

Искусственный интеллект: как углеродная жизнь создала кремниевую

А.Ю. Сторожук

*Институт философии и права Сибирского отделения РАН,
Новосибирск, Россия*

Аннотация

В статье рассматриваются основные этапы развития искусственного интеллекта (ИИ). Цель статьи – выявить влияние компьютеризации на общественную жизнь и мировоззрение в настоящее время и в будущем. Метод исторического анализа позволил выделить этапы развития нейросетей, провести сравнение прогнозов и реальных достижений в области моделирования ИИ. Охарактеризованы три главных периода: «первая весна», «вторая весна» и «зима» эры ИИ (период стагнации). Во время первых обнадеживающих разработок возникли трудности технического характера. На первом этапе разработчики пытались копировать природу, т.е. нейроны естественного мозга, что оказалось слишком сложным. Кроме того, первые алгоритмы обучения нейронных сетей сопровождалась принципиальными трудностями, такими как экспоненциальное затухание вектора ошибки при распространении во внутренние слои нейронной сети. Потребовались годы для нахождения новых алгоритмических решений, накопления данных для обучения нейронных сетей, совершения следующего технологического прорыва, известного как «вторая весна» ИИ. В процессе и по итогам анализа установлено, что прогнозы возможностей ИИ часто были более скромными, чем реальные достижения. В заключение в статье сделан новый прогноз, согласно которому вычислительная техника будет играть не только инструментальную роль, но и окажет существенное влияние на мировоззренческие ценности общества. Это произойдет как по причине усиления роботизации, так и за счет вытеснения машинами людей творческих профессий, что усилит сегрегацию общества и тенденции к индивидуализму. Проблема реализма и поиск критериев истины будет также обостряться по мере развития самопрограммируемых нейросетей.

Ключевые слова: философия искусственного интеллекта, нейронные сети, информация, технический прогресс, мышление, социальная философия, общество.

Анна Юрьевна Сторожук – доктор философских наук, ведущий научный сотрудник Института философии и права Сибирского отделения РАН.

stor71@mail.ru

<http://orcid.org/0000-0002-0893-6212>

Для цитирования: *Сторожук А.Ю.* Искусственный интеллект: как углеродная жизнь создала кремниевую // Философские науки. 2021. Т. 64. № 1. С. 134–148. DOI: 10.30727/0235-1188-2021-64-1-134-148

Artificial Intelligence: How Carbon-Based Life Has Created Silicon-Based Life

A.Yu. Storozhuk

*Institute of Philosophy and Law, Siberian Branch,
Russian Academy of Sciences, Novosibirsk, Russia*

Abstract

The paper considers the main stages of the development of artificial intelligence (AI). The aim of the paper is to reveal the influence of computerization on social life at present and in the future. The historical analysis method was used to identify the main stages in the development of neural networks and to compare some predictions and real achievements in the field of AI modeling. The article identifies three main periods, widely known as the first and second spring of the era of artificial intelligence and the “winter of AI” – a period of stagnation. The earliest promising developments ran into technical difficulties. At the first stage, the developers tried to copy nature, that is, the neurons of the natural brain, and the tasks turned out to be too complicated. In addition, the first algorithms for training neural networks encountered fundamental difficulties, such as exponential decay of the error vector when propagating into the inner layers of the neural network. It took years to find new algorithmic solutions, to accumulate data for training neural networks, for the next technological breakthrough, known as the “second spring of AI.” The result of the analysis showed that predictions of AI capabilities were often more modest than actual achievements. In the present paper, we make new predictions, according to which computer technology is going play not only an instrumental role but will also have a significant impact on the ideological values of society. This will happen not only due to increasing automation but also due to substitution of creative professions by machines, which will increase the segregation of society and the trend towards individualism. The problem of realism and the search for criteria of truth will also gain importance as self-programmable neural networks evolve.

Keywords: philosophy of artificial intelligence, neural networks, information, technical progress, thinking, social philosophy, society.

Anna Yu. Storozhuk – D.Sc. in Philosophy, Leading Research Fellow, Institute of Philosophy and Law, Siberian Branch, Russian Academy of Sciences.

stor71@mail.ru

<http://orcid.org/0000-0002-0893-6212>

For citation: Storozhuk A.Yu. (2021) Artificial Intelligence: How Carbon-Based Life Has Created Silicon-Based Life. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 64, no. 1, pp. 134–148.

DOI: 10.30727/0235-1188-2021-64-1-134-148

Введение

Термин «искусственный интеллект» появился в середине XX века, когда обсуждалась теоретическая возможность построить мыслящие машины. Предполагалось, что они смогут заметить человека во вредных условиях (шахта, космос) или будут выполнять монотонную работу, например, отбраковку деталей на заводах. При этом машине отводилась подчиненная роль послушного помощника.

Положительные ответы на вопрос «Может ли машина мыслить?» разделили на два класса. Речь идет о сильном и слабом искусственном интеллекте (ИИ). Последний был самым распространенным, т.к. предполагал лишь частичное превосходство машины в некоторых областях, таких как счет, память [Безлепкин 2020], поиск данных и т.п., что стало очевидным уже на ранних этапах развития вычислительной техники. Сильный ИИ имел немного сторонников, поскольку вплоть до начала нынешнего века не допускали мысли о возможном превосходстве машины над человеком в творчестве, интуиции, умении открывать новое знание [Чалмерс 2013]. Полагали, что машина не производит нового знания, но лишь быстро ищет в базе данных имеющиеся варианты или быстро считает [Фарберович 2019]. Кроме того, считалось невозможным для ИИ решение задач, требующих бесконечного перебора, в условиях действия непредсказуемых факторов и т.п. Например, нельзя не вспомнить о том, как компьютерная программа обыграла в шахматы чемпиона мира (шахматный суперкомпьютер *Deep Blue* в 1997 году выиграл у чемпиона мира Гарри Каспарова). Хотя знатоки утверждали, что компьютер никогда не

выиграет у человека, т.к. данная игра требует бесконечного перебора вариантов. По состоянию на 2016 год программа *AlphaGo Zero* с большим отрывом обыграла всех чемпионов мира и национальных чемпионов [Silver, Huang, Maddison et al. 2016].

Насколько оправдались ожидания и долговременные прогнозы, сделанные десятилетия назад? Каковы сегодня отношения человека и машины? Чего можно ожидать в ближайшее время? Каково влияние компьютеризации на общество? Как компьютеры изменили отношения между людьми?

Времена года: краткая история искусственного интеллекта

Сегодняшние дни принято называть «второй весной» ИИ. «Первой весной» именуют период энтузиазма, охвативший научное сообщество в середине XX века, когда появились первые алгоритмы и механические устройства, подобные «черепашке» Грея Уолтера [Walter 1950]. Волна энтузиазма нашла на камень трудностей технического характера, которые были непреодолимы во второй половине прошлого века, известного как «зима» ИИ.

Первые попытки создания искусственного разума представляли собой копирование природы. Архитектура искусственного нейрона создавалась по образу и подобию естественной нервной клетки, первая компьютерная нейронная сеть пыталась подражать нейронным сетям животных [Коннова 2020]. Подражание служило бледным подобием, т.к. естественная нервная клетка может иметь до двадцати тысяч синапсов, соединяющих ее с другими клетками. Кроме того, известно о перцептивных полях, т.е. объединениях нервных клеток сетчатки в определенные структуры. По образу и подобию организации нервных клеток в сетчатке кошки созданы первые искусственные нейронные сети, самой известной из которых является перцептон Розенблатта [Rosenblatt 1961]. Он представлял собой массив, состоящий из входного слоя, одного внутреннего слоя (т.е. однослойной нейронной сети) и выходного нейрона, куда записывались результаты вычислений. Каждый нейрон входного слоя был связан со всеми нейронами внутреннего слоя, который и производил вычисления. Часто это выглядело как суммирование значений входных нейронов, умноженных на коэффициент, который определял т.н. вес данной связи. Перцептон был обучаемой нейронной сетью. Обучение относилось к корректировке весов, т.е. на изменения коэффициентов умножали входные данные. Чтобы произвести корректировку весов, резуль-

тат нейронной сети сравнивали с заранее известным правильным ответом и вычисляли вектор ошибок, в зависимости от величины которого производили корректировку весов.

Очень быстро начались проблемы технического характера. Так, для однослойного перцептона не была доказана теорема сходимости в отношении некоторых логических операций [Minsky, Papert 1969]. Кроме того, его мощности не хватало для решения сложных задач. Появились сети с большим числом внутренних слоев. Но у многослойного перцептона возникли проблемы с обучением, т.к. вектор ошибки при распространении внутрь затухал экспоненциально, т.е. обучались практически один-два выходных слоя.

Как могла наступить «вторая весна»? Выделим три источника или составных части: появление новых алгоритмов, накопление больших массивов данных, возникновение новых вычислительных мощностей. Новые алгоритмы были нацелены на улучшение обучаемости нейронных сетей. Последние имели преимущества перед жестко заданными алгоритмами в своей гибкости. Нейронные сети позволяли параллельно обрабатывать массивы информации. В частности, наиболее известны рекурсивные нейронные сети, которые способствовали распространению вектора ошибок на внутренние слои [Rumelhart, Hinton, Williams 1986]. Нейронные сети обладали высокой помехоустойчивостью: нарушения приводили к снижению их производительности, но не к полному отказу.

Об этом писал и Н. Бостром: «Важно, что нейронные сети представляли собой самообучающиеся интеллектуальные системы, т.е. накапливали опыт, умели делать выводы из обобщенных примеров и находить скрытые статистические образы во вводимых данных. Это делало сети хорошим инструментом для решения задач классификации и распознавания образов. Например, создав определенный набор сигнальных данных, можно было обучить нейронную сеть воспринимать и распознавать акустические особенности подводных лодок, мин и морских обитателей с большей точностью» [Бостром 2016, 20]. Широкое распространение получил блочный подход – конструктор нейронных сетей, т.е. ряд блоков, из которых можно собирать нейронные сети [Варламов 2011]. Развитие последних стало возможным благодаря новым методам представления знаний в компьютерах как ядра ИИ. Методы поиска решений носят в основном переборный характер. Но в различных представлениях одна и та же задача может иметь разный

размер. Более компактное представление позволяет решать более сложные задачи при аналогичных вычислительных мощностях [Новиков 2019, 11].

Существенным условием, обеспечившим очередной рывок в технологиях ИИ, т.е. наступление «второй весны» ИИ, стало появление новых компьютерных технологий – методов эволюционных вычислений, таких как генетический алгоритм. Бостром объяснял происходящее следующим образом: «В эволюционных моделях в первую очередь создаются начальные популяции тех или иных решений (могут быть либо структуры данных, либо программы обработки данных), затем – в результате случайной мутации и размножения (“скрещивания”) имеющихся популяций – генерируются новые популяции. Периодически вследствие применения критерия отбора (по наличию целевой функции, или функции пригодности) количество популяций сокращается, что позволяет войти в новое поколение лишь лучшим решениям-кандидатам. В ходе тысяч итераций среднее качество решений в популяции постепенно повышается» [Бостром 2016, 20]. Генетические алгоритмы используют идею естественного отбора Дарвина и позволяют получать самые «жизнеспособные» продуктивные программы.

Огромную роль в приближении «второй весны» сыграло накопление большого количества данных. Для обучения нейронной сети требуется много образцов¹. На начальном этапе развития ИИ трудно было обеспечить такое количество. Сегодня, благодаря сети Интернет и социальным сетям, у многих пользователей появилась возможность выкладывать собственные изображения и тексты. Это обеспечило появление объема данных, достаточного для качественного обучения нейронных сетей.

Важную роль сыграло и увеличение вычислительных мощностей. Благодаря видеоиграм и криптовалютам получили широкое распространение скоростные процессоры с мощными видеокартами. Некоторые операции, например, сортировка, быстрее происходят на видеокартах, чем на процессорах. Это оказалось решающим фактором, т.к. обучение и работа нейросетей требует больших вычислительных мощностей. В итоге к 2020 году, по оценке М.С. Бурцева, если сравнивать с живой природой по коли-

¹ Например, чтобы научить нейронную сеть отличать кошку от собаки, необходимо около миллиона изображений. (См.: *Хуришудов А. Про котиков, собак, машинное обучение и deep learning // Хабр. 2014. 16 июня. – URL: <https://habr.com/ru/post/226347/>*)

честву нейронов, то общий уровень ИИ сравним с интеллектом мыши [Соколенко 2020].

Влияние прогресса технологий ИИ на развитие общества

Еще несколько десятилетий назад Р. Пенроуз в работе «Новый ум короля» (1989) писал, что компьютеры никогда не смогут заменить человека, поскольку у них нет понимания того, что они делают [Пенроуз 2020, 33]. По аналогии с мысленным экспериментом «китайская комната» Серла [Searle 2001] Пенроуз заключает, что нейросеть может быть хорошей имитацией, но она никогда не достигнет понимания. Кроме того, некоторые невербальные способы мышления не алгоритмизуемы [Пенроуз 2020, 338–344].

Однако сегодня машины прошли тест Тьюринга (случай испытания чат-ботов описан в 2014 году [Новиков 2019, 21]). Давно создана нейросеть, которая работает как психотерапевт [Weizenbaum 1966]. Нейросети занимаются творчеством. Они пишут статьи, сочиняют музыку, рисуют, декламируют и даже программируют другие нейросети, т.е. вытесняют человека из ряда типичных творческих специальностей.

В ранний период появления ИИ считали, что машина лишь комбинирует имеющиеся знания и не способна породить ничего нового. Однако, когда нейросеть выиграла у чемпиона мира партию в го [Metz 2016], специалисты, анализирующие партию, обнаружили непонятные ходы, сделанные нейросетью. Смысл этих ходов стал ясен только к середине партии. Это была позиционная игра, которая и позволила нейросети выиграть. Интересен тот факт, что подобные приемы в го ранее не использовались и не были известны. Профессиональные игроки в го нашли ходы нейросети очень изящными. Если говорить о стратегии именно этой игры, то машина в данном случае создала новое.

Еще одна задача для нейросети – научиться рисовать «под Ван Гога» [Gatys, Ecker, Bethge 2020]. В целях ее решения использовали две нейросети. Задача для одной из них состояла в том, чтобы на основе анализа картин Ван Гога научиться перерисовывать фотографии в стиле его живописи. Другая нейросеть должна была установить, подлинная ли картина. В результате совместного обучения первая нейросеть выдавала такие результаты, что человек уже не мог отличить подлинник от подделки. Иными словами, рисунок нейросети предстал таким, который мог бы принадлежать Ван Гогу. Только знание того, какие картины

были им написаны при жизни, позволило экспертам отличить подлинник от т.н. компьютерной подделки.

В настоящее время, на наш взгляд, влияние компьютеров на общество приобрело экзистенциальный характер. Компьютер дает возможность суррогатной жизни подобно тому, как соевые продукты представляют собой суррогатное мясо. Прежде всего важной частью является суррогатное общение. Усиливается индивидуализация общества, подогреваемая, помимо остального, активно насаждаемой моралью потребительского общества [Santos, Varnum, Grossmann 2017].

Существенное воздействие на психику и социальное поведение подростков оказывают также компьютерные игры, в которых «герой обычно не учитывает чувства или интересы других персонажей и демонстрирует утилитарное отношение к другим» [Noga 2016, 578]. Наблюдается разница в социальном поведении и системе ценностей между подростками, увлеченными играми с элементами агрессивности и жестокости, и подростками, которые в них не играют. Влияние медиакультуры на общество исследуется в специальных работах. Обратим внимание лишь на философски значимые аспекты, такие как влияние на формирование мировоззрения и новых стилей когнитивной активности, рост социального контроля, изменение системы ценностей и влияние на восприятие мира, понимание реальности [Perse, Lambe 2016, 1].

С развитием компьютерных технологий повышается точность воспроизведения образов. Ряд изображений, созданных нейросетями, люди уже не могут отличить от настоящих. Произошли серьезные изменения в кинематографе: во многих жанрах уже практически не снимаются картины без широкого использования компьютерной графики. На первый взгляд представляется, что компьютерное моделирование способствует повышению чувства реализма. Если проводить аналогию с развитием фотографии, то нельзя не отметить, что сначала, при переходе от черно-белых фотографий к цветным, степень реализма повышалась. Фотоискусство носило документальный характер, играло роль свидетельства, подтверждающего истинность словесной истории. С развитием программ для обработки изображений, таких как *Photoshop*, фотография теряет роль объективного свидетельства, поскольку подделка становится порой очень качественной, настолько, что ее сложно отличить от реально сделанной фотографии. Сегодня появляются нейросети, способные обрабатывать

фотографии и путем их комбинации получать практически любые изображения.

Тема реализма тесно связана с понятием истинности. В контексте научных теорий различали две основные теории: когерентную, предусматривающую отсутствие внутреннего противоречия, и корреспондентную, базирующуюся на идее о соответствии внешнему миру. В настоящее время, когда статья может быть написана программой, генерирующей тексты, а доступ к документальным свидетельствам опосредован компьютером, проблема отличия истины от лжи может оказаться неразрешимой.

Изначально техника создавалась, чтобы расширить возможности людей и облегчить им труд. Но в итоге такое расширение возможностей с помощью компьютерных технологий привело к подчиненному положению людей. Происходит замещение человека автоматикой. Люди, которые остаются, играют роль обслуживающего персонала, т.е. они нужны, чтобы помогать машинам, а не для того, чтобы машина помогала им [Маркофф 2016, 78]. Если это происходит сегодня, то ожидаемое потенциальное взрывоподобное развитие интеллекта, когда алгоритмы самостоятельно будут писать новое, более совершенное, данный процесс только усилит.

Заключение

В основе проектирования нейронных сетей, иных видов алгоритмов, актуальных при создании ИИ, находится подражание устройству живых организмов. Однако это подражание, по сути, происходит лишь в одном аспекте – информационном. По аналогии с естественной памятью создаются условия для хранения информации, по аналогии с естественной коммуникацией происходит передача, обработка, защита от помех информации в вычислительных машинах. Скорость технического прогресса намного опережает скорость биологической эволюции. Информационные технологии прогрессируют стремительно, намного медленнее развивается мир животных и растений. Многие обеспокоены тем, что человечество утратит пальму первенства в вопросе доминирования над другими биологическими видами, а также займет подчиненную роль, сравнимую с местом домашних животных [Butler 1863]. Финальной аналогией мы обязаны книге Ч. Докинза «Эгоистичный ген» [Докинз 2013]. Если мы заменим слово «ген» словом «информация», то, в сущности, создадим усло-

вия для более быстрого и широкого распространения последней. Если биологические существа, состоящие из белков, копировали и распространяли информацию в ограниченных масштабах, то полупроводниковые устройства намного превосходят углеводородные по скорости копирования и эволюции.

Фактически мы сами создаем преемников. Речь идет о следующей цивилизации, но уже на кремниевой основе. Основные типы прогнозов относительно будущего человечества в атмосфере доминирования ИИ можно разделить условно на три класса.

1. Оптимистичные сценарии, предполагающие улучшение условий жизни человека. Первая функция ИИ, как и любых других вычислительных мощностей, состоит в том, чтобы быть локомотивом экономики. Анализ цивилизаций за последние две тысячи лет показал [Колташов 2018], что причины экономических кризисов всегда были одними и теми же – падение покупательского спроса и истощение сельскохозяйственных земель. Выход из кризиса тоже был принципиально одинаковым, он связан с техническими новшествами, разогревающими спрос на рынках. Компьютерная индустрия является особенно значимым сегментом мировой экономики и практически единственным, показавшим рост в условиях эпидемии COVID-19. Взрыв спроса произошел на видеоигры, программы для дистанционного общения, проведения конференций и др. Тем более что данный сегмент рынка не является ресурсоемким: в основе вычислительной техники лежат полупроводниковые микросхемы, а кремний – третий по распространенности элемент земной коры, т.е. сырье для изготовления полупроводниковой техники общедоступно. Информация служит, возможно, единственным ресурсом многоразового использования, остальные ресурсы расходуются при пользовании ими, сокращаются. Т.н. развитые страны занимаются производством, скорее, моделей производства, которые воплощаются в других странах с индустриальными обществами [Федотова 2016, 185].

Ожидается значительное улучшение жизни людей как за счет оптимизации социальной организации, так и за счет свободного доступа к информационным ресурсам. Дж. Нэбит и П. Абурден в исследовании «Мегатенденции: десять новых направлений, преобразующих нашу жизнь» предсказали, что произойдет смена модели управления обществом: переход от централизованной иерархической структуры к децентрализованной сетевой структуре [Naisbitt, Aburdene 1991].

2. *Негативные сценарии, связанные с подчинением людей машинам.* В настоящее время ИИ во многих сферах интеллектуальной деятельности превосходит возможности человека. Неслучайно возникает следующий вопрос. Сложный для понимания ИИ уже сегодня, станет ли он понятнее через пять или сто лет? Предполагается, что одновременно со сложностью и превосходством будет расти и могущество машин, а значит, и способность влиять на жизнь людей. Негативные сценарии основаны на приписывании машинам человеческих слабостей: жажды денег и власти, тщеславия, иных. Насколько реалистичны такие тенденции?

Многие авторы, начиная с фантаста А. Азимова, писавшего о трех законах роботехники, указывают на необходимость передачи машинам моральных норм. В этом видится антропоморфизация техники: именно моральные нормы и совесть являются сдерживающей силой естественного интеллекта. Но как «воспитать» машину? Н. Бостром делает акцент на трудности формализации моральных правил, а также на том, что «зрелый», т.е. прошедший обучение ИИ, не склонен изменяться. Следовательно, моральные правила и систему запретов нужно закладывать изначально *in silico* [Бостром 2016].

Представляется надуманным приписывание технике человеческой мотивации. Тем не менее даже при ее отсутствии техника способна негативно влиять на жизнь людей. Вероятно, будет продолжаться вытеснение людей из ряда интеллектуальных и творческих профессий. Ранее существовало мнение о том, что машины заменят человека на тяжелых работах во вредных условиях. Однако стоимость оборудования для работы нейросети такова, что ее не отнести к вредным условиям. Скорее, человек самостоятельно пойдет в шахту добывать уголь, чтобы удовлетворить растущие потребности компьютеров в электроэнергии. При появлении первых вычислительных машин возник целый штат сотрудников по их обслуживанию: программисты, операторы, электрики, служба профилактики и технического обслуживания. Эта тенденция станет еще более активна. В сферу обслуживания машин будет вовлечено все больше людей.

3. *Радикальное изменение человеческого и социального.* Видение в черно-белых тонах является слишком примитивным пониманием истории. На наш взгляд, ни одну тенденцию нельзя охарактеризовать однозначно в позитивном или негативном ключе.

Но техника окажет и уже оказывает значительное влияние на жизнь человека, в чем-то облегчая его существование, а в чем-то требуя уступок и жертв.

Тенденции развития общества могут оказаться разнонаправленными: с одной стороны, происходит объединение человечества в информационном пространстве с медиакоммуникациями, позволяющими осуществлять общий контроль, вести пропаганду универсальных ценностей (по принципу равенства всех перед всеми). С другой стороны, как нам кажется, равенства не будет, развитие ИИ приведет к значительному расслоению общества и его дифференциации. Менее образованные будут занимать место пользователей и обслуживающего персонала, наиболее способные – выполнять функции программистов и наладчиков оборудования. Учитывая экспоненциально возрастающую скорость технического прогресса и тем, и другим придется постоянно учиться, чтобы сохранить свой *status quo*. Постоянное обучение в аспекте технической составляющей окажет влияние и на структуру сознания. Структурные и математические способности будут развиваться, творческие «правополушарные» способности и интуиция окажутся «за бортом» мейнстрима. Цифровая жизнь подчинит себе жизнь экономическую и политическую, а повсеместность цифровых технологий будет обеспечивать универсальность и унифицированность мышления, системы ценностей, что создаст условия для концентрации контроля и власти в одних руках.

Свобода будет широко пропагандироваться, но в действительности свобода перемещения, выбора профессии, самореализации, поиска себя будут в значительной степени ограничены. Произойдет существенное ослабление авторитета мудрых людей, т.к. техническая грамотность заменит мудрость, а жизненный опыт окажется не применимым в быстро изменяющихся условиях и технологиях. Кроме того, виртуальное (суррогатное) общение будет способствовать потере навыков реального общения, а значит, и развитию индивидуализма. Даже находясь в обществе, человек будущего не сможет продуктивно общаться, не станет членом коллектива. Рано или поздно эти эффекты, вызванные развитием техники, приведут к отказу от традиционных общественных ценностей (любви, дружбы, взаимовыручки), от чувства коллективизма. В этом проявляются, полагаем, основные риски цифровизации.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Безлепкин 2020 – *Безлепкин Е.А.* Сообщество разума: идеи М. Минского об интеллектуальных агентах // *Философия науки*. 2020. № 3 (86). С. 105–118.

Бостром 2016 – *Бостром Н.* Искусственный интеллект. Этапы. Угрозы. Стратегии. – М.: Манн, Иванов и Фербер, 2016.

Варламов 2011 – *Варламов О.* Миварный подход к разработке интеллектуальных систем и проект создания мультипредметной активной миварной интернет-энциклопедии // *Известия Кабардино-Балкарского научного центра РАН*. 2011. № 1 (39). С. 55–64.

Докинз 2013 – *Докинз Р.* Эгоистичный ген. – М.: АСТ, CORPUS, 2013.

Колташов 2018 – *Колташов В.* Экономические кризисы с III века до наших дней: цикличность, природа и логика современных потрясений // *Логос*. 2018. Т. 28. № 6 (127). С. 27–72.

Коннова 2020 – *Коннова Н. С.* Нейронные сети, или Как обучить искусственный интеллект // *Интернет изнутри*. 2020. № 3. С. 38–51.

Маркофф 2016 – *Маркофф Д.* Homo Roboticus? Люди и машины в поисках взаимопонимания. – М.: Альпина нон-фикшн, 2016.

Новиков 2019 – *Новиков Ф.А.* Символический искусственный интеллект: математические основы представления знаний. – М.: Юрайт, 2019.

Пенроуз 2020 – *Пенроуз Р.* Новый ум короля: о компьютерах, мышлении и законах физики. – М.: URSS, Ленанд, 2020.

Соколенко 2020 – *Соколенко О.* «Сегодня общий уровень ИИ сопоставим с уровнем интеллекта мыши или пчелы. Боймся ли мы, что мыши и пчелы захотят захватить мир?» Заведующий лабораторией нейронных систем и глубокого обучения МФТИ Михаил Бурцев – о рисках искусственного интеллекта для человечества // *Republic*. 2020. 25 сентября. – URL: <https://republic.ru/posts/97895>

Федотова 2016 – *Федотова В.Г.* Модернизация и культура. – М.: Прогресс-Традиция, 2016.

Чалмерс 2013 – *Чалмерс Д.* Сознательный ум. В поисках фундаментальной теории. – М.: Либроком, 2013.

Butler 1863 – *Butler S.* Darwin Among the Machines // *The Press*. 1863. June 13. P. 180–185.

Gatys, Ecker, Bethge 2020 – *Gatys L., Ecker A., Bethge M.* A Neural Algorithm of Artistic Style // *Journal of Vision*. 2016. Vol. 16. No. 12. P. 1–16.

Metz 2016 – *Metz C.* Why the Final Game Between AlphaGo and Lee Sedol Is Such a Big Deal for Humanity // *Wired News*. 2016. March 14. – URL: <https://www.wired.com/2016/03/final-game-alphago-lee-sedol-big-deal-humanity/>

Minsky, Papert 1969 – *Minsky M., Papert S.* Perceptrons. – Cambridge, MA: MIT Press, 1969.

Naisbitt, Aburdene 1990 – *Naisbitt J., Aburdene P.* Megatrends 2000. Ten New Directions for the 1990's. – New York: William Morrow, 1990.

Noga 2016 – *Noga H.* Selected Factors of the Socialization Impact of Computer Games and Their Content // *Society. Integration. Education: Proceedings of the International Scientific Conference*. 2016. Vol. 2. P. 569–578.

Perse, Lambe 2016 – *Perse E.M., Lambe J.* Media Effects and Society. – New York: Routledge, 2016.

Rosenblatt 1961 – *Rosenblatt F.* Principles of Neurodynamics. – Washington, DC: Spartan Books, 1961.

Rumelhart, Hinton, Williams 1986 – *Rumelhart D.E., Hinton G.E., Williams R.J.* Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations. – Cambridge, MA: MIT Press, 1986.

Santos, Varnum, Grossmann 2017 – *Santos H.C., Varnum M.E.W., Grossmann I.* Global Increases in Individualism // *Psychological Science*. 2017. Vol. 28. No. 9. P. 1228–1239.

Searle 2001 – *Searle J.* Chinese Room Argument // *The MIT Encyclopedia of the Cognitive Sciences* / ed. by R.A. Wilson, F.C. Keil. – Cambridge, MA: MIT Press, 2001. P. 115–116.

Silver, Huang, Maddison et al. 2016 – *Silver D., Huang A., Maddison C. et al.* Mastering the Game of Go with Deep Neural Networks and Tree Search // *Nature*. 2016. Vol. 529. No. 7587. P. 484–489.

Walter 1950 – *Walter W.G.* An Imitation of Life // *Scientific American*. 1950. Vol. 182. No. 5. P. 42–45.

Weizenbaum 1966 – *Weizenbaum J.* ELIZA – a Computer Program for the Study of Natural Language Communication between Man and Machine // *Communications of the ACM*. 1966. Vol. 9. No. 1. P. 36–45.

REFERENCES

Bezlepkin E. A. (2020) The Community of Reason: M. Minsky's Ideas about Intellectual Agents. *Filosofiya nauki*. No. 3, pp. 105–118 (in Russian).

Bostrom N. (2014) *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press (Russian translation: Moscow: Mann, Ivanov and Ferber, 2016).

Butler S. (1863, June 13) Darwin Among the Machines. *The Press*. Pp. 180–185.

Chalmers D. (1996) *The Conscious Mind. In Search of a Fundamental Theory*. New York: Oxford University Press (Russian translation: Moscow: Linrok, 2013).

Dawkins R. (1976) *The Selfish Gene*. Oxford: Oxford University Press (Russian translation: Moscow: AST, Corpus, 2013).

Fedotova V.G. (2016) *Modernization and Culture*. Moscow: Progress-Traditsiya (in Russian).

Gatys L., Ecker A., & Bethge M. (2020) A Neural Algorithm of Artistic Style. *Journal of Vision*. Vol. 16, no. 12, pp. 1–16.

Koltashov V. (2018) Economic Crises from the Third Century to the Present Day: Cyclicity, Nature and Logic of Modern Upheavals. *Logos*. Vol. 28, no. 6, pp. 27–72 (in Russian).

Konnova N.S. (2020) Neural Networks, or How to Train Artificial Intelligence. *Internet iznutri*. No. 3, pp. 38–51 (in Russian).

Markoff D. (2015) *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots*. New York: Ecco Books (Russian translation: Alpina non-fiction, 2016).

Metz C. (2016, March 14) Why the Final Game Between AlphaGo and Lee Sedol Is Such a Big Deal for Humanity. *Wired News*. Retrieved from <https://www.wired.com/2016/03/final-game-alphago-lee-sedol-big-deal-humanity/>

Minsky M. & Papert S. (1969) *Perceptrons*. Cambridge, MA: MIT Press.

Naisbitt J. & Aburdene P. (1990) *Megatrends 2000. Ten New Directions for the 1990's*. New York: William Morrow.

Noga H. (2016) Selected Factors of the Socialization Impact of Computer Games and Their Content. *Society. Integration. Education: Proceedings of the International Scientific Conference*. Vol. 2, pp. 569–578.

Novikov F.A. (2019) *Symbolic Artificial Intelligence: Mathematical Foundations of Knowledge Representation*. Moscow: Urait (in Russian).

Penrose R. (1989) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford: Oxford University Press (Russian translation: Moscow: URSS, Lenand, 2020).

Perse E.M. & Lambe J. (2016) *Media Effects and Society*. New York: Routledge.

Rosenblatt F. (1961) *Principles of Neurodynamics*. Washington, DC: Spartan Books.

Rumelhart D.E., Hinton G.E., & Williams R.J. (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*. Cambridge, MA: MIT Press.

Santos H.C., Varnum M.E.W., & Grossmann I. (2017) Global Increases in Individualism. *Psychological Science*. Vol. 28, no. 9, pp. 1228–1239.

Searle J. (2001) Chinese Room Argument. In: Wilson R.A. & Keil F.C. (Eds.) *The MIT Encyclopedia of the Cognitive Science* (pp. 115–116). Cambridge, MA: MIT Press.

Silver D., Huang A., Maddison C. et al. (2016) Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*. Vol. 529, no. 7587, pp. 484–489.

Sokolenko O. (2020, September 25) “Today, the general level of AI is comparable to the level of intelligence of a mouse or bee. Are we afraid that mice and bees will want to take over the world?” Head of MIPT Laboratory of Neural Systems and Deep Learning Mikhail Burtsev – on the risks of artificial intelligence for humanity. *Republic*. Retrieved from: <https://republic.ru/posts/97895> (in Russian).

Varlamov O.O. (2011) Mivar's Approach to Development of Intelligent Systems and Project of Development of Active Multi-Subject Online Mivar's Internet Encyclopedia. *Bulletin of Kabardino-Balkaria Scientific Center of RAS*. No. 1, pp. 55–64 (in Russian).

Walter W.G. (1950) An Imitation of Life. *Scientific American*. Vol. 182, no. 5, pp. 42–45.

Weizenbaum J. (1966) ELIZA – a Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM*. Vol. 9, no. 1, pp. 36–45.