

Философская мысль

Правильная ссылка на статью:

Розин В.М. — Две концепции искусственного интеллекта: реалистическая и утопическая // Философская мысль. — 2023. — № 2. DOI: 10.25136/2409-8728.2023.2.39739 EDN: DDLZIT URL: https://nbpublish.com/library_read_article.php?id=39739

Две концепции искусственного интеллекта: реалистическая и утопическая

Розин Вадим Маркович

доктор философских наук

главный научный сотрудник, Институт философии, Российская академия наук

109240, Россия, Московская область, г. Москва, ул. Гончарная, 12 стр.1, каб. 310

✉ rozinvm@gmail.com



[Статья из рубрики "Философия техники"](#)

DOI:

10.25136/2409-8728.2023.2.39739

EDN:

DDLZIT

Дата направления статьи в редакцию:

06-02-2023

Аннотация: В статье анализируются современные концепции искусственного интеллекта. Автор предлагает различать две основные: утопическую и реалистическую; в первой утверждается, что можно создать машины, мыслящие как человек и даже лучше его, во второй интеллект представлен как своеобразный психобиологический компьютер. Делается экскурс в историю развития искусственного интеллекта, в рамках которого рассматриваются три подхода, оказавших серьезное влияние на развитие этого явления: редукционалистский подход, психологический и компьютерный. Искусственный интеллект сравнивается с естественным, последний рассматривается как сложное семиотическое, культурное и социальное образование, предполагающее человеческое общение, развитие, филиацию идей и техники. Делается и обосновывается вывод, в соответствие с которым утопическая концепция искусственного интеллекта не может быть реализована, хотя усилия по ее реализации еще долго будут продолжаться. Одним из аргументов здесь выступают противоречия в этической трактовке искусственного интеллекта, которая обсуждается в форме виртуального диалога автора с конструктором современных роботов. Напротив, реалистическая концепция реализуется в форме перманентного проекта. В этой концепции практически отсутствуют антропологические претензии, а понятие интеллекта здесь редуцированное, т.е. под интеллектом понимается имитация только тех сложившихся интеллектуальных процессов, которые

можно промоделировать и алгоритмизировать, что является необходимым условием программирования.

Ключевые слова:

интеллект, мышление, сознание, робот, проект, реализация, программирование, моделирование, компьютер, утопия

Хотя у меня есть работы по философии техники (несколько книг и еще больше статей), в предыдущие годы тему искусственного интеллекта я не рассматривал. Но, поскольку искусственный интеллект имеет отношение к технике (конечно, не сводится к ней полностью), и о нем так много говорят и пишут, я решил, что необходимо все же эту тему продумать. Тем более что речь идет о важной тенденции современности. Стал читать и понял, что идеологи и теоретики искусственного интеллекта честно признаются, что сами не могут договориться, как понимать этот феномен. «Единого ответа на вопрос, чем занимается искусственный интеллект, – читаем мы в очень неплохой статье «Искусственный интеллект» в «Википедии», – не существует. Почти каждый автор, пишущий книгу об искусственном интеллекте, отталкивается в ней от какого-либо определения, рассматривая в его свете достижения этой науки» [\[4\]](#).

Точнее авторы указывают, что понимание искусственного интеллекта предполагает знание сущности и особенностей интеллекта и более широко творчества, а также наличие более-менее строгой, однозначной концепции, однако, соответствующие знания и концепцию в настоящее время нельзя считать удовлетворительными. «В философии не решён вопрос о природе и статусе человеческого интеллекта. Нет и точного критерия достижения компьютерами “разумности”, хотя на заре искусственного интеллекта был предложен ряд гипотез, например, тест Тьюринга или гипотеза Ньюэлла – Саймона... Природа человеческого творчества ещё менее изучена, чем природа интеллекта» [\[4\]](#).

Несмотря на большую критику, до сих пор популярна и серьезно воспринимается концепция искусственного интеллекта, в которой утверждается, что можно создать машины, мыслящие как человек и даже лучше его. При этом под машинами понимаются компьютеры и современные роботы, действующие на основе программ; и то и другое подводится под понятие современная технология. Будем эту концепцию называть «антропокомпьютерной». Вот всего лишь один пример. «Искусственный интеллект – это технология, а точнее направление современной науки, которое изучает способы обучить компьютер, роботизированную технику, аналитическую систему *разумно мыслить также как человек* ... Исследования в сфере ИИ ведутся путем изучения умственных способностей человека и переложения полученных результатов в поле деятельности компьютеров... Нейросеть – математическая модель, которая имитирует строение и функционирование нервных клеток живого организма. Соответственно в идеале – это самостоятельно обучаемая система» [\[5\]](#). (курсив наш. – В.Р.)

Только на первый взгляд, это одна концепция, фактически здесь их несколько. Ведь не одно и то же: обучение (самообучение), мозг, понимаемый как сеть нейронов, умственные способности индивида, их исследование, «переложение полученных результатов в поле деятельности компьютеров», психологическое истолкование мышления (понимание, способы решения задач, процессы анализа, синтеза, рефлексия и др.), компьютерная, в том числе вычислительная версия мышления – в науке все это разное, как по понятиям, так и дискурсам. Но в литературе по искусственному

интеллекту эти и сходные концепты идут «через и», легко объединяются в комплексы и определения. Примерно то же мы видим в уже цитированной статье «Искусственный интеллект»

«Проблема состоит в том, – писал еще в 1956 г. Джон Маккарти, – что пока мы не можем в целом определить, какие вычислительные процедуры мы хотим называть интеллектуальными. Мы понимаем некоторые механизмы интеллекта и не понимаем остальные. Поэтому под интеллектом в пределах этой науки понимается только вычислительная составляющая способности достигать целей в мире».

В то же время существует и точка зрения, согласно которой интеллект может быть только биологическим феноменом... Даются следующие определения искусственного интеллекта:

- Научное направление, в рамках которого ставятся и решаются задачи аппаратного или программного моделирования тех видов человеческой деятельности, которые традиционно считаются интеллектуальными.

- Свойство интеллектуальных систем выполнять функции (творческие), которые традиционно считаются прерогативой человека. При этом интеллектуальная система — это техническая или программная система, способная решать задачи, традиционно считающиеся творческими, принадлежащие конкретной предметной области, знания о которой хранятся в памяти такой системы. Структура интеллектуальной системы включает три основных блока — базу знаний, решатель и интеллектуальный интерфейс, позволяющий вести общение с ЭВМ без специальных программ для ввода данных.

- Направление в информатике и информационных технологиях, задачей которого является воссоздание с помощью вычислительных систем и иных искусственных устройств разумных рассуждений и действий.

- Способность системы правильно интерпретировать внешние данные, извлекать уроки из таких данных и использовать полученные знания для достижения конкретных целей и задач при помощи гибкой адаптации.

Одно из частных определений интеллекта, общее для человека и «машины», можно сформулировать так: «Интеллект — способность системы создавать в ходе самообучения программы (в первую очередь эвристические) для решения задач определённого класса сложности и решать эти задачи»» [\[4\]](#).

Подобный материал, а он типичный, заставляет поставить два принципиальных вопроса. Может быть, все же есть целостная концепция искусственного интеллекта, но она необычная, как сегодня говорят, гибридная, или другой вариант – она, по В.С. Степину, выполнена в рамках постнеклассической рациональности? Если же такой концепции нет, то правильно ли существование многих концепций искусственного интеллекта, а также как в этом случае объяснить подобную множественность? Например, Л.С. Выготский в 20-х годах прошлого столетия считал, что существование нескольких психологий, по-разному объясняющих психику человека, недопустимо, при этом он исходил из естественнонаучного понимания и идеала познания [\[13, с. 44-54\]](#). Однако я показываю, он не прав, поскольку, во-первых, психология без противоречий и других проблем не может быть подведена под указанный идеал, во-вторых, каждая версия психологии обслуживает и обеспечивает определенный тип личности, характерный для новоевропейской культуры (и таких типов, действительно, несколько) [\[13, с. 219-233\]](#). Но

можно ли сравнивать искусственный интеллект с типами новоевропейской личности? Подвесим пока ответы на эти вопросы и рассмотрим под влиянием каких подходов развивалась концепция искусственного интеллекта.

Первый подход (назовем его «редукционалистским») был задан программой реформирования логики, предполагавшей сведение ее к своего рода математике. Здесь предтечей был Лейбниц, который сформулировал задачу построения символического исчисления, позволяющего вместо рассуждений вычислять (вспомним и Маккарти, писавшего, что «под интеллектом в пределах этой науки понимается только вычислительная составляющая»). Лейбниц рассказывает, что «напал на ту достойную удивления мысль, что можно найти известный алфавит человеческих мыслей, и что, комбинируя буквы этого алфавита и анализируя составленные из них слова, можно, как все вывести, так и все обсудить<...>". "Всеобщая характеристика", понимаемая в логическом плане, представляет собой систему точно установленных знаков, посредством которых в логике и других дедуктивных науках должны обозначаться простые элементы объектов, составляющих предмет исследования данной науки. Эти знаки, во-первых, должны быть краткими и сжатыми по форме; они должны заключать максимум смысла в минимуме протяжения. Во-вторых, эти знаки должны изоморфно соответствовать обозначаемым ими понятиям, представлять простые идеи как можно более естественным способом. Сложные идеи будут представимы посредством соединений или сочетаний элементарных идей. На языке "всеобщей характеристики" абстрактные тезисы логики предстанут в виде наглядных правил, регулирующих действия с символами. Эти правила описывают формальные свойства знаковых преобразований и имеют своим источником привычки наглядного представления <...> единственное средство улучшить наши умозаключения, сделать их, как и у математиков, наглядными, так чтобы свои ошибки находить глазами, и, если среди людей возникнет спор, нужно сказать: "Посчитаем!"; тогда без особых формальностей можно будет увидеть, кто прав» (чит. по [\[16, с. 122\]](#)).

Как я показываю, путь от указанного замысла к его реализации занял несколько веков и предполагал: «1) истолкования понятий как математических функций (Фреге и Рассел), что позволило обозначать символами отдельные высказывания и понимать их как переменные, 2) сведение всех значений высказываний к двум основным – истинности и ложности (Фреге) и 3) создание на основе двух предыдущих изобретений *таблиц истинности* (своего рода порождающей системы), с помощью которых строится автономный язык символической логики (Витгенштейн); 4) построение на основе аксиоматического метода моделей логических систем» [\[10, с. 133\]](#) Но была и серьезная плата за редукцию мышления к математике: приведение обычного языка высказываний в соответствие с искусственным символической логики, ограничение символического дискурса только типичными случаями, математическая интерпретация событий, сведение функций истинности к двум – истинности и ложности [\[10, с. 138-139\]](#). Не осознавалась, не только редукция предметных смыслов к математическим, но и 1) различие логики как инструмента построения правил, позволяющих *непротиворечиво рассуждать*, и логики как инструмента обоснования, позволяющего организовать в единую систему (теорию) уже сложившиеся поля рассуждений, 2) противоречия между ощущением автономии символической логики и зависимости её от традиционной логики, 3) отсутствие философского осмысления новой дисциплины [\[10, с. 140-142\]](#).

Второй подход, оказавший серьезное влияние на концепцию искусственного интеллекта, – «психологический», в рамках которого понимание, мышление и творчество (соответственно, интеллект) были сведены к способностям индивида. В отличие от

философского истолкования этих феноменов, предполагавшего анализ более широкого целого, включавшего Других, Бога, культуру, сознание, опыт, ценности, рефлексия, развитие и ряд других факторов, психологи свели интеллект к психологическим процессам, что оказалось очень удобным для моделирования и программирования (это *третий подход* , собственно, «компьютерный»). Компьютерное моделирование и программирование тоже предполагает редукцию, но более сложную, не только к математическим построениям, но также к схемам, процессам, техническим конструкциям [\[11, с. 140-143\]](#). Отдельный вариант психологического подхода – «физиологический» («биологический»), здесь интеллект редуцируется к работе мозга и далее «нейронных сетей».

Указанные три влияния обусловили формирование второй концепции искусственного интеллекта – «реалистической», в соответствии с которой интеллект это своеобразный *психобиологический компьютер* . Вспомним: «интеллектуальная система – это техническая или программная система, способная решать задачи... знания о которой хранятся в памяти такой системы...включает три основных блока – базу знаний, решатель и интеллектуальный интерфейс, позволяющий вести общение с ЭВМ без специальных программ для ввода данных...направление в информатике и информационных технологиях, задачей которого является воссоздание с помощью вычислительных систем и иных искусственных устройств разумных рассуждений и действий».

Соответственно, первую, указанную выше, концепцию искусственного интеллекта можно назвать не только антропокомпьютерной, но и «утопической», и вот почему. Какие бы оговорки не делать, искусственный интеллект – это все такая техника, пусть и с включением антропологических составляющих (В отличие, скажем, от Интернета, который представляет собой планетарную антропо-социо-техническую форму жизни. [\[14\]](#)) А технику научить мыслить и творить как человека невозможно. Но чтобы понять последнее утверждение, нужно обсудить, что собой представляет «естественный интеллект» в отличие от искусственного. Ограничимся рассмотрением мышления и творчества человека, понимая, что в понятие естественного интеллекта входят и некоторые другие составляющие, например семиотическая и техническая филиация идей. Рассмотрим сначала один кейс с целью предоставить читателю материал, на который он мог бы опираться, уясняя, что такое, с точки зрения автора, естественный интеллект.

Диалог Платона в «Пире» вполне может считаться одним из первых примеров и творчества, и мышления, правда, еще только становящегося. Герои «Пира», в частности, дают два принципиально новых определения любви: это «поиск своей половины и стремление к целостности», а также «вынашивание духовных плодов – прекрасного, блага и бессмертия» [\[7, с. 100\]](#). Кстати, если следовать реалистической концепции искусственного интеллекта, то подобное представление любви можно истолковать процессуально и оно достаточно для моделирования, алгоритмизации и программирования. Однако анализ показывает, что эти определения указывают только на верхушку айсберга, его целое скрыто под водой. Целое можно понять, если реконструировать ситуацию и проблему, которые заставили Платона вводить новые представления о любви, учесть реакцию афинских граждан, привыкших считать, что любовь вызывают боги-любви (Афродита и Эрос), (а Платон исходил из того, что индивид должен сам выбирать, кого ему любить), объяснить, каким образом Платон вышел на новые представления о любви, и кто на него повлиял.

Я показываю, что Платон в «Пире» ориентировался на новую антропологическую

ситуацию (имел в виду потребности становящейся античной личности) и разрешал проблему невозможности для личности любить в старой социальной модели любви. Для этого он создавал новые схемы и определения любви, переключающие ее на личность и самостоятельный выбор, учитывал возможности понимания новых представлений, поэтому всячески принижает образ традиционной богини-любви, называя ее Афродитой вульгарной и простонародной, противопоставляя ей Афродиту небесную. Платон стремился мыслить диалектически, что предполагало выполнение ряда условий – именование любви, построение ее определений, построение схем, непротиворечивые рассуждения, доброжелательные вопросы и ответы, наконец, своего рода озарение (инсайт), позволяющее припомнить идею любви, которую душа созерцала до своего воплощения в земное тело человека. Сюда же входила рефлексия, так в «Федре», это следующий за «Пиром» диалог, Платон пишет, что давал любви определения, относил их к идее любви (что предполагало согласование этих определений) и это позволило мыслить любовь без противоречий. [\[9, с. 19-20; 8, с. 176\]](#) Не стоит думать, что Платон был первооткрывателем всего. Нет, он опирался на пифагорейское учение и реализовал программу построения мышления Парменида, а также ряд положений, высказанных Сократом. [\[9, с. 23-34\]](#)

Приведенный кейс – это всего лишь один пример, но и другие исследования мышления и творчества, например, творчества и мышления Галилея, подтверждают основные полученные при анализе «Пира» положения. А именно, целое, позволяющее рационально объяснить строение, функционирование и развитие естественного интеллекта, включает в себя не только процессы мышления и творчества, которые моделируются в реалистической концепции, но также коммуникацию, в рамках которой эти процессы разворачиваются, ситуации и проблемы, заставляющие создавать новые представления, механизмы разрешения этих проблем, реакцию на новые представления пользователей, коррекцию этих представлений, изменение условий, ситуаций и проблем, формы осознания и концептуализации. Все это обусловлено в общении людей, исторически и культурно. В создание новых форм и содержаний интеллектуальной деятельности делает вклад не только семиотика, язык, дискурсивные построения, но и личность с ее чувствами, отношениями, пониманием социальности.

Я не претендую в данном случае на приоритет, только на более объемную, развернутую картину, ряд указанных здесь характеристик естественного интеллекта в сравнении с искусственным были уже указаны в научной литературе. «Решение искусственного интеллекта, – пишет, например, Соня Шпильберг, – выглядит как перетасовывание и выбор из загруженных вариантов “опыта”, тогда как человек анализирует и выбирает наилучший вариант не только эмпирически, но и эмоционально. Хомо сапиенс также обладает ассоциативным мышлением, поэтому его решение онтологически верное, тогда как у ИИ нет своего феноменального опыта – то есть представления о логике развития культуры. Поэтому его решение может шокировать нас «неэтичностью» [\[15\]](#).

Можно сослаться и на книгу Х. Дрейфуса «Чего не могут машины». «Основным предположением, лежащим в основе исследований в области искусственного интеллекта, является предположение, согласно которому человек действует подобно устройству для символической обработки информации. В свою очередь, данное предположение распадается на четыре допущения, которые Дрейфус последовательно опровергает: 1. *Психологическое допущение*. «Мышление можно рассматривать как переработку информации, заданной в бинарном коде, причем переработка происходит в соответствии с некоторыми формальными правилами» [\[2, с. 105\]](#). 2. *Эпистемологическое допущение*. «Все знания могут быть формализованы, то есть все, что может быть понято,

может быть выражено в терминах логических отношений, точнее, в терминах булевых функций – логического исчисления, задающего правила обращения с информацией, заданной в двоичном коде» [\[2, с. 105\]](#). 3. *Онтологическое допущение*. «Машинная модель мышления предполагает, что все сведения о мире, все, что составляет основу разумного поведения, должно в принципе допускать анализ в терминах множества элементов, безразличных к ситуациям» [\[2, с. 105\]](#). Таким образом, все происходящее в мире можно представить в виде множества фактов, каждый из которых логически не зависит от остальных. 4. *Биологическое допущение*. «На некотором уровне – обычно полагают, что на уровне нейронов – операции по переработке информации носят дискретный характер и происходят на основе некоторого биологического эквивалента переключательных схем» [\[2, с. 105\]](#).

Психологическое допущение оказывается возможным лишь на основании смешения обычного смысла слова «информация со специальным значением, имеющим место в кибернетической теории информации. Опровержение психологического допущения, однако, нисколько не мешает продолжать существовать допущению эпистемологическому. В данном случае имеет место неоправданное распространение методологии естественных наук на область психических явлений. Однако, пишет Дрейфус, «полное опровержение эпистемологического допущения потребовало бы доказательства того, что мир принципиально не может быть проанализирован в терминах четко определенных данных» [\[2, с. 164\]](#), а подобное доказательство одновременно является опровержением онтологического допущения. По мнению Дрейфуса, онтологическое допущение в большей степени не согласуется с нашим опытом, Оно явилось следствием необходимости понимания мира и умения управлять им, принуждавшим западную традицию к упрощению действительности, в то время как она гораздо сложнее. Распространению этого стремления к упрощению способствовали успехи в физике. Однако, хотя в рамках физической теории мир и может быть представлен в виде совокупности множества атомарных фактов, однако, данное представление, выходя за пределы данной теории, плохо согласуется с нашим опытом. То, что мир может быть «разделен» на атомарные факты, еще не означает, что, попытавшись «собрать» из этих элементов целостную картину мира, мы действительно ее получим. В рамках биологического же допущения лежит интерпретация нейронного импульса в качестве единицы информации, циркулирующей в мозгу, подобно машинному биту. Возможность подобной интерпретации, однако, опровергается Дрейфусом, что отменяет и биологическое допущение.

В силу того, что опровергнутые Дрейфусом допущения лежали в основании предположения о том, что человек действует подобно устройству для символической обработки информации, то опровергнутым оказывается как данное предположение, так и вся концепция классического ИИ, для которого оно является фундаментальным. Все те серьезные трудности, которые встали перед разработками в области ИИ, уже не могут быть решены путем увеличения быстродействия и объемов памяти. Опровержение же вышеуказанного фундаментального предположения ИИ, не оставляет никаких оснований для уверенности в возможности моделирования человеческого поведения. А в силу этого излишним будет даже само появление вопроса о возможности прохождения машиной теста Тьюринга» [\[6, с. 129-131\]](#). Действительно, возможность спутать поведение человека с действием машины очень далека от указанных Дрейфусом и нами различий естественного и искусственного интеллекта.

Иногда ссылаются на то, «что в мае 1997 года компьютер Deep Blue в матче из 6 партий

обыграл чемпиона мира по шахматам Гарри Каспарова. Но такое отождествление, на наш взгляд, не совсем законно. Во-первых, потому, что мышление человека – феномен психологический, семиотический и социальный (мозг только субстрат мышления), а не машинный. Во-вторых, сторонники компьютерного мышления не понимают, что программы Deep Blue включали в себя не только переборы и сравнения ходов и их следствий, но также обобщение лучших шахматных партий и даже практику подготовки к шахматным чемпионатам, которую разработал и успешно использовал наш чемпион мира М. Ботвинник. В этом смысле Каспаров сражался не с машиной, а с воплощенной в машинной среде шахматной культурой; культура же на порядок мощнее индивида» [\[12, с. 264\]](#). И обыграла Каспарова усиленная машиной шахматная культура.

Естественный интеллект – образование принципиально антропологическое, историческое и культурное. Поэтому не только законосообразное, но и сингулярное. С этой точки зрения, выражение «обучить компьютер, роботизированную технику, аналитическую систему *разумно мыслить также как человек*» является неосмысленным, а как задача утопической. Искусственный интеллект – образование принципиально программируемое и конструируемое, выступающее в функции средства. Но и «социального тела» человека, поскольку начинает задавать его новые возможности, способности и компетенции.

Альберт Ефимов в своей диссертации попытался сблизить все три версии интеллекта – естественную, реалистическую и утопическую. С одной стороны, он соглашается с критикой утопической версии, говоря, что Дж. Сёрл и Т. Нагель «критиковали машинный функционализм за игнорирование эпистемологических вопросов ценностно-смыслового плана (подчеркивая отсутствие понимания компьютером смысла воспринимаемой информации), за крайне упрощенную трактовку понятия естественного интеллекта. Этими философами и их последователями были высказаны принципиальные возражения против создания машинного интеллекта, имеющего возможности, сопоставимые с мышлением человека» [\[3, с. 15-16\]](#). С другой стороны, Ефимов предлагает концепт «Общего искусственного интеллекта», который в основном близок к концепции реалистического искусственного интеллекта. Однако, вероятно, находясь под влиянием своего научного руководителя профессора Давида Дубровского, основные работы которого посвящены изучению сознания, Ефимов, с третьей стороны, обогащает свой концепт антропологическими обертонами, утверждая, что можно создать такие программы и технические устройства (речь идет о роботах), которые воспроизведут человеческое сознание, эмоции, мышление.

«Современная специализация искусственного интеллекта и его новейшие результаты, – пишет Ефимов, – обнаруживают существенный разрыв между теоретическими исследованиями в области философии и методологии искусственного интеллекта и практическими усилиями исследователей создавать интеллектуальные машины (роботов), которые *будут обладать способностями, не уступающими человеческим <...>* Данный подход значительно увеличивает сложность задачи вследствие универсализма разрабатываемых систем. Он требует более широкого использования для создания соответствующих когнитивных архитектур *результатов современных феноменологических исследований динамической структуры явлений сознания* (Д.И. Дубровский). В определенной степени такой подход отражает процесс формирования психики в ходе биологической эволюции...

«Робот «Э.ЛЕНА» является виртуальным гуманоидом, который обладает виртуальным гуманоидным телом, наделённым *мимикой*, может использовать естественный язык, *полностью поддерживая русский язык* (виземы, фонемы), является автономным (не

требующим действий оператора при выполнении базовых функций), *обладающим собственной персоной* (в зависимости от персоны чат-бота) и имеющим первичную возможность для обучения изменениям внешнего мира (в роботе реализовано распознавание лиц собеседников робота). Впервые в научной литературе представлена архитектура подобного робота. Уникальность её заключается в том, что при её построении использовался не только ряд существенных характеристик *субъективной реальности человека (опиравшихся на современные исследования феноменологии сознания)* , но и сложный комплекс внешних проявлений сознательной деятельности человека, исполняющего роль ведущего на телевидении: мимика, выражение глаз, жесты, интонации голоса и т. д...Моделирование и программирование таких свойств роботов, которые бы отвечали нашим юридическим и этическим принципам, полностью исключали бы их агрессивность и «недружественные интенции», вероятно, потребует создания «виртуальных людей» – программ, которые будут *эмоционально отождествлять себя с людьми, обладать чертами человеческого самосознания и самости* ». [\[3, с. 7, 19-21\]](#) (курсив наш. – В.Р.)

Работа Ефимова – пионерская и весьма содержательная (защищена была в Институте философии блестяще), но трудно согласиться с вводимыми им антропологическими обертонами; человеческие эмоции, смыслы, сознание, самость – все это из естественного интеллекта, они, на мой взгляд, не могут быть идентично воссозданы в форме искусственного интеллекта.

На первый взгляд, мечта идеологов и конструкторов искусственного интеллекта создать технику (роботов, компьютеры), мыслящую как человек, очень понятная и ясная, но только на первый взгляд. Если продумывать ее, выявляются серьезные противоречия, что видно, например, по этической проблематике. Представим для пояснения такой виртуальный разговор между автором и идеологом-конструктором (кратко *И-К*), уровня примерно Ефимова.

– *Автор* . Почему вы думаете, что идеальный робот, например, превосходящий в психологическом отношении «Э.ЛЕНУ», может, вдруг, пойти против человека и навредить ему, что является бесконечной темой многих рассказов и фильмов?

– *И-К* . Как же, ведь он может мыслить и вообще вести себя как человек, только человек искусственный, созданный нами. А некоторые люди совершают преступление, идут против общества, воруют, насилуют, убивают. Если робот мыслит и ведет себя как человек, то почему он тоже не может свихнуться?

– *Автор* . Зачем тогда создавать робота, который может, как вы говорите, свихнуться и нарушить первые два закона робототехники?

– *И-К* . А как иначе создать робота с нужными компетенциями, например, чтобы он помогал человеку и общался с ним. А свихнуться робот может не потому, что такая возможность в нем предусмотрена, а в силу сложности и неясности для нас до конца его строения, ведь, скажем, вы работаете на компьютере или разговариваете по мобильному телефону, не зная точно, как они устроены.

– *Автор* . Я предпочитаю общаться не с техникой, а с близкими или интересными мне людьми. Если вы создаете для робота искусственный интеллект, идентичный человеческому, то вам придется смириться с мыслью, что робот может навредить человеку, ведь сами пишете про самосознание и самость «Э.ЛЕНУ». Если уж самость, будьте готовы к незапрограммированным поступкам.

– *И-К* . Вот для этого случая и предназначены законы робототехники. Возможно, они несовершенны, и лучше говорить об этике и нравственности искусственного интеллекта, чтобы на их основе сформулировать более эффективные законы.

– *Автор* . Странно звучит «этика и нравственность для техники», я всегда думал, что они относятся только к человеку. Кроме того, не будут ли ваши роботы напоминать рабов Древнего мира, для них тоже существовали правила и законы, однако, как известно, рабы периодически восставали, вспомните хотя бы Спартака. Не восстанут ли и ваши роботы против человека?

– *И-К* . Не исключено, нужно продумать эту возможность и ввести в искусственный интеллект роботов дополнительные программы. Кроме того, у всякой новой техники со временем обнаруживаются негативные последствия. Всего сразу не предусмотреть.

– *Автор* . Так ваши роботы – это искусственные люди или техника? Если люди, то для них, чтобы они вели себя как люди, а не бездушная техника, нужны не только законы, но и человеческая среда – другие люди, общение, ощущение своей конечности и ограниченности (люди не боги), изменения, история, настоящее и будущее и многое-многое другое, что присуще естественному интеллекту. Но если это техника, пусть и очень сложная и внешне похожая на людей, то не нужно стремиться, чтобы она обладала самосознанием и самостью, напротив, помнить, что это всего лишь техника, что нельзя ослаблять или убирать вообще управление. Ну да, в современных роботах человек управляет посредством специальных программ, а не непосредственно с помощью пультов управления, но суть от этого не меняется – управляет роботом не он сам, а человек. В конечном счете, нормальная программа это тоже орудие (естественно, более сложное), тоже техника, но семиотическая.

Невозможность реализовать утопическую концепцию искусственного интеллекта не означает сомнение в эффективности и перспективах реалистической концепции. В этой концепции практически отсутствуют антропологические претензии, а понятие интеллекта здесь редуцированное, т.е. под интеллектом понимается имитация только тех сложившихся интеллектуальных процессов, которые можно промоделировать и алгоритмизировать, что является необходимым условием программирования [\[11, с. 140-144\]](#).

С развитием искусственного интеллекта и Интернета интеллект стал быстро меняться. Помимо естественного интеллекта и развития (они сохраняются во многих местах в условиях решения традиционных задач) все большее влияние приобретают, с одной стороны, программирование, с другой – мышление, творчество и филиация идей в интернет-среде.

Хотя утопическая концепция искусственного интеллекта не может быть реализована, она заражает разработчиков и пользователей, воодушевляя их работать в целях осуществления подобной концепции. Что означает длительное существование утопической концепции искусственного интеллекта?

Вообще мы имеем дело с гуманитарной и социальной ситуацией, а следовательно, что отмечал З. Бауман, с множеством интерпретаций и популятивностью (множественностью) существования феномена (в данном случае, искусственного интеллекта) [\[11\]](#). На мой взгляд, например, трактовка мышления и творчества в биологических терминах ведет к эпистемологическому тупику, но поскольку есть научное сообщество, разделяющее биологические взгляды на человека и социальность, попытки не просто создать

нейронный компьютер, но и свести интеллект к работе нейронов будут продолжаться. Будут продолжаться и другие истолкования искусственного интеллекта и работы по созданию соответствующих типов техники.

«В последнее время, – отмечает Ефимов, – несмотря на успешное развитие науки и технологий в области искусственного интеллекта, у исследователей и инженеров нет ясного ответа на вопрос, как развивать технологии в перспективе следующих десятилетий. Причина в том, что теоретико-методологическая и философская основы такого планирования разработаны весьма слабо. Целиком сохраняется актуальность мысли К. Ясперса, что “вся дальнейшая судьба человека зависит от того способа, посредством которого он подчинит себе последствия технического развития и их влияние на его жизнь”.

Наблюдаются постоянные расхождения в прогнозах создания Общего искусственного интеллекта. Вычислительные мощности, которые казались А. Тьюрингу достаточными для создания “мыслящей машины”, давно достигнуты, возможности создания, хранения и передачи информации в интернете превышают самые смелые мечты полувековой давности. Однако всё это лишь частные результаты, далекие от создания Общего интеллекта» [\[3, с. 5\]](#).

Если под Общим искусственным интеллектом понимать утопическую концепцию, то этот проект принципиально не может быть осуществлен, но если реалистическую, то он уже реализован и в качестве «перманентного проекта» развивается. При этом нетрудно предсказать, проект реалистического искусственного интеллекта будет еще не раз уточняться и даже, возможно, пересматриваться, однако, все же в процессе его осуществления.

Библиография

1. Бауман З. Мыслить социологически. М., 1996. М.: Аспект-Пресс, 1996. 254 с.
2. Дрейфус Х. Чего не могут вычислительные машины. Критика искусственного разума. М.: Прогресс, 1978.
3. Ефимов А.Р. Философско-методологические основы посттьюринговой интеллектуальной робототехники. Автореферат дис. на соиск. уч. степени. М, 2021. 22 с.
4. Искусственный интеллект https://ru.wikipedia.org/wiki/Искусственный_интеллект
5. Искусственный интеллект <https://www.calltouch.ru/blog/glossary/iskusstvennyj-intellekt/>
6. Кляшторный А.С. Критика концепции «сильного» ИИ в работе Х.Дрейфуса «Чего не могут вычислительные машины?» chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://elar.urfu.ru/bitstream/10995/93892/1/978-5-7525-4025-7_2011_037.pdf
7. Платон. Пир. Соб. соч. в 4 т. Т. 2. М.: Мысль, 1993. С. 81-135.
8. Платон. Федр. Соб. соч. в 4 т. Т. 2. М.: Мысль, 1993. С. 135-192.
9. Розин В.М. «Пир» Платона: Новая реконструкция и некоторые реминисценции в философии и культуре. М.: ЛЕНАНД, 2015. 200 с.
10. Розин В.М. Логика и методология: От «Аналитик» Аристотеля к «Логико-философскому трактату» Л. Витгенштейна. М.: ЛЕНАНД, 2014, 272 с.
11. Розин В.М. Проектирование и программирование: Методологическое исследование. Замысел. Разработка. Реализация. Исторический и социальный контекст. М.:

ЛЕНАНД, 2018. 160 с.

12. Розин В.М. Техника и технология: от каменных орудий до Интернета и роботов. Йошкар-Ола: ПГТУ, 2016. 280 с.
13. Розин В.М. Методологические проблемы психологии. 2-е изд. М.: Юрайт, 2019. 358 с.
14. Розин В.М., Голубкова Л.Г. Интернет как техническое сооружение и живой планетарный организм /Розин В.М. Техника и технология: от каменных орудий до Интернета и роботов. Йошкар-Ола: ПГТУ, 2016. С. 235-247.
15. Семь смертных грехов искусственного интеллекта
<https://trends.rbc.ru/trends/social/5eb299089a79476e9fd77f5c>
16. Стяжкин Н.И. Формирование математической логики. М.: Наука, 1967. 508 с.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Автор рецензируемой статьи, обращаясь к рассмотрению одной из самых актуальных и вызывающих широкий интерес проблем – проблемы искусственного интеллекта, указывает на зависимость характера оценки разработок в этой сфере от того или иного понимания «интеллекта» как такового, поскольку только последний способен выступать в качестве критерия успешности любых конкретных разработок. Однако интеллект (в непосредственном смысле) представляет собой лишь одну из сторон творческой деятельности человека, о приближении к которой, по существу, и говорят исследователи, когда ставят проблему создания искусственного интеллекта. А определение творчества и деятельности, по-видимому, оказывается ещё более трудной задачей. Анализируя состояние разработанности этих проблем, автор статьи справедливо указывает на то, что носители «оптимистических» представлений о возможностях искусственного интеллекта, как правило, не дифференцируют в должной степени характеристики тех задач, которые сегодня уже решают компьютеры или роботы. Основное теоретическое содержание статьи представлено в виде характеристик подходов, предлагавшихся для решения проблемы искусственного интеллекта. Один из них (который автор именует редукционистским в связи с присущим ему сведением логики к «исчислению», то есть своего рода математике) восходит к идее «всеобщей характеристики» Лейбница (заметим, что разработки Лейбница оказали и «обратное» влияние и на математику, положив начало так называемому «логицизму» в математике). Ему противопоставляется психологический подход, основывающийся на представлении о способностях индивида как критерии понимания интеллекта и творчества (и здесь снова необходимо указать, что понимание души как «системы способностей» также уходит вглубь веков, например, оно доминировало в философской психологии 18 века). От него пролегает путь и к собственно «компьютерному» подходу, удобному для решения задач программирования и моделирования. На основании этих трёх подходов и сформировалась, утверждает автор, так называемая «реалистическая» концепция искусственного интеллекта, в которой последний понимается как «психобиологический компьютер». Насколько можно понять последующие примеры автора, он склонен считать, что «реалистичность», о которой в данном случае идёт речь, указывает на принципиальную и неискоренимую зависимость «машинного» интеллекта от интеллекта «человеческого»; «утопичность», в противоположность «реалистичности», начинается с того, что эта зависимость игнорируется: «Естественный интеллект – образование

принципиально антропологическое, историческое и культурное. Поэтому не только законосообразное, но и сингулярное. С этой точки зрения, выражение «обучить компьютер, роботизированную технику, аналитическую систему разумно мыслить также как человек» является неосмысленным, а как задача утопической. Искусственный интеллект – образование принципиально программируемое и конструируемое, выступающее в функции средства, и т.д.». «Если под Общим искусственным интеллектом понимать утопическую концепцию, – резюмирует свои размышления автор, – то этот проект принципиально не может быть осуществлен, но если реалистическую, то он уже реализован и в качестве «перманентного проекта» развивается». Не вызывает сомнения, что рецензируемая статья вызовет интерес читателей, однако, до публикации необходимо исправить опечатки (например, уже в первом предложении), имеются также и некоторые погрешности в пунктуации. Рекомендую статью к публикации в научном журнале.