

УДК 001.11: 004.8: 008.2

ОТ ТЕХНОЛОГИЧЕСКОЙ СИНГУЛЯРНОСТИ – К БУДУЩИМ СЦЕНАРИЯМ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Галстян Р.В., Михайлова Т.Л.

ФГБОУ ВО «Нижегородский государственный технический университет им. Р.Е. Алексеева»,
г. Нижний Новгород, e-mail: tmichailova2012@yandex.ru

В статье рассматриваются будущие сценарии развития искусственного интеллекта как репрезентанта междисциплинарных проблем современной технауки, что повышает значимость методологического потенциала философии как «ничейной», пограничной сферы знания. Приводится методологический обзор качественно новых форм искусственного интеллекта, рубеж отсчета которых предлагается начать с точки «невозврата», именуемой технологической сингулярностью. Контекстуальное поле обзора форм искусственного интеллекта включает исследование качественных различий между машиной и человеком, дифференциация которых позволяет обозначить специфику человеческого мышления, в частности, его гибкость, умение работать с неточными идеями. В качестве практического результата авторы рассматривают перспективность направления, связанного с самообучением машин, что по прогнозам Рея Курцвейла, приведет к неотличимости человеческого и компьютерного интеллекта. Итог статьи – вывод о принципиальном отличии человеческого интеллекта, суть которого не только в его гибкости, но и социокультурной детерминации.

Ключевые слова: технологическая сингулярность, искусственный интеллект, экспертная система, самообучение машин, искусственные нейронные сети, человеческое мышление, воображение, междисциплинарная проблема, технаука, граница, эвристика.

FROM TECHNOLOGICAL SINGULARITY TO FUTURE SCRIPTS OF DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

Galstyan R.V., Mikhailova T.L.

Nizhny Novgorod State Technical University n.a. R.E. Alekseev, Nizhny Novgorod,
e-mail: tmichailova2012@yandex.ru

The article considers future scenarios for the development of artificial intelligence as a representative of interdisciplinary problems of modern technology, which increases the importance of the methodological potential of philosophy as a «no man's land», the borderline sphere of knowledge. A methodological review of qualitatively new forms of artificial intelligence is presented, the reference point of which is proposed to start from the point of "no return", called the technological singularity. The contextual field of the review of forms of artificial intelligence includes the study of the qualitative differences between a machine and a man, the differentiation of which allows us to identify the specificity of human thinking, in particular, its flexibility, the ability to work with inaccurate ideas. As a practical result, the authors consider the perspective of the direction connected with the self-learning of machines, which according to Ray Kurzweil's predictions will lead to the indistinguishability of human and computer intelligence. The conclusion of the article is a conclusion about the fundamental difference between the human intellect, the essence of which is not only in its flexibility, but also in sociocultural determination.

Keywords: technological singularity, artificial intelligence, expert system, self-learning machines, artificial neural networks, human thinking, imagination, interdisciplinary problem, technological science, boundary, heuristics.

Актуальность выбора темы обоснована потребностью изучения растущего воздействия компьютерных и коммуникационных технологий, создающих новые конфигурации общества и формирующих универсум человеческого бытия. Возможности и угрозы технологического прогресса находят отражение в фантастических фильмах и книгах. Как правило, такие произведения повествуют о будущем, в котором технологический прогресс в области компьютерных и коммуникационных технологий вывел общество на качественно новый уровень жизни. Люди в этом мире имеют иные ценности и взгляды. Заменив человеческий труд машинами, человек начал жить значительно лучше, нежели самые богатые люди живут сегодня. Вместе с искусственными помощниками они ищут ответы на самые разноо-

бразные вопросы бытия. Этому качественно новому уровню жизни предшествует некий технологический рубеж – технологическая сингулярность [5]. Контекстом рассмотрения будущего искусственного интеллекта и является изучение технологической сингулярности, возможных путей ее возникновения посредством системного подхода через выделение комплекса критериев. Дадим определение технологической сингулярности. Технологической сингулярностью называется предположительная точка во времени, после которой развитие технического прогресса станет настолько быстрым и сложным, что сама техника окажется недоступной пониманию человека [6].

В книге «Об интеллекте» Джефф Хоккинс представляет свою теорию строения мозга, основанную на знаниях в областях

нейробиологии, психологии и кибернетики. Он доказывает, что вся «верхняя» работа мозга базируется на системе память-предсказание и системе «паттернов» [7], из которых состоит наш мир, демонстрируя это на примерах. **Джефф Хокинс** объясняет, почему попытки создания искусственного интеллекта до сих пор не увенчались успехом, полагая, что за этим стоит не техническая, а принципиальная проблема. Принципиальность обозначенной трудности создания И.И. инициирует наш интерес к теме.

Давайте разберемся, чем же занимается наш мозг, когда мы пассивно слушаем чью-то речь или читаем. Информация поступает в мозг, но не выходит оттуда. Что с ней происходит? Наше внешнее поведение в данный момент остается на базовом уровне – мы дышим, совершаем движения глазами яблоками. В то же время наш мозг продельывает колоссальную работу, благодаря которой возможны чтение и понимание прочитанного. Понимание есть результат деятельности нейронов. Но какой деятельности? Чем именно занимаются нейроны в процессе осознания? Конечно, мозг не разговаривает сам с собой, составляя прогнозы, и он не составляет прогнозы в серийном виде. Кроме того, он не составляет прогнозы относительно каких-то отдельных объектов. Мозг составляет прогнозы о самой структуре мира, в котором мы живем, и делает это, используя параллельные формы. Он с одинаковой готовностью распознает непривычную поверхность, деформированный нос или неожиданное движение. Сразу становится понятным, насколько распространенными являются такого рода несознательные прогнозы и почему на них так долго не обращали никакого внимания. Они происходят без малейших заминок, автоматически, и нам просто уловить, что происходит внутри нашего черепа. Прогностическая функция настолько органична для мозга, что наше восприятие мира не основывается исключительно на сигналах, которые мы непрерывно получаем от органов чувств. На самом деле восприятие действительности является комбинацией наших ощущений и прогнозов, составляемых мозгом на основе воспоминаний [7, с. 45]. Действительно, когда мы возвращаемся домой, хватаясь за ручку двери, чтобы открыть её, мы на подсознательном уровне уже знаем, с какой силой надо потянуть за ручку двери, чтобы она открылась. Ведь если, потянув за ручку двери, она не поддастся нам, то мы

удивимся. Наш мозг уже составил предсказание, что дверь должна открыться, но этого не случилось. Это и есть прогноз-предсказание, о котором говорил Д. Хокинс.

Цель работы – обозначить основные различия между машиной и человеком, чтобы осуществить рефлексии относительно будущего. Как известно, именно дифференциация феноменов и есть исходный пункт любого исследования. В чем эти различия? Различие между человеком и машиной, прежде всего, заключается в том, что в организме человека число элементов по порядку величин во много раз больше, чем обладает машина. По научным данным, человеческий мозг содержит около 270 основных «вычислительных» узлов нейронов, которые соединяют около 250 связей синапсов. Из этого естественно вытекает, что организация элементов в организме настолько сложна, что при помощи наших современных логических средств мы не можем еще овладеть этой сложностью. Феномен сложности – один из самых обсуждаемых в современной эпистемологии, сконцентрированной на междисциплинарных проблемах науки. Проблема будущего искусственного интеллекта как раз репрезентирует междисциплинарную область современной технауки.

Как известно, современные вычислительные системы быстро приближаются по своим вычислительным возможностям к человеческому мозгу. Искусственные нейронные сети контролируют сложнейшие системы управления и слежения, проявляют способности в области распознавания изображения вплоть до возможности создания интеллектуальных автопилотов. Помимо различия в количестве объема нейронной составляющей, между машиной и человеком существует еще и одно качественное отличие. Преимущество человека состоит в его гибкости, в его умении работать с неточными идеями, хотя создание нейронных сетей и было задумано для имитации структуры нейронов человеческого мозга и работы с зашумленной информацией. Для реализации искусственного интеллекта сейчас используются тщательно спроектированные алгоритмы перебора. Этот новый подход и состоит в построении интеллектуальных программ с использованием моделей, имитирующих структуры нейронов в человеческом мозге или эволюцию разных альтернативных конфигураций, как это делается в генетических алгоритмах и искусственной жизни. Нейронные архитектуры как сред-

ства реализации интеллекта привлекательны по многим причинам. Традиционные программы искусственного интеллекта (ИИ) могут быть слишком неустойчивы и чувствительны к шуму. Человеческий интеллект куда более гибок при обработке такой зашумленной информации, например, как лицо в затемненной комнате или разговор на шумной вечеринке. Нейронные архитектуры более пригодны для сопоставления зашумленных и недостаточных данных, поскольку они хранят знания в виде большого числа мелких элементов, распределенных по сети. Это означает, что человек обладает фантазией, другими словами, он создает понятие. Да, пока машины не обладают фантазией, но у них есть другие преимущества – скорость и точность.

На ранних этапах исследований, поиски ИИ совершались на основе таких распространенных настольных игр, как шашки, шахматы и пятнашки. Вдобавок к свойственному им «интеллектуальному» характеру такие игры имеют некоторые свойства, делающие их идеальным объектом для экспериментов. Большинство игр ведутся с использованием четко определенного набора правил – это позволяет легко строить пространство поиска и избавляет исследователей от неясности и путаницы, присущих менее структурированным проблемам. Позиции фигур легко представимы в компьютерной программе, они не требуют создания сложных формализмов, необходимых для передачи семантических тонкостей более сложных предметных областей. Игры могут порождать необычайно большие пространства состояний. Для поиска в них требуются мощные методики, определяющие, какие альтернативы следует рассматривать. Такие методики называются «эвристиками», составляющими значительную область исследований искусственного интеллекта.

Эвристика – стратегия полезная, но потенциально способная упустить правильное решение. Примером эвристики может быть рекомендация проверять, включен ли прибор в розетку, прежде чем делать предположение о его поломке, или, например, выполнять рокировку в шахматной игре, чтобы попытаться уберечь короля от шаха. Но большая часть того, что мы называем разумностью, по-видимому, опирается на эвристики, используемыми в решении задач. Это повышает особую значимость «ничейной территории», территории философии науки, интенция которой состоит в обозначении/решении междисциплинарных фило-

софских вопросов. В нашем случае интерес к будущему искусственного интеллекта и искусственной жизни – закономерно выводит на одну из таких «ничейных проблем» – проблему соотношения искусственного и естественного вообще, активно обсуждаемую и на уровне современных философов [4], и в студенческих аудиториях [1]. Проблема будущего ИИ – это частный случай проблемы соотношения искусственного и естественного. Рассмотрение взаимовлияния искусственного интеллекта и мышления, по-видимому, многое может прояснить относительно будущего ИИ.

Представляется целесообразным рассмотреть вопрос об обучении машин, ибо это обозначает комплекс задач, возникающих перед теми, кто «заглядывает» за горизонт. Ведь даже сейчас, в XXI веке, обучение остается «крепким орешком» искусственного интеллекта. Важность обучения, тем не менее, несомненна, поскольку эта способность является одной из главных составляющих разумного поведения. Экспертная система может выполнять долгие и трудоемкие вычисления для решения проблем. Но, в отличие от человеческих существ, если дать ей такую же или подобную проблему второй раз, она не «вспомнит» решение. Она каждый раз вновь и вновь будет выполнять те же вычисления – вряд ли это похоже на разумное поведение.

Одним из главных достижений ранних исследований по ИИ стало осознание важности специфического для некоторой предметной области знания. Экспертное знание – это сочетание теоретического понимания проблемы и набора эвристических правил для ее решения, являющихся, как показывает опыт, эффективными в данной предметной области. Экспертные системы создаются с помощью заимствования знаний у человеческого эксперта, кодирования их в форму, применяемую компьютером для решения похожей проблемы. Стратегии экспертных систем основаны на знаниях человека-эксперта. Хотя многие программы пишутся самими носителями знаний о предметной области, большинство экспертных систем есть плод сотрудничества между экспертом (врач, химик, геолог или инженер) и независимым специалистом по ИИ. Эксперт предоставляет необходимые знания о предметной области, описывая свои методы принятия решений, демонстрируя эти навыки на тщательно отобранных примерах, специалист по ИИ (или инженер по знаниям, как часто называют разработчиков

экспертных систем) отвечает за реализацию этого знания в программе, которая должна работать эффективно и внешне разумно. Большинство экспертных систем были написаны для специализированных предметных областей, эти области довольно хорошо изучены и располагают четко определенными стратегиями принятия решений. Проблемы, определенные на нечеткой основе здравого смысла, подобными средствами решить будет уже сложнее. Несмотря на воодушевляющие перспективы экспертных систем, было бы ошибкой переоценивать возможности этой технологии.

Выделим **основные проблемы**:

1. Недостаток здравого смысла и гибкости. Если людей поставить перед задачей, которую они не в состоянии решить немедленно, то они обычно сначала исследуют основные принципы, вырабатывая какую-то стратегию для перехода к проблеме. Экспертным системам этой способности не хватает.

2. Неспособность предоставлять осмысленные объяснения. Поскольку экспертные системы не владеют глубоким знанием своей предметной области, их пояснения обычно ограничиваются описанием шагов, которые система предприняла в поиске решения. Но они зачастую не могут пояснить, почему был выбран данный конкретный подход.

3. Трудности в тестировании. Хотя обоснование корректности большой компьютерной системы достаточно трудоемко, экспертные системы проверять особенно тяжело. Это важная проблема, ибо технологии экспертных систем применяются для таких серьезных задач, как управление воздушным движением, ядерными реакторами и системами оружия.

Очевидное решение этих проблем – заставить программы учиться самим на опыте, аналогах или примерах. Несмотря на эти ограничения, экспертные системы доказали свою ценность во многих важных областях. Хотя обучение является трудной областью, существуют и некоторые программы, которые опровергают опасения о ее неприступности. Одной из таких программ является АМ – Автоматизированный Математик, разработанный для открытия математических законов. Отталкиваясь от заложенных в него понятий и аксиом теории множеств, Математику удалось вывести из них такие важные математические концепции, как мощность множества, целочисленная арифметика и многие результаты теории чисел. АМ строил теоремы, модифицируя свою базу знаний, и

использовал эвристические методы для поиска наилучших из множества возможных альтернативных теорем. Из недавних результатов можно отметить программу Коттона, которая изобретает интересные целочисленные последовательности [Cotton и др., 2000].

Из всего этого можно допустить предположение, что машины смогут обучаться. Вычислительные машины могут учиться улучшать свою работу путем ее анализа. Существует ли вероятность того, что машины смогут учиться больше, чем человек? Если такое и произойдет, то только лишь потому, что люди перестанут учиться. Человеку учиться легче, чем машине. Если же люди начнут поклоняться машине и оставлять ей все, то за возможные неприятности и негативные последствия нам нужно будет «благодарить» только самих себя. Обучающаяся машина – это такая машина, которая не просто играет в какую-нибудь игру по твердым правилам, с неизменной стратегией, но периодически или непрерывно рассматривает результаты этой стратегии, чтобы определить, нельзя ли изменить с пользой те или иные параметры, величины или стратегии. Успешность программ машинного обучения наводит на мысль о существовании универсальных принципов, открытие которых позволило бы конструировать программы, способные обучаться в реальных проблемных областях.

О перспективах развития искусственного интеллекта говорит Рей Курцвейл, изобретатель, писатель и футурист, стоящий у истоков таких направлений компьютерной науки, как распознавание речи и символов, синтез музыки, виртуальная реальность и искусственный интеллект – в своем интервью журналу «Computerworld», изложив свой взгляд на будущее ИИ. Согласно его работе «Сингулярность рядом: когда человечество выйдет за границы биологии» («The singularity is near: when humans transcend biology», Viking Adult, 2005) [2, с.225]. Курцвейл предсказывает, что, в дальнейшем, человеческий и компьютерный интеллекты сольются, став неотличимыми. Объем данных о мозге человека, собираемых учеными, с каждым годом увеличивается почти вдвое; по мере получения все новых данных о конкретных участках мозга, люди довольно скоро и быстро смогут создавать подробные математические модели этих участков. По самым скромным подсчетам, к концу 2020-х годов мы получим очень точную, детальную эмуляцию работы всех участков

мозга. Десять квадрильонов (т.е. 10¹⁶) операций в секунду – вполне достаточно для того, чтобы эмулировать все участки мозга человека. Люди научатся объединять преимущества человеческого интеллекта, в частности, способность распознавать образы, с теми возможностями, в которых, как уже очевидно, машины нас превосходят. Более того, к концу 2040-х годов один кубический дюйм микросхемы на нанотрубках будет в 100 миллионов раз более мощным, чем мозг человека. Что касается программного обеспечения, то в 3030-х годах машины смогут обращаться к собственному исходному коду и совершенствовать его в рамках все более ускоряющегося цикла проектирования. Поэтому, в конце концов, эти системы станут значительно более интеллектуальными, чем люди, сочетая в себе преимущества биологического и небологического интеллекта [3]. Например, нанороботические белые кровяные тельца смогут загружать программное обеспечение для конкретного патогенного микроорганизма в теле человека, разрушая его буквально за несколько секунд, притом, что наши биологические кровяные тельца тратят на это несколько часов. Нанороботы, направленные в мозг, позволят человеку значительно расширить возможности своего интеллекта, мы сможем выйти за пределы биологических границ и заменить имеющееся у нас «человеческое тело версии 1.0» на кардинально обновленную версию 2.0, тем самым радикально увеличить продолжительность жизни [3]. В XXII веке, по прогнозу Курцвейла, мы будем использовать возможности материи и энергии на Земле и вокруг нее для поддержки вычислительных процессов, и интеллектуальные вычисления начнут распространяться по остальной части вселенной [3].

Подведем итог. В создании искусственного интеллекта как модели некоторого интеллекта кибернетика достигла больших успехов: созданы программы «эксперты», формирующие общие правила для решения частичных задач, самообучающиеся программы, диагностические экспертные системы, программы для различного рода игр

(морской бой, шахматы), ведутся работы области «машинного зрения», т.е. распознавания образов. Но усложнение интеллектуальных функций ЭВМ имеет свои границы, связанные не только с огромной сложностью и «тонкостью» устройства человеческого мозга как биологического феномена – продукта двух миллионов лет эволюции, но и социальной природой естественного интеллекта [8]. Поэтому техника, даже самая совершенная, всегда была и будет лишь средством деятельности человека определяющего цель функционирования техники. Вместе с тем, развитие техники идет в направлении все большей замены трудовых функций человека техническими устройствами. Не в этом ли смысл научно-технического прогресса? Однако при конструировании новых машин мы всегда должны сознавать, какие возможны последствия от их применения. Программы для этих машин должны быть всегда точно определены, в противном случае могут быть не только положительные, но и вредные последствия. Человек, несомненно, изменяет окружающую среду, а делает ли он это свыше своей способности, мы узнаем довольно скоро.

Список литературы

1. Захаров, А.С. Естественное и искусственное: противостояние продолжается... / А.С. Захаров, Т.Л. Михайлова // Международный студенческий научный вестник. 2016. № 3-4. – С. 577-581.
2. Курцвейл, Р. Эволюция разума [пер. с англ. Т.П. Мосоловой] / Р. Курцвейл. – М. Издательство «Э», 2015. – 352 с.
3. Курцвейл, Р. Сингулярность рядом: когда человечество выйдет за границы биологии / Р. Курцвейл. – Viking Adult, 2005. – Режим доступа: <https://pikmuvi.jimdo.com/2017/04/24/singularity-is-near> (дата обращения: 29.11.2016).
4. Кутырев, В.А. Человеческое и иное: борьба миров / В.А. Кутырев. – СПб.: Алетейя, 2009. – 264 с.
5. Сидоренко, О.О. Технологическая сингулярность как неминуемое событие: позитивная и негативная стороны вопроса // О.О. Сидоренко, Т.Л. Михайлова // Международный студенческий научный вестник. 2015. № 4-4. – С. 622-628.
6. Технологическая сингулярность / Википедия – свободная энциклопедия [Электронный ресурс]. – <https://ru.wikipedia.org/wiki> (дата обращения: 05.12.2016).
7. Хокинс, Дж. Об интеллекте / Дж. Хокинс, С. Блейкли. – Москва-СПб-Киев: Издательский дом «Вильямс»; 2007.
8. Цыбульская, Е.В. Негативное влияние искусственного интеллекта – через призму этики / Е.В. Цыбульская, Т.Л. Михайлова // Международный студенческий научный вестник. 2016. № 3-4. – С. 590-593.