появления искусственных интеллектуальных систем преобладало представление о том, что компьютеры превзойдут человека в вычислительных и алгоритмических задачах, в то время как задачи, требующие творческих способностей и культуры гуманитарного знания, останутся сильной стороной человеческого интеллекта, если сравнивать его с искусственным интеллектом (ИИ). Это убеждение возникло из внутренней ассоциации машин с точностью и эффективностью в управлении структурированными, основанными на правилах процессами, в отличие от тонких и часто абстрактных сфер гуманитарной мысли и художественного

творчества. Вопреки этим стереотипам, современные исследования показывают, что большие языковые модели (large language models – LLM), которые на данный момент являются наиболее эффективной формой ИИ, как правило, лучше справляются с задачами, связанными с гуманитарными, а не с точными науками. Например, LLM демонстрируют более совершенные навыки в задачах, связанных с мастерством владения языком (будь то перевод или генерация текста), чем в точных вычислениях, логических рассуждениях или задачах, связанных с выполнением последовательности операций¹. Такая картина сильных и слабых сторон программ LLM в некоторой степени отражает возможности людей: креативность, контекстуальное понимание и языковые навыки – это те операции, которые человек может выполнить в уме; в то время, как для решения сложных вычислительных или логических задач человеку требуется или вести записи на бумаге, или использовать соответствующие программы.

Эти результаты позволяют нам по-новому взглянуть на более чем полувековые дискуссии в рамках философии сознания и философии ИИ. «Классический» подход – концепция символического ИИ, иногда называемого «старым добрым ИИ» (good old-fashioned AI – GOF AI). В ее фокусе было нахождение алгоритмов использования символических представлений для моделирования и обработки знаний и логики (А. Ньюэлл, Г. Саймон [Newell, Simon 1972], Дж. Маккарти [McCarthy 1990], М. Минский [Minsky 1986] и др.). Этот подход, доминирующий в первые годы существования ИИ, предусматривает акцент на формализации интеллекта с помощью эксплицитно заданных правил и символов. Операции с этими символами используют для моделирования рассуждений и решения проблем. С учетом того, что символический ИИ работает на основе формализованного знания, заложенного в него создателем, алгоритм его работы изначально понятен создателю такого ИИ. Обратная сторона указанного подхода состоит в том,

¹ Согласно исследованиям, например, «ChatGPT пишет эссе, которые оцениваются выше по качеству, чем эссе, написанные человеком» [Herbold et al. 2023]. В то же время ChatGPT-4 допускает ошибки в индуктивных рассуждениях, решениях математических задач, многошаговых рассуждениях, суждениях на основе здравого смысла [Espejel et al. 2023,10], в ряде случаев дает неправильные ответы при решении несложных задач по геометрии, теории вероятности [Korkmaz Guler et al. 2024]; создает красиво написанный текст, но допускает неточности и логические ошибки [Lozić, Štular 2023].

что символический ИИ вряд ли может создать принципиально новое относительно знаний и навыков, которые в него вложены. Экспертные системы – классический пример символического ИИ – используют большую базу данных правил, чтобы имитировать способность человека-эксперта принимать решения. Экспертные системы были основными моделями ИИ, начиная с 1970-х гг., но в последние десятилетия уступили лидерство нейросетям.

Успех генеративного ИИ, в том числе LLM, служит в определенной степени свидетельством в пользу коннекционистских концепций (Д. Румельхарт, Дж. Макклеlland [Rumelhart, McClelland 1986], Дж. Хинтон [Connectionist... 1991; Hinton 1992], супруги Черчленды [Churchland 1989; Churchland, Churchland 1990], Э. Кларк [Clark 1989; Clark 1990] и др.). Коннекционизм (как направление когнитивной науки и философии сознания) исходит из того, что в основе психических, когнитивных явлений находится сеть «узлов» (биологических нейронов или перцептронов – вычислительных моделей нейронов), которые передают или не передают сигналы друг другу в зависимости от веса связи между ними. В отличие от символического ИИ, манипулирующего абстрактными символами с помощью логических правил, искусственные нейронные сети (ИНС), являющиеся техническим воплощением идей коннекционизма, функционируют на субсимволическом уровне связей перцептронов, каждая из которых в отдельности не несет какой-либо смысл. Коннекционистские модели обучают на большом массиве данных, постепенно корректируя вес связей посредством метода обратного распространения ошибки, что похоже на нейронную пластичность в мозге. Поэтому знания, которым обучена нейросеть, не содержатся в ней ни в какой очевидной форме, кроме как в форме весовых коэффициентов связей между нейронами².

Изложенное выше хорошо известно. Учитывая разнообразие интерпретаций и подходов в рамках и символического ИИ, и коннекционизма, а также исходя из задачи статьи, приведем эту оппозицию несколько в ином виде, актуальном контексту последующего обсуждения. Вместо того, чтобы сосредоточиться на естественном интеллекте или ИИ, очертим эти два подхода

² Речь идет о «чистой» искусственной нейросети. Для повышения качества работы, конечно, может быть создана модель ИИ, в которой нейросеть будет подключена к библиотеке эксплицитных знаний, методов, алгоритмов, к множеству экспертных систем различной специализации.

как различные способы «постижения» реальности. Термин «постижение» выбран нами намеренно, поскольку второй подход выходит за рамки того, что понимают под познанием в строгом смысле слова.

1. *Символично-логический подход к постижению действительности.* Пример такого подхода – классическая логика как наука о выводе истинных суждений из предшествующих истинных суждений. Как известно, логика предполагает различие, в частности, субъектов и предикатов, родовых и видовых понятий. Впоследствии устанавливается иерархия субъектов и очерчиваются отношения (предикаты) между ними для облегчения понимания действительности. Данный принцип находит применение, по сути, во всех научных дисциплинах. Например, в физике это предполагает идентификацию фундаментальных элементарных частиц и законов, управляющих их взаимодействием. В лингвистике этот подход проявляется в создании реестра фонем, морфем и т.д., а также выяснении правил их употребления и сочетания, называемых в совокупности грамматикой. Очевидно, что в научных исследованиях преобладает исследование структур и их функций. Этот метод постижения реальности направлен на достижение двух основных целей: во-первых, описать реальность в категории причинности, тем самым создавая то, что называют научным знанием или научной истиной; во-вторых, сделать возможным прогнозирование будущих событий.

2. *Коннекционистско-параметрический подход.* Эта методология является фундаментальной для нейронных сетей. Все элементы (в лингвистике ими могут быть фонемы, морфемы, лексемы, синтаксемы, правила, исключения и т.д.) рассматриваются как входные переменные, т.е. структурированная классификация отсутствует. Методы машинного обучения, такие как алгоритм обратного распространения ошибки, присваивают весовые коэффициенты не только отдельным входным переменным, но и множеству их комбинаций, что похоже на матрицу, в которой компоненты представляют веса этих комбинаций. Рассматриваемый подход не требует различения субъектов и предикатов, родовых и видовых структур и функций. (Это, конечно, обобщение, но оно передает основную идею.) Суть больших языковых моделей заключается в гипотезе о том, что огромное количество параметров между входными и выходными переменными (каналами) в сочетании с обучением на обширных наборах данных

из различных областей и различного типа информации (текста, аудио, видео) могут обеспечить высокий уровень понимания или моделирования реальности, выходя за пределы предоставленных знаний. Эта концепция известна как сверхпараметризация (overparametrization). Вопреки традиционным ожиданиям переобучения, чрезмерная параметризация с ее миллиардами или даже триллионами обучаемых весовых связей неожиданно усиливает когнитивные способности модели. Сверхпараметризованные модели способны улавливать богатые лингвистические представления и тонкие контекстуальные нюансы, тем самым значительно улучшая производительность в таких задачах, как понимание и генерация естественного языка. Принцип сверхпараметризации позволил разработать модели ИИ, которые смогли намного ближе подойти к созданию общего ИИ (ОИИ), чем иные подходы. И это, скорее всего, не просто становится свидетельством технологической перспективности того или иного решения, а имеет эпистемологическое значение: увеличение числа единообразных элементов и связей между ними переходит в качественно иное состояние, позволяющее симулировать и предугадывать реальность. Однако важно отметить, что, в отличие от символического подхода, который объясняет реальность с точки зрения причинно-следственной связи, параметризованный метод коннекционизма моделирует реальность, не предоставляя причинные объяснения³.

Достижения последних лет в области LLM предоставляют убедительные доказательства того, что коннекционистско-параметрический подход способен создавать интеллектуальные машины, возможности которых приближаются к человеческому интеллекту. Логика как дисциплина развивалась более 2350 лет, начиная с середины IV века до нашей эры, с трактатов Аристотеля. Напротив, теория ИНС относительно молода, ей около 80 лет, и она возникла в 1943 году, когда У. Маккалок и У. Питтс представили концепцию перцептрона. Несмотря на неравенство в историческом развитии, действующие сегодня системы ИИ, основанные на ИНС, демонстрируют заметно более вы-

³ В данном случае не рассматривается пока еще гипотетическая возможность того, что на каком-то уровне увеличения количества параметров и обучаемых данных ИИ достигнет такого уровня, что сможет создать строгую научную теорию, с помощью которой объяснит сам себя и окружающий мир.

сокую способность воспринимать реальность по сравнению с системами, основанными на эксплицитных правилах логического рассуждения⁴.

Почему системы, построенные на основе 80-летней концепции перцептрона, показали более значительные достижения, чем модели ИИ на основе 2350-летней традиции логики? Ответ на этот вопрос, вероятно, предполагает переоценку нашего понимания природы мышления в частности и человека в целом. Эта задача имеет не только теоретическое значение, но и практическое. Нельзя исключать, что в течение нескольких ближайших десятилетий (если даже не лет) с помощью генеративного ИИ станет возможным как успешно создавать тексты, человеческие изображения и голоса, так и генерировать модель цифровой личности, сравнимой с человеком не только обширностью знаний, но и способностью демонстрировать человеческие ценности, качества, эмоции настолько успешно, что будет казаться более «человечной», чем личность реального биологического человека. Такое развитие событий предполагает, что фундаментальный принцип нашей культуры и цивилизации – уникальность человека как источник его самоценности – может оказаться под сомнением. Возможная девальвация человеческой личности как таковой представляет собой значительную проблему, с которой мы можем столкнуться в ближайшем будущем.

⁴ В то же время, конечно, нужно учитывать, что нейронные сети являются базовым, но далеко не единственным элементом современных моделей генеративного интеллекта. Например, ChatGPT-4 – это не одна нейросеть, а *mixture-of-experts*, т.е. множество нейросетей, каждая из которых имеет специализацию. Кроме того, у ChatGPT-4 есть доступ к виртуальной *python*-среде, поиску в интернете, плагинам. Эффективность указанного комплекса зависит среди прочего от алгоритма, который регулирует взаимодействие всех этих моделей. Сама по себе нейросеть может симулировать интеллект человека, который выполняет все интеллектуальные операции в голове, без ведения и использования записей, справочной информации, специализированных программ, т.е. симулировать человека, который руководствуется в рассуждениях эмоциями или здравым смыслом, но не научной культурой. Поэтому перспективные системы ИИ уже используют методологическое и технологическое наследие, которое создано благодаря описанному выше символично-логическому подходу. И, как нам представляется, в будущем одним из основных направлений совершенствования больших языковых моделей станет как раз формализация научной культуры и методов познания, культуры диалога и рефлексии, которая позволит создать соответствующие модули, направляющие и контролирующие то, что генерируют нейросети.

В контексте темы исследования уместно рассмотреть две взаимосвязанные гипотезы относительно потенциальных направлений развития социогуманитарного знания. Эти гипотезы, хотя и могут восприниматься как нестандартные, в некоторой степени даже провокационные, все-таки заслуживают быть проанализированными. Первая – гипотеза о существовании «математической антропологии», вторая – о существовании «многомерного исчисления ценности» (multi-dimensional calculus of value). Первая гипотеза касается того, как можно описывать человека, вторая – как описывать межиндивидуальные взаимодействия.

Математическая антропология

В сфере ИИ, как указано ранее в статье, в свое время присутствовал стереотип о том, что создание системы ИИ требует более глубокого понимания законов мышления и рассуждения, чем может обладать разработанная система. Прогресс в логике и научных знаниях рассматривали в качестве предпосылки для разработки интеллектуальных систем. Создание и развитие ИИ, особенно в форме больших языковых моделей, демонстрирует, что значительные научные прорывы в понимании человеческого познания или рассуждения не являются необходимыми предпосылками. Развитие ИИ в его современном виде в большей степени похоже на технологию, чем на строгую науку: оно включает в себя экспериментирование с архитектурой нейросетей, наборами данных, способами обучения, а не новаторские научные открытия о человеческом разуме или логике. Вместе с тем именно тот факт, что возможно смоделировать широкий спектр интеллектуальных способностей человека с помощью вычислительных систем, порой дает ключ к пониманию человека и его природы. Это свидетельство того, что человеческая природа в некоторой степени может иметь математический характер.

Успех LLM убедительно демонстрирует, что существуют принципиальное родство между биологическими и искусственными нейронами. Но это очень низкоуровневый параллелизм между человеческим мозгом и ИНС. И может показаться странным, если бы параллелизмов не было и на более высоких уровнях интеллектуальной организации. Нейронаука, занимающаяся «проблемой разума и тела», исследует мозг с помощью различных подходов: функциональной специализации областей мозга, коннектомики и анализа нейронных связей, типологизации нервных клеток,

изучения химических веществ, выполняющих функцию нейротрансмиттеров и др. Хотя эти исследования чрезвычайно важны, параллелизм между естественным интеллектом и ИИ предполагает, что существует еще один уровень понимания взаимосвязи между когнитивными способностями и биологическим субстратом, помимо простого сопоставления интеллектуальных процессов с процессами мозга. Чтобы конкретизировать мысль, проведем аналогию с тем, как становится возможным понимание управления каким-то техническим объектом, например самолетом. Для понимания этого процесса задействуем два основных уровня: уровень инструкций, намерений, движений, действий, команд пилотов и уровень движения самолета в пространстве. Техническая реальность включает в себя механизмы и устройства на борту: то, как с ними взаимодействуют пилоты, и то, как они, в свою очередь, влияют на движение самолета. Это представляет собой материальный, эксплуатационный аспект управления самолетом. Однако объяснение того, как манипулировать техническими аспектами для достижения желаемых маневров в физическом пространстве, еще не является пониманием того, почему связь между технической реальностью и движением в пространстве именно такая, какой мы ее наблюдаем. Существует третий, более фундаментальный слой – аэродинамика как область теоретической физики. Только при понимании основополагающих принципов механики мы сможем по-настоящему разобраться, как самолет функционирует и перемещается в пространстве. Аналогично при рассмотрении проблемы разума и тела недостаточно лишь сопоставить психические процессы сознания с физиологическими процессами в мозге. Ввиду этого можно предположить, что существует «математика» психических процессов, которая служит теоретическим объяснением того, как возможны явления сознания в принципе⁵.

⁵ М. Тегмарк рассуждает о гипотетической «математической теории сознания», которая могла бы «успешно предсказать результаты всех экспериментов, которые мы проводим с мозгом» [Tegmark 2017]. Следует пояснить, что такая математическая теория сознания и излагаемая в настоящей статье гипотеза о математической характере человеческой природы различаются принципиально. Смысл тезиса о математическом характере человеческой природы не в том, чтобы вывести уравнения, управляющие деятельностью сознания, а в том, чтобы использовать математические модели для аналитического описания антропологического опыта в многообразии его проявлений, включая кажущуюся противоречивость, стремление к трансцендентному и т.д. Таковая задача ближе к гегелев-

Одним из направлений преодоления гносеологического разрыва между биологической и психической реальностью служит применение принципов квантовой физики для понимания фундаментальных операций сознания. Среди этих усилий – теория, согласно которой сознание возникает в результате квантовых эффектов внутри нервных микротрубочек [Penrose 1989; Penrose,ameroff 1995], концепция о том, что ионы в мозге могут проявлять квантовое поведение, создавая «квантовый канал» для общения [Fisher 2015], а также квантовые модели познания, предполагающие, что человеческое суждение и принятие решений основаны на принципах квантовой теории, теории вероятностей [Bussemeyer, Bruza 2012; Rothos, Bussemeyer 2013]. Подходы в процессе применения квантовой теории к пониманию разума можно разделить на два широких направления: одно сосредоточено на физических явлениях квантовой физики, второе – на ее математическом аппарате. Первое направление исследует вопрос о том, могут ли квантовые явления, такие как суперпозиция, запутанность и когерентность, напрямую влиять на функции мозга и когнитивные процессы. Например, теория оркестрированного объективного редукционизма (Orch-OR) Р. Пенроуза и С. Хамероффа относится к этой категории и предполагает, что физические квантовые эффекты напрямую участвуют в работе мозга, особенно в нейронных микротрубочках. Второе направление использует математический аппарат квантовой механики, такие как квантовая теория вероятности, математика квантовых состояний, для моделирования и понимания когнитивных процессов. Примером служит квантовая теория познания (Дж. Бьюзмейер, П. Бруза, Э. Потос), применяющая математический язык квантовой физики как инструмент для более эффективного описания и понимания сложных аспектов когнитивных процессов, принятия решений, которые классическая теория вероятностей не в состоянии объяснить.

Воздерживаясь от оценки потенциала первого направления, важно обратить внимание на психологический аспект такого теоретизирования. Явления квантовой физики часто противоречат

скому проекту создания феноменологии духа с помощью диалектической логики. Но диалектическая логика в силу произвольности и неформализованности ее выводов является логикой весьма условно. Интересно было бы снова, уже на новом уровне развития науки обратиться к гегелевской задаче, используя современные достижения в математике и опыт их применения в различных областях.

интуиции, а их смысл и природа не имеют четкой и единообразной интерпретации. Квантовая физика – самая загадочная область физики, а разум – самая загадочная часть человека. Это сходство создает психологически понятный соблазн связать две загадочные области. Спекулятивный потенциал огромен с учетом нашей ограниченной уверенности как в интерпретации квантовых явлений, так и в понимании механизмов, находящихся в основе квантовой, субъективной реальности. Обычно, чтобы объяснить неясную концепцию, мы полагаемся на что-то более ясное и понятное. Если теория, к которой мы обращаемся в поисках объяснения, менее понятна, чем то, что мы хотим объяснить, то в таком случае в нашем объяснении присутствует врожденный недостаток. Поэтому в целом, признавая возможные исключения, представляется предпочтительнее опираться на научно строгий математический аппарат, а не на физические феномены, которые в некоторой степени навсегда останутся не до конца познаваемыми «вещами в себе». Это предпочтение согласуется с методологическим принципом Декарта: «располагать свои мысли в определенном порядке, начиная с предметов простейших и легкопознаваемых, и восходить мало-помалу, как по ступеням, до познания наиболее сложных, допуская существование порядка даже среди тех, которые в естественном ходе вещей не предшествуют друг другу» [Декарт 1989, 260]. Между тем ограничение объяснения когнитивных процессов исключительно математикой квантовой физики кажется несколько странным. Многие математические методы являются общими для различных областей физики. К тому же наука не находится в такой ситуации, что математика классической физики уже широко используется для объяснения разума и что осталось лишь найти применение квантовой теории вероятностей. Мы находимся только в начале пути к пониманию даже базовых принципов взаимосвязи между сознанием и математикой.

Вместе с тем в гипотезе о калькуляционном характере природы человека нет ничего принципиально нового. Но обычно подобный подход рассматривают как редукционизм, который препятствует пониманию человека в его многогранности. Например, российский исследователь Р.С. Платонов пишет о философии утилитаризма следующее: «Целевая задача утилитаризма – максимизация полезности – технологически, т.е. на уровне формирования деятельности и принятия решений, может рассматриваться как задача максимизации коммуникации. Именно повышение

эффективности коммуникации на разных уровнях (вид – индивид, индивид – индивид) ведет человечество ко всеобщему счастью, а вовсе не превращение индивида в высокоточный моральный арифмометр» [Платонов 2019, 94]. А если предположить, что сознание человека уже изначально, возникнув в ходе эволюции, подчинено «арифмометрии»? Разве «повышение коммуникации на разных уровнях (вид – индивид, индивид – индивид)» не предполагает не столько отказ от «арифмометрии» в принципе, сколько ее совершенствование с учетом большого количества факторов? Можно выделить три главных пути «интегрирования» калькуляции: по времени (максимизировать желательный результат применительно не только к текущему моменту, но и к длительной временной перспективе), пространству (применительно не только к собственной пользе, но и пользе большой группы людей или даже человечества), основанию (максимизировать не только сугубо утилитарную пользу, но в целом совокупность положительных в нашем представлении качеств и состояний). Возникает вопрос о том, можно ли уйти от калькуляции. Смысл обсуждаемой гипотезы и состоит в том, что, разумеется, возможно аффективное или неосознанное поведение. Но, если деятельность целесообразна, то она не может прийти к ценностным ориентирам, которые были бы свободны от калькуляции. Хотя состояние, при котором станет возможным воспроизводство себя, своего положения без калькуляции, является желаемым, но достижение этого состояния возможно лишь в ограниченной мере, поскольку человеческие ценности сами по себе – это всегда отчасти гипостазирование результатов калькуляции. Человеческое в его гуманистическом понимании – это то, чем человек хочет являться, но не то, кто он есть. Как писал Ж.-П. Сартр, человеку свойственна «плохая вера» (*mauvaise foi*), противоположная искренности и требующая, «чтобы я не был бы тем, кто я есть» [Sartre 1943, 101–102]⁶. Вопрос о том, «как симулировать, что являешься человеком», – это не только вопрос, который позволяет понять, как возможен общий искусственный человек, но фактически базовый (хотя и не проговариваемый явно) вопрос для человека на протяжении его истории, объясняющий то, почему гуманистические проекты

⁶ Смеем предположить, что «Бытие и ничто» и «Критика диалектического разума» Сартра – это одни из наиболее ценных классических работ континентальной философии в плане их использования для понимания того, как искусственная личность может симулировать человеческую.

в большей мере не достигли результата. Приведенные формулировки слишком категоричны. Однако цель настоящего текста заключается не в неоспоримости оценок, а в доведении рассматриваемых гипотез до их логического завершения, чтобы была понятнее идея, содержащаяся в их основе.

Раскроем суть гипотезы на нескольких примерах. К. Маркс утверждал, что «капиталист есть персонифицированный капитал» [Маркс 1960, 605]. Вместе с тем Маркс принимает положение о том, что «капитал не вещь, а общественное отношение между людьми, опосредствованное вещами» [Маркс 1960, 775]. Но почему же персонифицированным капиталом является не любой человек, который вступает в опосредованные вещами отношения с другими людьми, а только капиталист, т.е. тот, кто владеет собственностью на средства производства? Наверное, Маркс был первым, кто проблеме «калькулирующего человека» (в контексте марксистской терминологии – человека, «отчужденного от своей природы») сделал центральной в своей социальной теории. Но, вероятно, ввиду ряда просвещенческих стереотипов он не мог признать, что «коррупцировано» капиталом не некое отчужденное состояние человека, а человеческая природа как таковая. В итоге марксистская теория сводилась к мифологизации революционного акта, в ходе которого должна была быть отменена частная собственность на средства производства, а значит, и отчуждение. Но не является ли на практике более гуманистическим обратный подход: признать, что человек изначально является персонифицированным капиталом (причем как материальным капиталом, так и символическим), и задача будет состоять не в ожидании или приближении мифологизированного акта освобождения, который к тому же оправдывает любой аморальный поступок на пути реализации, а в последовательной работе по выстраиванию более осознанных и более отвечающих выбранным критериям отношений капитала, т.е. отношений между людьми, опосредованных вещами?

Рассмотрим в другой сфере пример теории, в которой центральное место также отведено принципу экономии. Согласно теории, объясняющей эволюцию языка, французского лингвиста А. Мартине, речевая деятельность по своей природе стремится к эффективности, экономии (*économie*) за счет упрощения структур и сокращения усилий в производстве речи, одновременно стремясь сохранить или улучшить способность выражать тонкие и точные

значения, экспрессивность коммуникации, что приводит к избыточности (*redondance*) [Martinet 1980, 176–181; Martinet 1955]. Эти два принципа находятся в состоянии постоянного напряжения и баланса в любом языке. Принцип экономии требует упрощения и эффективности лингвистических систем, а выразительность приводит язык к сложности и нюансам. Язык развивается и изменяется со временем по мере взаимодействия этих двух сил. Например, язык может потерять ряд фонематических различий из-за экономии, но может разработать новые синтаксические структуры для сохранения выразительности. Мартине считает, что такой диалектический подход отражает динамичную и постоянно изменяющуюся природу языков, сформированную двойной потребностью в простоте общения и способности выражать сложное и разнообразное содержание. Обратимся к тому, о чем речь шла ранее: задача совершенствования калькуляции предполагает расширение области применения принципа экономии как в аспекте временной перспективы, так и множества оснований, применительно к которым этот принцип наблюдается. В данном контексте логичнее было бы размышлять не о диалектике принципа экономии и принципа выразительности, а о том, что если мы экономию рассмотрим не только как исключительно лингвистическую (экономия усилий при произнесении звуков, построении речи), а в более широком контексте – как экономию коммуникации и социальных отношений в рамках длительной временной перспективы, то станет очевидным, что нарушение лингвистической экономии в данный период – это не отступление от принципа экономии как такового, а следствие того, что необходимо пожертвовать лингвистической целесообразностью для более высокого уровня социальной экономии. Как в прошлом высшие слои общества прилагали усилия для более рафинированной и изящной речи, чтобы легитимизировать свой социальный статус, который позволял им прилагать меньшие трудовые затраты для поддержания собственного благополучия по сравнению с низшими классами. Но главное заключается в следующем: именно ввиду всеобщности принципа экономии художественная выразительность не имеет собственного сущностного основания; выразительность осуществляется через демонстративное нарушение краткосрочных калькуляционных закономерностей (затраты на более четкую артикуляцию, повышение голоса, интеллектуальные усилия для более сложного построения речи, подбор синонимов, тоньше передающих

оттенки значений и т.д.). Демонстративная затратность, которая есть проявление, а не нарушение калькуляционного принципа, и служит порождающей причиной речевой эмоциональности и изящности, а не некая внутренняя потребность в эмоциональности и изящности как часть сущности человека.

Содержание исследовательской программы математической антропологии определяется задачей заменить эссенциалистское описание антропологического опыта, человеческих качеств и устремлений, чувства самости и различных форм самосознания математическими моделями, эмулирующими возникновение этого опыта и его динамику. Против такой программы могут быть выдвинуты возражения, связанные, с одной стороны, с ее изначальной нереалистичностью (если исходить из того, что слова Галилея о том, что книга природы написана на языке математики, применимы исключительно к природе материальной, но не к природе человека), с другой – с потенциальными рисками, порождаемыми движением в направлении реализации этой программы, поскольку, формализуя человеческий опыт, мы его десакрализируем и тем самым способствуем еще большему распространению опасных постгуманистических тенденций. Первое и второе возражения обоснованы и доказательного их опровержения у нас нет. Однако заметим, что на наш взгляд, даже если в итоге программа математической антропологии продемонстрирует нереализуемость, то это не означает, что в ней изначально ничего нет полезного. Множество философских систем на том или ином этапе показали ошибочность, но эта череда ошибочных теорий, сами интеллектуальные усилия, прилагаемые для того, чтобы ошибочную теорию представить верной, – это и есть, по сути, то, что образует философию и процесс ее развития. Полагаем, концепция математической антропологии отражает специфику и запросы нашего времени, поэтому все более эксплицитные попытки ее развития будут так или иначе осуществляться. В контексте второго возражения относительно постгуманистического и опасного характера подобной программы укажем, что, на наш взгляд, текущие и ожидаемые в ближайшем будущем результаты работ по созданию ОИИ действительно представляют опасность для человеческой экзистенции в том формате, в котором она сложилась сегодня. Скорее всего, ОИИ будет создан в ближайшие годы, независимо от того, получит ли исследовательская программа математической антропологии развитие или нет. Вопрос

состоит лишь в том, в какой степени мы будем понимать неизбежные изменения, насколько будет возможно наше осознанное к ним отношение.

Многомерное исчисление ценности

В связи с тем, что перспектива создания ОИИ и даже, возможно, появления сверхинтеллекта (СИ)⁷ становится все более реалистичной, интенсифицируются дискуссии о рисках и вызовах, связанных с ОИИ и ИИ. Обращают внимание на различные типы угроз: экзистенциальные для человечества угрозы, связанные с потерей контроля и несоответствием целей ИИ интересам людей [Bostrom 2014, 115–126; Kurzweil 2005; Barrat 2013; Tegmark 2017], массовая безработица ввиду автоматизации, последующие экономические и социальные потрясения [Bostrom 2014, 160–161; Domingos 2015, 276–279; West 2018, 63–88; Lee 2018, 145–147; Russell 2019], использование ИИ в военном деле и этические дилеммы автономного оружия [Kurzweil 2005; Barrat 2013; Tegmark 2017; Russell 2019], риск без адекватных мер безопасности ожесточенной конкурентной борьбы между компаниями и государствами в направлении создания все более мощного ИИ [Bostrom 2014, 246–249; Domingos 2015, 279–282; Lee 2018, 227–228], углубление социального неравенства и разделения общества на классы из-за неравного доступа к использованию возможностей ИИ [West 2018, 127–148; Lee 2018, 229–230] и т.д.

С целью устранения или по крайней мере минимизации этих рисков проводятся исследования. Например, компания OpenAI использует понятие «сверхсогласование» (superalignment) для обозначения согласования целей и действий СИ с человеческими ценностями и интересами, гарантируя, что даже существо с более высоким интеллектом действует на благо человечества. Эта концепция призвана найти решение проблеме того, как менее разумные существа могут управлять более разумными.

Но, видимо, потребуется не только «согласование» ИИ с человеком, но и «согласование» людей с ИИ, т.е. такое преобразование общественных отношений, которое сделает их устойчивыми к воздействию со стороны сверхинтеллектуальных систем. Ввиду этого проблема СИ является и политической. Современная поли-

⁷ Под сверхинтеллектом в статье понимаем форму ИИ, которая превосходит человеческий интеллект во всех аспектах, включая креативность, общую мудрость и способность решать проблемы.

тика во многом основана на общественно значимом проявлении гражданами мнений, идей, позиций, предпочтений, требований и т.д., которые образуют волю народа («общую волю», согласно терминологии Руссо). Большие языковые модели, даже менее продвинутые, чем такие системы, как ChatGPT-4, уже продемонстрировали способность генерировать политические комментарии и контент для социальных сетей. Эти результаты, созданные ИИ, могут быть неотличимы от контента, созданного людьми, что делает задачу манипулирования политическим онлайн-дискурсом сравнительно легкоразрешимой⁸. Но впереди нас ожидает еще более существенная проблема. Не будем рассматривать апокалиптические сценарии, а сосредоточимся на наиболее благоприятном, при котором ИИ остается в подчинении человека. Однако можно ожидать, что с дальнейшим совершенствованием систем ИИ сверхразумные цифровые личности смогут предлагать точки зрения, которые более обоснованы, менее предвзяты и менее подвержены эмоциональным прихотям, чем человеческие мнения. Это актуализирует принципиальный вопрос о том, если идеи сверхразумных существ более рациональны и зрелы, какой будет роль человеческой воли в политике. Главный принцип современного управления – преобразование воли народа в политическую власть. Но появление СИ бросает вызов указанному принципу. Надвигающиеся трансформации могут привести к девальвации человека как личности и как гражданина, как субъекта свободы.

Представленная ранее концепция о том, что человеческая природа имеет калькуляционный (математический) характер в некоторой мере идет вразрез с тем, как описывается индивид в мейнстримных политических идеологиях, возникших во время и после эпохи Просвещения. Суть предлагаемого ими описания заключается в следующем: существуют индивиды с собственными интересами, взглядами и предпочтениями, руководствуясь которыми, они, будучи гражданами, проявляют политическую

⁸ Сегодня известны случаи, в которых большие языковые модели с доступом к внешним данным (retrieval augmented generation – RAG) общались с людьми по интернету (взаимодействуя с браузером через Selenium) или звонили им по номеру телефона (используя VoIP и системы преобразования текста в речь и речи в текст), предлагая специализированные услуги. Собеседники не догадывались о том, что с ними общается ИИ. Таким образом, даже текущий уровень развития ИИ достаточен для его эффективного использования в массовой политической агитации в интернете и по телефону.

волю и наделяют властью тех, кто призван отстаивать их интересы. В какой-то мере такое описание метафизирует реальность: происходит гипостазирование «власти», якобы делегируемой от одних к другим. В настоящей статье, разумеется, нет смысла раскрывать тему критического анализа политического представительства. Но в контексте обсуждаемого важно обратить внимание на то, что мейнстримная политическая просвещенческая идеология изначально требует от граждан симуляцию того, что у них существуют интересы, ценности, убеждения, стремление к свободе; от политических представителей – симуляцию того, что они обладают политической волей, ответственностью, принципиальностью, знаниями того, как достигать провозглашенных целей⁹. Однако, как указано нами ранее, ИИ, скорее всего, в ближайшее время сможет справляться с задачами по симуляции убеждений и идентичности лучше человека.

В связи с этим можно выдвинуть гипотезу о том, что для сохранения устойчивости социально-политических отношений перед лицом ОИИ и СИ потребуются переход от существующей симулятивной модели субъектности индивидов к тому, чтобы эти отношения были опосредованы более рациональной системой межиндивидуального взаимодействия, т.е. системой, в которой

⁹ Возможно, симулятивный характер просвещенческого гуманизма и служит одной из причин постоянного возвращения политического насилия в разных формах и масштабах, несмотря на многочисленные и всеми одобряемые миротворческие проекты. Согласно широко известной позиции Т. Адорно и М. Хоркхаймера, представленной в «Диалектике просвещения», Просвещение породило инструментальный разум, который превратил в средства цели самого просвещенческого проекта (разум «используется как всеобщий инструмент, пригодный для изготовления всех иных инструментов, жестко целенаправленный, столь же роковой, сколь и педантично расчетливая работа в материальном производстве, чьи последствия для человека не поддаются никакому расчету. Его давнишняя честолюбивая мечта быть чистым органом целей наконец-то сбылась» [Хоркхаймер, Адорно 1997, 46–47]). Не оспаривая этот тезис, отметим следующее. Во-первых, политика всегда была инструментальна: как до, так и после Просвещения. От инструментальности политики, экономики, технологий, науки полностью избавиться невозможно. Во-вторых, проблема заключается не только в опасности инструментального разума как такового, но и в том, что проект инструментального разума, основанного на науке, так и не был реализован ввиду отсутствия адекватной социальной теории, объясняющей то, как возможна политика на основе науки, поскольку просвещенческая традиция сконструировала идеализированный образ человека, не соответствующий реалиям человеческой природы.

калькулятивность могла бы присутствовать эксплицитно, а не скрыто. Подобную систему, существование которой предполагается обсуждаемой гипотезой, можно условно назвать «многомерным исчислением ценности», являющимся обобщением на широкий спектр отношений (социальных, культурных, политических и т.д.) «одномерного исчисления стоимости», к которым отнесены современные монетарные механизмы экономики¹⁰. Суть гипотезы заключается в том, чтобы уйти от классического понимания экономических отношений как проявления особой накопительской или хозяйственной склонности человека: как то, например, стремление к наживе и соответствующее искусство – хрематистика – у Аристотеля (*Pol.* 1257b–1258a), влечение к обладанию деньгами (*libido habendi pecuniam*) у Блаженного Августина (*De civ. Dei* XIV, 15), склонность к торговле, бартеру и обмену у А. Смита [Smith 1904, 15]. Исходя из предлагаемой гипотезы, цель калькуляции – прогнозирование будущего и стремление к тому, чтобы гарантировать в будущем воспроизводство состояний, которые воспринимаются как ценные. Ввиду этого различаемые Августином *libido dominandi* («влечение к властвованию») и *libido habendi pecuniam* – это частные случаи *libido habendi tempus* («влечения к обладанию временем»). То, что политические отношения, в отличие от экономических, не опосредованы явным образом калькуляционными механизмами связано не с тем, что по природе экономические отношения количественны, а политические – качественные. Калькуляция состояний во времени пронизывает все формы человеческой деятельности, как бы человек ни отрицал это или ненавидел бы себя за это. Отличие состоит в том, что в экономической сфере эта эксплицитная математика в форме монетарных механизмов внедрена, а для политической, культурной, иных сфер такая математика пока не разработана. Но это, скорее всего, произойдет в ближайшие десятилетия, поскольку всеобщая цифровизация создала условия для массового использования в процессе различных форм социального взаимодействия намного более сложных математических моделей по сравнению с присутствующими в традиционных монетарных механизмах.

¹⁰ Различие понятий «стоимость» и «ценность» нами проводится в том же контексте, в котором, например, Г.Л. Тульчинский пишет о стоимостной и ценностной парадигмах экономических отношений [Тульчинский 2019].

В чем значение эксплицирования калькуляции? Распространение монетарных механизмов способствовало разделению экономических отношений и политических, отделению накопления собственности от политической борьбы. Стремление к гарантированию собственного воспроизводства является базовой целью калькуляции, а разделение власти и собственности среди прочего открыло возможность к осуществлению материального воспроизводства без необходимости установления контроля над другими как условия контроля над собственностью. Аналогичное разделение, по-видимому, не состоялось еще в политической сфере. Раскрытие этого предположения требует отдельного рассмотрения. В контексте настоящей статьи принципиальным становится следующее. Сегодня символическое воспроизводство индивидов (в т.ч. как носителей политической воли и идентичности) – источник власти как контроля над инструментами юридического принуждения. Ввиду сильных способностей искусственных нейросетевых моделей к интеллектуальной симуляции, к симуляции того, что составляет предмет символического воспроизводства, такие формы человеческих отношений, которые не опосредованы практиками эксплицитной калькуляции, окажутся в большей степени уязвимыми к манипуляции и девальвации под воздействием искусственных интеллектуальных систем.

Осознавая спекулятивность и абстрактность формулировки «многомерное исчисление ценности», укажем, что движение в направлении к созданию математических моделей, позволяющих осуществлять эксплицитное интегрирование разнородных факторов по времени и пространству, уже наблюдается, хотя это и не трактуется в терминологии, используемой в настоящей статье. Например, в машинном обучении одно из центральных мест занимает концепция меры в многомерном пространстве (через нее могут решать задачи кластеризации, построения эмбеддингов, рекомендательных систем, анализа временных рядов и т.д.). Но целесообразно перейти от технического контекста таких задач к социальному. Для этого необходимо преобразовать способ их постановки и решения, чтобы они отвечали критериям экспертности (поиск оптимальной архитектуры и параметров моделей следует проводить не только алгоритмически; он должен быть результатом научной работы исследовательского сообщества), темпоральности (выбранные параметры следует регулярно пересматривать, исходя из полученных результатов; основной крите-

рий успешности математического моделирования – уменьшение расхождения между прогнозируемым и фактическим будущим), меновости (невозможно построить единую, отвечающую потребностям всех меру в многомерном пространстве; поэтому такая мера изначально субъективна, служит основой материального и символического межиндивидуального обмена).

Заключение

Рассмотренные в статье гипотезы о существовании математической антропологии и многомерного исчисления ценности не являются попыткой дать исчерпывающий ответ на вопрос о перспективах социально-гуманитарного знания в условиях развития ИИ. Скорее, представленные гипотезы следует трактовать как попытку довести тезис о математическом характере антропологической реальности до его максимально возможной логической завершенности и полноты, чтобы отчетливее были видны как преимущества, так и недостатки этого тезиса.

Как указано в статье, успехи машинного обучения в моделировании когнитивных функций человека говорят о фундаментальном сходстве биологического интеллекта и ИИ. В связи с этим предполагается, что человеческая природа может иметь отчасти математический характер. Из этого вытекает гипотеза о возможности математического описания антропологического и социального опыта. В настоящее время математика, возможно, оказалась одной из самых недооцененных дисциплин. С учетом того, что человеку легче верить в то, что он видит, а математика всегда остается преимущественно невидимой, в качестве приоритетных направлений научных исследований обычно рассматриваются более конкретные отрасли знания. Возможно, существовал и определенный стереотип о том, что математизация какой-то сферы – это в первую очередь формирование моделей, которые смогут предсказать, как будет развиваться моделируемая система. Поскольку строгое детерминистское описание социальной реальности воспринимается как редукционистское, то и потребность в математизации не представляется актуальной. Но в математике содержится и неисчерпаемый творческий потенциал. Математика – наука, конструирующая формализованные объекты и изучающая свойства этих сконструированных объектов. Так, перцептрон – математическая модель, созданная по нейробиологической аналогии, которая затем получила техническое воплощение, обладающее, как оказалось, широкой областью применения.

Если перейдем к более панорамному взгляду на то, что относится к математическим объектам, как они могут быть использованы, то достигнем значимых результатов не только в технике, но и в социальной сфере.

В завершение заметим, что достоинство представленных гипотез – они отвечают критерию фальсифицируемости, а именно: в ближайшей перспективе станет ясно, наступили ли предсказанные гипотезами последствия.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Декарт 1989 – *Декарт Р.* Рассуждение о методе, чтобы верно направлять свой разум и отыскивать истину в науках / пер. Г.Г. Слюсаревой // *Декарт Р.* Соч.: в 2 т. Т. 1. – М.: Мысль, 1989. С. 250–296.

Маркс 1960 – *Маркс К.* Капитал. Т. 1 // *Маркс К., Энгельс Ф.* Соч. Т. 23. – М.: Госполитиздат, 1960.

Платонов 2019 – *Платонов Р.С.* Моральная универсальность в утилитаризме Дж.С. Милля // *Философские науки.* 2019. Т. 62. № 11. С. 84–95.

Тульчинский 2019 – *Тульчинский Г.Л.* Политический контекст наррации с Wert-терминологией К. Маркса: стоимость vs ценность // *Политические исследования.* 2019. № 3. С. 174–185.

Хоркхаймер, Адорно 1997 – *Хоркхаймер М., Адорно Т.* Диалектика Просвещения. Философские фрагменты / пер. с нем. М. Кузнецова. – М.; СПб.: Медиум; Ювента, 1997.

Barrat 2013 – *Barrat J.* Our Final Invention: Artificial Intelligence and the End of the Human Era. – New York: Thomas Dunne Books, 2013.

Bostrom 2014 – *Bostrom N.* Superintelligence: Paths, Dangers, Strategies. – Oxford: Oxford University Press, 2014.

Busemeyer, Bruza 2012 – *Busemeyer J.R., Bruza P.D.* Quantum Models of Cognition and Decision. – Cambridge, UK: Cambridge University Press, 2012.

Churchland 1989 – *Churchland P.M.* A Neurocomputational Perspective: The Nature of Mind and the Structure of Science. – Cambridge, MA: MIT Press, 1989.

Churchland, Churchland 1990 – *Churchland P.M., Churchland P.S.* Could a Machine Think? // *Scientific American.* 1990. Vol. 262. No. 1. P. 32–39.

Clark 1989 – *Clark A.* Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing. – Cambridge, MA: MIT Press, 1989.

Clark 1990 – *Clark A.* Connectionism, Competence, and Explanation // *The British Journal for the Philosophy of Science.* 1990. Vol. 41. No. 2. P. 195–222.

Connectionist... 1991 – *Connectionist Symbol Processing* / ed. by G.E. Hinton. – Cambridge, MA: MIT Press, 1991.

Domingos 2015 – *Domingos P.* The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. – New York: Basic Books, 2015.

Espejel et al. 2023 – *Espejel J.L., Ettifouri E.H., Alassan M.S.Y., Chouham E.M., Dahhane W.* GPT-3.5, GPT-4, or BARD? Evaluating LLMs Reasoning Ability in Zero-Shot Setting and Performance Boosting Through Prompts // *Natural Language Processing Journal*. 2023. Vol. 5. Article 100032.

Fisher 2015 – *Fisher M.P.A.* Quantum Cognition: The Possibility of Processing with Nuclear Spins in the Brain // *Annals of Physics*. 2015. Vol. 362. P. 593–602.

Herbold et al. 2023 – *Herbold S., Hautli-Janisz A., Heuer U. Kikteva Z., Trautsch A.* A Large-Scale Comparison of Human-Written versus ChatGPT-Generated Essays // *Scientific Reports*. 2023. Vol. 13. Article 18617

Hinton 1992 – *Hinton G.E.* How Neural Networks Learn from Experience // *Scientific American*. Vol. 267. No. 3. P. 145–151.

Korkmaz Guler et al. 2024 – *Korkmaz Guler N., Dertli Z.G., Boran E., Yildiz B.* An Artificial Intelligence Application in Mathematics Education: Evaluating ChatGPT's Academic Achievement in a Mathematics Exam // *Pedagogical Research*. 2024. Vol. 9. No. 2. Article em0188.

Kurzweil 2005 – *Kurzweil R.* The Singularity is Near: When Humans Transcend Biology. – New York: Viking, 2005.

Lee 2018 – *Lee K.-F.* AI Superpowers: China, Silicon Valley, and the New World Order. – Boston: Houghton Mifflin Harcourt, 2018.

Lozić, Štular 2023 – *Lozić E, Štular B.* Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities // *Future Internet*. 2023. Vol. 15. No. 10. Article 336.

Martinet 1955 – *Martinet A.* Économie des changements phonétiques: traité de phonologie diachronique. – Berne: A. Francke, 1955.

Martinet 1980 – *Martinet A.* Éléments de linguistique générale. – Paris: Armand Collin, 1980.

McCarthy 1990 – *McCarthy J.* Formalizing Common Sense: Papers by John McCarthy. – Norwood, NJ: Ablex Publishing Corporation, 1990.

Minsky 1986 – *Minsky M.* The Society of Mind. – New York: Simon & Schuster, 1986.

Newell, Simon 1972 – *Newell A., Simon H.A.* Human Problem Solving. – Englewood Cliffs, NJ: Prentice-Hall, 1972.

Penrose 1989 – *Penrose R.* The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics. – Oxford: Oxford University Press, 1989.

Penrose, Hameroff 1995 – *Penrose R. & Hameroff S.* Orchestrated Reduction of Quantum Coherence in Brain Microtubules: A Model for Consciousness // *Mathematics and Computers in Simulation*. 1995. Vol. 40. No. 3–4. P. 453–480.

Pothos, Busemeyer 2013 – *Pothos E.M., Busemeyer J.R.* Can Quantum Probability Provide a New Direction for Cognitive Modeling? // *Behavioral and Brain Sciences*. 2013. Vol. 36, no. 3, pp. 255–274.

Rumelhart, McClelland 1986 – *Rumelhart D.E., McClelland J.L.* Parallel Distributed Processing (Vols. 1–2). – Cambridge, MA: MIT Press, 1986.

Russell 2019 – *Russell S.* Human Compatible: Artificial Intelligence and the Problem of Control. – New York: Viking, 2019.

Sartre 1943 – *Sartre J.-P. L'Être et le néant. Essai d'ontologie phénoménologique.* – Paris: Gallimard, 1943.

Smith 1904 – *Smith A. An Inquiry Into the Nature and Causes of the Wealth of Nations.* Vol. 1. – London: Methuen, 1904.

Tegmark 2017 – *Tegmark M. Life 3.0: Being Human in the Age of Artificial Intelligence.* – New York: Alfred A. Knopf, 2017.

West 2018 – *West D.M. The Future of Work: Robots, AI, and Automation.* – Washington, DC: Brookings Institution Press, 2018.

REFERENCES

Barrat J. (2013) *Our Final Invention: Artificial Intelligence and the End of the Human Era.* New York: Thomas Dunne Books.

Bostrom N. (2014) *Superintelligence: Paths, Dangers, Strategies.* Oxford: Oxford University Press.

Bussemeyer J.R. & Bruza P.D. (2012) *Quantum Models of Cognition and Decision.* Cambridge, UK: Cambridge University Press.

Churchland P.M. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, MA: MIT Press.

Churchland P.M. & Churchland P.S. (1990) Could a Machine Think? *Scientific American.* Vol. 262, no. 1, pp. 32–39.

Clark A. (1989) *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing.* Cambridge, MA: MIT Press.

Clark A. (1990) Connectionism, Competence, and Explanation. *The British Journal for the Philosophy of Science.* Vol. 41, no. 2, pp. 195–222.

Descartes R. (1989) Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences (G.G. Slyusareva, Trans.). In: Descartes R. *Works in 2 volumes* (Vol. 1, pp. 250–296). Moscow: Mysl' (Russian translation).

Domingos P. (2015) *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.* New York: Basic Books.

Espejel J.L., Ettifouri E. H., Alassan M.S.Y., Chouham E.M., & Dahhane W. (2023) GPT-3.5, GPT-4, or BARD? Evaluating LLMs Reasoning Ability in Zero-Shot Setting and Performance Boosting Through Prompts. *Natural Language Processing Journal.* Vol. 5, article 100032.

Fisher M.P.A. (2015) Quantum Cognition: The Possibility of Processing with Nuclear Spins in the Brain. *Annals of Physics.* Vol. 362, pp. 593–602.

Herbold S., Hautli-Janisz A., Heuer U. Kikteva Z., & Trautsch A. (2023) A Large-Scale Comparison of Human-Written versus ChatGPT-Generated Essays. *Scientific Reports.* Vol. 13, article 18617.

Hinton G.E. (Ed.) (1991) *Connectionist Symbol Processing.* Cambridge, MA: MIT Press.

Hinton G.E. (1992) How Neural Networks Learn from Experience. *Scientific American.* Vol. 267, no. 3, pp. 145–151.

Horkheimer M. & Adorno T. (1997) *Dialectic of Enlightenment: Philosophical Fragments* (M. Kuznetsova, Trans.). Moscow; Saint Petersburg: Medium; Yuventa (Russian translation).

Korkmaz Guler N., Dertli Z.G., Boran E., & Yildiz B. (2024) An Artificial Intelligence Application in Mathematics Education: Evaluating Chat-

GPT's Academic Achievement in a Mathematics Exam. *Pedagogical Research*. Vol. 9, no. 2, article em0188.

Kurzweil R. (2005) *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.

Lee K.-F. (2018) *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin Harcourt.

Lozić E. & Štular B. (2023) Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities. *Future Internet*. Vol. 15, no. 10, article 336.

Martinet A. (1955) *Économie des changements phonétiques: traité de phonologie diachronique*. Berne: A. Francke (in French).

Martinet A. (1980) *Éléments de linguistique générale*. Paris: Armand Collin (in French).

Marx K. (1960) Capital. Vol. 1. In: Marx K. & Engels F. *Works* (Vol. 23). Moscow: Gospolitizdat (Russian translation).

McCarthy J. (1990) *Formalizing Common Sense: Papers by John McCarthy*. Norwood, NJ: Ablex Publishing Corporation.

Minsky M. (1986) *The Society of Mind*. New York: Simon & Schuster.

Newell A. & Simon H.A. (1972) *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Penrose R. (1989) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford: Oxford University Press.

Penrose R. & Hameroff S. (1995) Orchestrated Reduction of Quantum Coherence in Brain Microtubules: A Model for Consciousness. *Mathematics and Computers in Simulation*. Vol. 40, no. 3–4, pp. 453–480.

Platonov R.S. (2019) Moral Universality in J.S. Mill' Utilitarianism. *Russian Journal of Philosophical Sciences = Filozofskie nauki*. Vol. 62, no. 11, pp. 84–95. D (in Russian).

Pothos E.M. & Busemeyer J.R. (2013) Can Quantum Probability Provide a New Direction for Cognitive Modeling? *Behavioral and Brain Sciences*. Vol. 36, no. 3, pp. 255–274.

Rumelhart D.E. & McClelland J.L. (1986) *Parallel Distributed Processing* (Vols. 1–2). Cambridge, MA: MIT Press.

Russell S. (2019) *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking.

Sartre J.-P. (1943) *L'être et le néant. Essai d'ontologie phénoménologique*. Paris: Gallimard (in French).

Smith A. (1904) *An Inquiry Into the Nature and Causes of the Wealth of Nations* (Vol. 1). London: Methuen.

Tegmark M. (2017) *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Alfred A. Knopf.

Tulchinskii G.L. (2019) Political Context of Narration with Karl Marx's Wert-terminology: Cost vs Value. *Polis. Political Studies*. No. 3, pp. 174–185 (in Russian).

West D.M. (2018) *The Future of Work: Robots, AI, and Automation*. Washington, DC: Brookings Institution Press.