

Исчислимо ли бытие человека, или Антропология искусственного интеллекта. Методологический аспект*

С. А. Смирнов

Институт философии и права СО РАН,
Российская Федерация, 630090, Новосибирск, ул. Николаева, 8

Для цитирования: Смирнов С. А. Исчислимо ли бытие человека, или Антропология искусственного интеллекта. Методологический аспект // Вестник Санкт-Петербургского университета. Философия и конфликтология. 2023. Т. 39. Вып. 3. С. 478–491.
<https://doi.org/10.21638/spbu17.2023.306>

В статье представлен анализ антропологических оснований доминирующих моделей, связанных с разработкой искусственного интеллекта. Делается вывод, что базовым, лежащим в основании данных разработок, является принцип вычислимости (исчислимости), введенный со времен первых разработок искусственно интеллекта, включая работы А. Тьюринга. Показывается, что этот принцип связан с определенной онтологической установкой, описанной в том числе в работах М. Хайдеггера, связанной с отношением к сущему и человеку в том числе как к потребляемому и в пределе уничтожаемому сущему. Принципиальная исчислимость сущего нужна для того, чтобы его употребить. Это, в свою очередь, ведет к тому, что к технике, в том числе к «умной машине», человек относится как к устройству, создаваемому по логике имитации и копирования. Показывается, что стратегия копирования при разработке искусственного интеллекта является весьма неэффективной и ущербной. В отличие от названного принципа вычислимости предлагается принцип опосредования, разработанный в свое время культурно-исторической психологией. В ее рамках само представление о мышлении и интеллектуальной операции строится не в категориях активности мозга, а в категориях предметного действия, опосредованного знаком и речью. Следствием этого принципа опосредования выступает и иная стратегия разработки искусственного интеллекта — разработка умных цифровых помощников для человека. Но для этой стратегии необходимо сначала перестраивать существующие сферы жизнедеятельности (как то: здравоохранение, образование, управление). Поэтому разработка искусственного интеллекта не является сугубо технологической проблемой. Она затрагивает глубинные устои бытия человека, которое не может быть исчислено.

Ключевые слова: искусственный интеллект, человек, антропологические основания, принцип вычислимости, принцип опосредования, сфера жизнедеятельности.

* Исследование выполнено за счет гранта Российского научного фонда, проект № 21-18-00103, <https://rscf.ru/project/21-18-00103/>

...Предложение «Машина мыслит (воспринимает, жаждет)» кажется чем-то бессмысленным. Как если бы спросили: «Есть ли у числа 3 цвет?»

Л. Витгенштейн

Введение: вызовы и допущения

Искусственный интеллект исследователи периодически называют последним изобретением человечества. Слово «последний» в данном случае употребляется в двойственном смысле, оно может означать: 1) что искусственный интеллект — венец человеческого творения, свидетельствующий о высшем уровне научной и технологической мысли, или 2) что своим творением человек сам себя поставил под вопрос, бросил самому себе вызов, столкнувшись с тем, что его же собственное изобретение впоследствии может его самого, создателя, убить, т. е. уничтожить как существо, заняв его место.

Мы не намерены обсуждать апокалиптические сценарии, связанные с грядущим «восстанием машин». Нам бы хотелось обсудить проблему искусственного интеллекта (далее — ИИ) не в технологическом, инженерном или узконаучном смысле, а в категориях антропологических оснований.

Какие антропологические основания закладываются в представления разработчиков ИИ? Что они понимают под мышлением человека, его интеллектом, прежде чем приступить к разработкам в сфере ИИ? Как они представляют себе человеческую психику, какие ее модели строят? Ведь сначала они как-то полагают себе мышление человека, а уже потом приступают к разработке ИИ, допуская, что машину можно научить мыслить.

Для начала заметим, что при любых допущениях вопрос, относящийся к разработке и внедрению ИИ, связан действительно с радикальными вызовами для человека, ставящими его самого в ситуацию онтологического испытания, испытания собственными границами. Выделим здесь два момента.

1. ИИ впервые поставил под вопрос уникальность и исключительность человека [1, с. 321]. Впервые творение человека, ИИ, может делать что-то лучше своего создателя: считать, прогнозировать, обрабатывать данные, делать диагностику, играть в шахматы и игру Go и т. д.

2. Приступив к разработке моделей ИИ, человек получил реальную возможность посмотреть на себя в зеркало. Он фактически в лице ИИ получил цифрового двойника, свое отражение. Строя модели ИИ, своего двойника, человек получил возможность выстраивать мысленный эксперимент над самим собой, а значит, ставить под вопрос себя как существо, и, как следствие, он может изменить понимание себя, переосмысляя, что значит быть собой, человеком.

В таком случае все зависит от человека-создателя — какими средствами, представлениями о себе самом он пользуется, выстраивая эти зеркальные модели самого себя? Если модельные копии неадекватны оригиналу, то значит, неадекватны прежде всего его представления о себе самом, выставляющие его всякий раз в редуцированном, участненном, урезанном виде. В связи с вышесказанным важно понять, какие принципиальные онтологические и методологические допущения делает автор-разработчик ИИ, какие критерии разумности, осмысленности, рациональности он вводит, строя те или иные модели ИИ.

Вычислимость как принцип

Со времен А. Тьюринга (A. Turing) были заложены основополагающие принципы создания «умных машин», согласно которым эти машины не должны ошибаться, они должны быть точными в вычислениях и их действия должны строиться на основе вычислений согласно пошаговым инструкциям. Сам принцип следования пошаговой инструкции (алгоритму) выводится из иного, более рамочного положения: *мышление есть вычисление*, которое осуществляет человеческий мозг. А само мышление локализовано в человеческом мозге. Тайна мышления заложена именно в активности мозга. Она там скрыта, полагали и полагают разработчики ИИ. Поэтому они десятки лет ставили перед собой главную задачу: построить модель работы человеческого мозга¹. Собственно, с тех пор и стали говорить о «думающих машинах», «умных машинах», введя метафору ИИ как искусственного мозга.

Тем самым укоренилось допущение, согласно которому мышление есть счет, вычисление, исчисление сущего, выраженное в математических моделях, на языке которых можно описать даже и такие тонкие материи, как совесть, сострадание, нравственность т. д. Именно поэтому стали делать ставку на быстрый счет, на быстрое действие. Машина должна уметь быстро считать, перебирая миллиарды вариантов в секунду. Поэтому быстродействующий компьютер в итоге обыграл человека сначала в шахматы, потом в игру Go².

Хотя сам А. Тьюринг ставил задачу шире. Он хотел создать не компьютер. Сначала он хотел решить математическую задачу, так называемую проблему разрешения Гильберта: разработать пошаговую процедуру, определяющую истинность или ложность любого отдельно взятого математического выражения. Если таковая процедура возможна, то возможен в таком случае (в принципе!) искусственный интеллект, способный дать ответ на любое логическое выражение. В основании этого подхода лежит принципиальная допустимость логико-математического выражения любого сущего, включая и человеческое мышление. Таковая процедура стала называться далее алгоритмом. И далее Тьюринг предложил устройство, «машину Тьюринга», которая показывала бы эту процедуру в действии, впоследствии эта машина стала называться компьютером, машиной, действующей согласно пошаговой процедуре по определенным правилам.

В дальнейшем при разработке всех поисковых систем в интернете, при организации интернет-торговли этот принцип и был заложен — машина обучалась за счет быстрого действия, включая новые данные в свой алгоритм, и могла уже предугадывать действия многих клиентов. На этом строится любой маркетинг. Стоит кому-то в интернете в поисковике задать какой-то поиск, ему тут же предложат много вариантов. И в следующий раз у него на экране монитора будут всплывать

¹ А. Тьюринг был уверен, что мозг ребенка можно смоделировать на компьютере. Именно потому, что его активность понималась им как набор вычислительных операций, которые можно разложить на последовательность шагов (программу).

² Заметим, что программа смогла обыграть человека именно за счет скорости вычислений, которые позволяли просчитывать 200 миллионов шахматных комбинаций в секунду. Но такая стратегия при разработке ИИ — тупиковая. Она энергозатратна и не дает ответа на главный вопрос: что есть мышление как субъектное действие? Ответа при такой модели нет и быть не может. Пока ИИ таковым субъектом и не выступает. ИИ лишен субъектности, он есть искусственная математическая модель с заложенными в нее параметрами и действиями по алгоритму.

окна с разными предложениями (от женского белья до дорогих машин и домов). Тем самым человек в целом рассматривается как экономическая, точнее, покупающая единица, которую можно подсадить на иглу рынка. Программа, с помощью которой осуществляются интернет-продажи, выступает в роли самообучающегося цифрового торгового агента, изучающего разные группы покупателей и ловящего их на разных капризах и желаниях. Машина обучается обобщению по схеме статистической обработки больших массивов данных, будь то онлайн-знакомства или игра в шахматы. В основании всех разработок современного ИИ лежит обработка больших данных и так называемое машинное обучение нейросети, происходящее путем проб и ошибок, т. е. принцип вычислимости [2, с. 18].

Надо заметить, что этот принцип в качестве методологического базового основания был введен в философию раньше. В конце 20-х годов прошлого века между представителями Венского кружка, отстаивавшими идеи логического позитивизма, с одной стороны, и представителями метафизики, главной фигурой коих был признан М. Хайдеггер (M. Heidegger), с другой, шел интенсивный спор [3, с. 352]. Внешне он выглядел как спор о предмете философии, ее будущем. Р. Карнап (P. R. Carnap) и другие обосновывали свою позицию тем, что философия должна быть логически и математически обоснованной наукой, строиться на логических основаниях. А коль скоро метафизика, ее основные понятия не могут быть логически ни обоснованы, ни опровергнуты, то они не имеют смысла, не являются научными. И поэтому философия как мышление о бытии не может быть основой для выстраивания будущей философии.

В основании спора сторонами закладывались разные онтологические установки относительно того, что есть сущее. Для Р. Карнапа сущее сводилось к эмпирической реальности, реальности существования эмпирических индивидов и натуральных объектов, которые могут быть *описаны, исчислены, подсчитаны, спрогнозированы*. Такая реальность должна быть понятной, данной в процедурах описания и расчета. В сущем не должно быть тайны. Все сущее должно быть научно, т. е. логически, экспериментально, опытно, обосновано и просчитано, исчислено. В основание сущего тем самым закладывался *принцип вычислимости*, т. е. расчета, охвата, ухвата. Сущее должно быть схвачено, просчитано и представлено как готовый объект. Равно как и сам человек как объект должен быть просчитан и схвачен.

М. Хайдеггер пытался показать, что в логике его оппонентов происходит подмена, замена онтологического поиска математическим доказательством, а истине отведена роль просчитанной достоверности, в силу чего происходит редукция, сужение самого предмета разговора. В результате сама философия превращается в «грамматическое предприятие», а философствование — в логические доказательства [3, с. 46–47]. Концепт бытия редуцируется к приставке «есть» как связке в суждении. Сущее исчезает, остается суждение. Спор идет о суждениях о предмете, а самого предмета уже и нет. Но логика не может справиться с онтологическим вопрошанием, она «расплывается в водовороте более изначального вопрошания» [4, с. 37].

В этом споре важна базовая интенция. Онтологическая установка, носителями которой были представители позитивизма, направленная на познание и логическое обоснование мира и человека, предполагает не просто сам по себе расчет мира, а дальнейшее его, мира и человека, употребление, в пределе — уничтоже-

ние в угоду выгоде, прибыли, капитализации. Этим словом Хайдеггер не употреблял, но основание обозначил: «Всякий расчет сводит исчислимое к расчисленному, чтобы употребить его в последующих расчетах... Это потребляющее употребление сущего выдает истребляющую природу расчета» [4, с. 71]. В пределе сущее потребляется ради полного расходования. Потребляется также и человек как сырье: «Потребление всех материалов, включая сырье человек, для технического производства неограниченной возможности изготовления всего поэтапно обусловлено полной пустотой, в которой взвешено сущее, материал для действительного. Эта пустота требует заполнения» [4, с. 252–253].

Но пустоты бытия никогда не заполнить полностью только сущего, поэтому возникает необходимость нарастания «обесцеленной деятельности» (сиречь *имитации и симуляции*). В этом случае именно техника подходит для обеспечения этого недостатка. Техника выступает заместителем, эрзацем сущего, в пределе — протезом для человека, подпирющим его собственную онтологическую нехватку. Человек, растерявший себя, привыкает к этому протезу, ему в его бытийной оставленности нравится замещать, заменять себя техникой-протезом. И чем дальше, тем больше этот протез становится все более функционально нагруженным, в пределе — «умным», т. е. собственно искомым искусственным интеллектом.

Исчисляемое и потребляемое сущее, будучи исчислено, становится «представленным», удобным для потребления и «более доступным» [5, с. 605]. Но в кажущейся доступности потребления это сущее на самом деле начинает сворачиваться, поскольку в стремлении потребления его всегда мало, его всегда не хватает. Точнее, человеку всегда не хватает собственного бытия, он в погоне за исчисленным и потребляемым, доступным сущим всегда остается голодным, испытывая онтологическую нехватку, попадая в яму хронического дефицита, из которой чисто технически не выбраться, полагал М. Хайдеггер.

Тем самым задается онтологическое основание для вполне определенной стратегии по разработке ИИ, построенной на исчислении и употреблении. В настоящее время именно эта стратегия доминирует при разработках разных моделей ИИ. Мы не говорим о ней оценочно, негативно или позитивно. Мы говорим о том, что в ее основании заложена вполне определенная онтологическая установка, направленная на исчисление и, соответственно, редукцию человека³.

За этим и стоит вполне определенное представление об ИИ. Приведем примеры определений ИИ, данные разными авторами и опубликованные в разных документах.

³ В философии математики вышеназванные вопросы связаны с проблемой вычислимости и невычислимости: «...вычислимость предполагает редукцию в описании, которая избегает или выводит за скобки алгоритмически несжимаемые фрагменты вселенной» [6, с. 11]. Допускается при этом, что «подлинная модель вселенной включает невычислимые элементы», с которыми как раз имеют дело неточные науки, которые потому и неточные, поскольку «сфера их внимания не подлежит алгоритмической сжимаемости». Последнее, т. е. феномен алгоритмической сжимаемости, становится основой для «пошагового процесса улучшения описания», т. е. операции вычисления в смысле Тьюринга [6, с. 10]. При этом математически доказывается, что «никакая компьютерная программа, которая занимается доказательством, не может достичь того, что достигается математиком, который занят доказательством в процессе математического мышления» [6, с. 443]. Мышление в принципе носит не компьютерный характер.

1. «ИИ — комплекс технологических решений, позволяющий *имитировать когнитивные функции человека* (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые как минимум с результатами интеллектуальной деятельности человека» [7].

2. «Направление исследований в современной компьютерной науке, целью которого является *имитация и усиление интеллектуальной деятельности человека* посредством компьютерных систем» [8, с. 316].

3. «В настоящее время к ИИ фактически относят различные программные системы и применяемые в них методы и алгоритмы, главной особенностью которых является способность решать интеллектуальные задачи *так, как это делал бы размышляющий над их решением человек*. К числу наиболее популярных направлений применения ИИ относятся прогнозирование различных ситуаций, оценка цифровой информации и выведение из нее заключений, анализ различных данных с поиском скрытых закономерностей» [9, с. 81].

4. «ИИ — теория и реализация компьютерных систем, способных *копировать интеллектуальное поведение человека*, выполнять задачи, обычно требующие человеческого интеллекта». Такие, как визуальное восприятие, распознавание речи, принятие решений и перевод с одного языка на другой [10].

5. «ИИ — способность цифрового компьютера или управляемого компьютером робота выполнять задачи, обычно связанные с разумными существами. Термин часто применяется к проекту разработки систем, наделенных интеллектуальными процессами, характерными для человека, такими как способность рассуждать, находить смысл, обобщать или учиться на прошлом опыте» [11].

Во всех приведенных определениях ядром выступает выделение имитации и копирования того действия, которое ставится в основание для последующей разработки моделей ИИ. Возникает вопрос методологического свойства: это единственная возможная стратегия при разработке ИИ или возможны и другие?

Дело в том, что прежде всего при таких определениях редуцируется сам человек. ИИ определяется фактически через редукцию человека к отдельной функции и допущение искусственной имитации этой функции:

- определение человека как существа, способного к интеллектуальной деятельности;
- выделение этого отдельного типа деятельности, интеллектуальной деятельности, в отдельную функцию, и отделение ее от человека;
- допущение, что эту функцию можно воспроизвести, копировать, имитировать, строить симуляционные, имитационные искусственные модели, наделенные этой функцией;
- допущение, что интеллектуальная деятельность может быть описана в категориях вычислений.

Фактически уже прошло более 60 лет, а тот тренд в определении ИИ, который был задан с середины 50-х годов прошлого века, так и остается базовым: определение ИИ через копирование человека, разумная деятельность которого описывается в категориях вычислений. Например, Дж. Маккарти (J. McCarthy) в 1956 г. так и полагал, признаваясь, что это лишь косвенное определение человеческого интеллекта, вынужденная редукция. Мы, указывая Маккарти, 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 наших знаний понимается только «вычислительная составляющая способности достигать целей в мире» [12].

А. Тьюринг делал то же самое. Но он не стремился строить философские обобщения. Он, будучи математиком, построил для себя логическую задачку-модель на имитацию и заложил в ней ряд допущений, из которых строятся в принципе любые математические задачи. Из этих допущений нельзя делать далеко идущие выводы и строить на их основании онтологические характеристики и концепты, обсуждать проблему природы мышления и человека в целом. Он на это не претендовал, предложив тогда, в 1950 г., *поиграть в имитационную игру* [13]. А ответ на вопросы, что такое «машина», что такое «мыслить» и умеет ли машина мыслить, он давать не спешил.

Вопрос «Могут ли машины мыслить?» он заменил на другой: «Могут ли цифровые вычислительные машины с помощью человека играть в имитацию?», т. е. как бы притворяться человеком? Тьюринг был уверен, что через 50 лет цифровые машины настолько быстро будут уметь считать, что обычный человек, находящийся за стеной и не видящий, с кем (чем) он разговаривает — с человеком или машиной, — не сможет их отличить. И все приведенные им самим в своей статье возражения противников относительно возможности машин мыслить он парировал не просто аргументами, а именно рефлексивной позицией: он выступал как разработчик, ставящий мысленный эксперимент, строящий имитационную модель, но не более того. В этой части он действовал как обычный классический ученый, ставящий исследовательскую задачу и не ведущий онтологических споров, действуя по принципу «если, то...», «представим себе...». Поэтому все вопросы и возражения относятся не к ИИ, не к некоей за пределом человека находящейся реальности умных машин, а к модели, предложенной автором, Тьюрингом. Поэтому разговор здесь становится сугубо проектно-рефлексивным.

В этой связи рассуждения о том, что есть ИИ, важны скорее своей рефлексивностью, нежели спорами о том, возможен ли ИИ. Они носят характер *методологических и мировоззренческих поисков и допущений*, а не объектных характеристик. Поэтому имеет смысл их обсуждать с точки зрения не того, что есть на самом деле ИИ и возможен ли он (и через него косвенно — что есть сам человек, так рассуждающий), а того, что и как делает автор высказываний, рассуждающий об ИИ. Смысл и польза многообразных концептов ИИ в этом плане имеет сугубо проектно-рефлексивный и методологический характер. Какие мы полагаем допущения относительно ИИ? Какими мыслительными средствами пользуемся?

Базовые редукции

При обсуждении и понимании человеком того или иного сложного явления он вынужден прибегать к разного рода аналогиям, метафорам и так или иначе допускает упрощенные схемы и модели понимания. Так ему удобнее. Вот и в случае ИИ его разработчики так или иначе прибегают к ряду базовых редукций, с их точки зрения помогающих понять свой собственный предмет. Каковы они?

Согласно *первой редукции и допущению*, по внешней линии поведения можно судить о существовании происходящего. На этом основан и тест Тьюринга: компьютер мыслит, если он *ведет себя точно так же, как человек* в момент его размышлений.

Мыслит, т. е. *ведет себя так же*. Мысленный акт уподоблен акту поведения. На этом основан простой актерский пример. Актерствующий субъект *изображает*, что он думает. Он сидит, молчит или ходит по комнате взад и вперед, курит, снова сидит, молчит, морщит лоб, смотрит вдаль в окно, мы смотрим на него и полагаем, что он думает. Уподобление мысленного акта и наблюдаемого акта поведения — сильнейшая редукция и базовое допущение всех натуралистов и бихевиористов.

Мы в таком случае вынуждены описывать акт мышления в категориях поведенческих реакций, в категориях внешне наблюдаемых действий объекта, сравнивая — похож, не похож, так ведет себя или не так, ловим внешние признаки поведения, даем характеристики т. д. и тем самым упускаем главное — внутренний процесс, присущий акту мышления, скрытому от внешнего наблюдателя.

Согласно *второму допущению*, интеллектуальное действие сродни логической операции, которую проделывает вычислительное устройство. Тем самым полагается, что умственная деятельность — это выполнение некоторой хорошо выстроенной последовательности логических операций, называемой алгоритмом. То есть мышление — это действие по алгоритму. А значит, его может осуществить машина с заложенным в нее алгоритмом. Чем лучше алгоритм и чем больше данных можно обработать с его помощью, тем лучше машина будет «мыслить», т. е. считать. Соответственно, ставится задача создать машину, которая считала бы быстро и не совершала ошибок. В настоящее время ИИ опережает человека по быстродействию и по числу искусственных нейронов. Означает ли это, что она мыслит? Причем лучше человека?⁴

Р. Пенроуз (R. Penrose) полагал, что именно этот пункт — самый важный при создании ИИ. И совершенно не важно, по его мнению, кто (что) приводит в действие этот алгоритм — человеческий мозг, компьютер, целое государство индийцев или механическое устройство, состоящее из колесиков и шестеренок [14, с. 61]. Главное — построение логической структуры алгоритма, а его физическая основа никакой роли не играет. Алгоритм, полагал Р. Пенроуз, обладает бесплотным существованием, не связанным с физической реализацией⁵. Тем самым мы подходим к идее бесплотности, идеальной сущности математического алгоритма.

В данном случае допускается опять же редукция, согласно которой у мышления есть физический носитель. Но проблема в том, что само по себе наличие физического носителя, мозг это или машина, не является достаточной для осуществления события мышления. Просто потому, что природа мышления не редуцируется к носителю. Она событийна. Один носитель мозга мыслит, другой такой же

⁴ К слову, модели ИИ уже сейчас количественно перегнали человека по объему нейронов. Компания Digital Reasoning создала и обучила нейронную сеть, состоящую из 160 млрд цифровых нейронов. Но мыслить она не умеет. Она умеет быстро перебирать данные и выдавать прогнозы и решения. Но не умеет ответить на вопрос, почему она принимает те или иные решения.

⁵ В свое время методолог Г. П. Щедровицкий также допускал, что мышление как субстанция может паразитировать на разных физических носителях, неважно, на людях или на смешанных системах, состоящих из людей и машин. Главное — есть мышление или нет, а на чем оно реализуется — не важно. В нашем мире случайно оно живет на людях, в другом мире — на пингвинах, в третьем — на железках [15, с. 561].

носитель не мыслит, а просто ведет себя как двуногое и бескрылое животное. Но смоделировать событие мышления — совсем другая задача, на сегодняшний день даже труднопредставимая.

Согласно *третьему допущению*, коль скоро мышление аналогично поведению и мыслит машина программно, то можно построить имитационную модель такого мыслительного поведения, т. е. модель ИИ. И сама схема выстраивается по логике имитации. Есть оригинал, затем строится копия, имитирующая оригинал, его поведение. Тем самым все упирается в выстраивание связки: что берется за оригинал и что значит построить его копию.

В соответствии с четвертым допущением, коль скоро ставится задача, связанная с построением имитационной модели, в качестве исходного оригинала берется активность мозга. Допускается, что тайна мысли скрыта в активности мозга, в активности его нейронов.

По логике имитации и симуляции фактически и создается модель ИИ в виде искусственной нейронной сети. Нейрон — это клетка, которая обрабатывает, хранит и передает информацию за счет электрических сигналов через синаптические связи. Система миллиардов нейронов создает такую сеть⁶.

Нейрофизиологи полагают, что здесь и скрыта тайна интеллектуальной деятельности — сеть нейронов, объединенных в миллиарды связей, и есть машина, коммутатор, порождающий мышление. Поэтому искусственная нейронная сеть становится базовой моделью для создания ИИ. Именно способность биологической нейронной сети обучаться и исправлять свои ошибки легла в основу создания ИИ⁷.

Но этот ход — также сильная редукция, поскольку мыслит не мозг. Мыслит сложно устроенный «функциональный орган», состоящий из мозга, органа тела человека (например, руки), глаза (уха), предмета, которым оперирует та же рука (например, ложка или молоток), опосредованного словом предметного действия. Это действие с предметом человек совершает много раз, чтобы освоить его и чтобы оно отложилось в локальной зоне мозга. Совершая действие, закрепляя его, опосредуя знаком-словом, человек овладевает собой, своими аффектами, своим поведением. Так, ребенок много раз пробуя совершить необходимое действие с ложкой, чтобы научиться есть кашу ложкой, овладевает этим предметным действием.

Принцип опосредствования

⁶ Человеческий мозг имеет порядка 86 млрд нейронов, имеющих в сумме 2^{14} контактов, в мозге ежесекундно происходит обмен более 10^{15} нервных импульсов, своего рода звонков между нейронами [16, с. 26].

⁷ Должен заметить, что уже в 1958 г., в период создания первых искусственных нейросетей, перцептронов, их создатель Ф. Розенблатт (F. Rosenblatt), основатель новой дисциплины — нейродинамики, сам признавался, что перцептроны, конечно, плохие копии биологически сетей, но они нужны для изучения деятельности мозга. Перцептроны выступали прежде всего моделью мозга (пусть и сильно упрощенной) для проведения экспериментов по изучению активности мозга, а не как устройства ИИ. Их создатели осознавали степень упрощения и редукции, но подобные модели могли подвергаться анализу и были удобны при проведении экспериментов [17].

Глубинная редукция, согласно которой мышление локализуется в мозге, идущая со времен И. П. Павлова (точнее, со времен «машины Р. Декарта») и первых бихевиористов, описывающая интеллектуальное действие в схеме стимул — реакция (S — R), против которой выступал Л. С. Выготский и его ученики и соратники, до сих пор доминирует в методологических концептах авторов-разработчиков моделей ИИ. Они все убеждены, что мышление происходит в мозге, причем «автор» мышления — нейросеть, поэтому надо ее моделировать, создавать имитационную, симуляционную модель биологической нейросети. Более того, мысль понимается как сугубо химическая реакция, происходящая в мозге, сравнимая с электрическим сигналом. Такой пещерный натурализм загоняет разработки в области ИИ в тупик, поскольку равным счетом никакого отношения такие допущения к природе мышления не имеют [18].

Здесь нет возможности пересказывать основы культурно-исторической концепции Л. С. Выготского, тем более что она давно и много раз описана в его собственных трудах и работах его учеников, соратников и последователей. Скажем только, что в ней предложена принципиально иная модель формирования мышления, его структуры, его устройства [19].

Для Л. С. Выготского «центральной проблемой при объяснении «высших психологических функций» (как то: мышления, памяти, воли), является проблема средств, с помощью которых человек овладевает процессом собственного поведения [20, с. 126]. Главным таким средством выступает знак, слово, как «психическое орудие», выполняющее в развитии высших психических функций роль посредника. Высшие формы поведения суть процессы опосредованные. Поэтому важно понять то, как действие, мысль, знак-слово, орган тела человека (рука) и функция мозга, в которой откладывается память предметного действия, соединены в эту операцию опосредования, совершая которую человек овладевает своим натуральным, стихийным поведением, выстраивая высшие, осознанные формы поведения. В структуре процесса опосредования мышление и может быть именно языковым и деятельностным образованием. Отдельно взятого мышления нет. Мысль «не выражается, но совершается в слове», писал Выготский [21, с. 307].

В этой связи ключевой проблемой становится проблема *моделирования процесса и процедур актов опосредствования действия с помощью знаков*. Такой ход выглядит принципиально иным, отличным от имитации и копирования нейросетей.

Как мы отметили выше, машина действует по принципу быстрой обработки больших данных, т. е. действует максимально неэффективно. Она не умеет делать обобщения, строить понятия, вырабатывать стратегии или совершать ходы по диагонали, наперекор алгоритму. Обобщение у ИИ на самом деле пока не получается. Объединение баз данных и статистический перебор вариантов не есть обобщение. Последнее предполагает выход на понятийный уровень, что делает уже ученик начальной школы.

Машина же делает обобщение по принципу натурального объединения признаков: стул имеет определенное количество ножек и функцию. Человек же задает иное полагание: он оперирует не только признаками и функциями, но и смыслами. Поэтому в качестве стула могут быть избраны и стол, и спина папы, и кровать, и любая доска. Ребенок может перепредмечивать предмет, играя смыслами, отрываясь от натурального ряда вещей и признаков (тапок — пароход). И наобо-

рот — неспособность отрываться от натурального поля, неспособность оперировать символами и смыслами показывает, что сознание натурально. Человек страдает психическим расстройством, будучи «рабом зрительного поля»⁸. Машина в этой связи где-то похожа на больного деменцией. Она хоть и играет в игру Go, но она не осознает, что играет в Go. Она не понимает значений символов, которыми виртуозно и быстро оперирует.

Заключение

Имеет смысл выходить на разработку иных моделей ИИ, строящихся не по логике имитации внешнего поведения и симуляции активности человеческого мозга, а по логике построения формирующей предметной мыследеятельности. Необходимо строить не модели мозга, а *модели предметной деятельности опосредования*, в которой объединены: предмет (орудие) — орган — зона специализации мозга — предметное действие — знак, опосредующий действие.

Разработка такого рода цифровых моделей, в которых описывается не только активность биологической нейросети, но и структура знаково опосредованного предметного действия, выступает более продвинутым ходом при разработке моделей ИИ. И тогда само машинное обучение имеет смысл выстраивать по логике вышеназванной методики развивающего обучения, по логике опосредования⁹.

В настоящее же время и ИИ, и нейронные сети, и машинное обучение выступают как метафоры, используемые для построения имитационных моделей ИИ. Они не могут играть роль строгих понятий. Равно как и реальность ИИ есть реальность наших представлений о мышлении и в целом о человеке. Мы больше приписываем ИИ интеллект, редуцируя и понимая, что так называемый интеллект машин слабее интеллекта трехлетнего ребенка.

Поэтому главным остается вопрос: когда человек приступает к разработке систем ИИ, как он представляет самого себя и какой новый интерфейс он собирается выстраивать с умной машиной?

Литература

1. Хэвен, Д. (ред.) (2019), *Искусственный интеллект. Что стоит знать о наступающей эпохе разумных машин*, пер. с англ. Сайфутдинова, О. Д., М.: АСТ.
2. Маркус, Г. и Дэвис, Э. (2021), *Искусственный интеллект: Перегрузка. Как создать машинный разум, которому действительно можно доверять*, пер. с англ. Скворцов, В., М.: АльпинаПРО.

⁸ См. о явлении деменции при болезни Пика: [22].

⁹ Фактически нейронауки уже вышли на постановку таких задач. Например, А. Каплан показывает, что в человеческом мозге формируется так называемая динамическая ментальная модель внешнего мира, которой не требуется жесткая привязка к нейронам [16, с.31]. Учитывая то, что ежедневно в мозге умирают тысячи нейронов, возникает вопрос: если нейроны специализируются на тех или иных операциях, то как нейросеть справляется с такой динамикой? Например, если в телефонной сети происходит сбой, то не работает вся сеть. Но получается, что над нейросетью формируется динамическая модель, ответственная за творчество, открытия, интуицию. В ней фиксируются «репрезентации внешнего мира», становящиеся самостоятельными «агентами» поведения. Подобная надсетевая модель формируется, разумеется, в живом реальном опыте, она от рождения человеку не дается. Но в таком случае необходимо изучать и строить модели этого опыта. От структур опыта, точнее, предметной деятельности, зависят структуры, морфология человеческого мозга. И не иначе.

3. Фридман, М. (2021), *Философия на перепутье: Карнап, Кассирер и Хайдеггер*, пер. с англ. Целищев, В. В., М.: Канон+ РООИ «Реабилитация»
4. Хайдеггер, М. (2007), *Что такое метафизика?*, пер. с нем. Библихин, В. В., М.: Академический проект.
5. Хайдеггер, М. (2020), *К философии (О событии)*, пер. с нем. Сагетдинов, Э., М.: Изд-во Института Гайдара.
6. Ершов, Ю. Л. и Целищев, В. В. (2012), *Алгоритмы и вычислимость в человеческом познании*, Новосибирск: Изд-во СО РАН.
7. Указ Президента РФ, (2019), *Указ Президента РФ «О развитии искусственного интеллекта в РФ»* от 19.10.2019, № 490. URL: <http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=102608394> (дата обращения: 24.02.2022).
8. Финн, В. К. (2009), Искусственный интеллект, в: *Энциклопедия эпистемологии и философии науки*, М.: Канон+, РООИ «Реабилитация», с. 316–318.
9. Гусев, А. В. и Добридняк, С. Л. (2017), Искусственный интеллект в медицине и здравоохранении, *Информационное общество*, № 4–5, с. 78–93.
10. *Oxford English Dictionary* (2022). URL: <https://www.oxfordlearnersdictionaries.com> (дата обращения: 24.02.2022).
11. Copeland, B. J. (2022), Artificial Intelligence, *Encyclopedia Britannica*, URL: <https://www.britannica.com/technology/artificial-intelligence> (дата обращения: 24.02.2022).
12. McCarthy, J. (2022), What is artificial intelligence? URL: <http://www-formal.stanford.edu/jmc/whatisai/whatisai.html> (дата обращения: 24.02.2022).
13. Turing, A. (1950), Computing Machinery and Intelligence, *Mind*, vol. 59, pp. 433–460.
14. Пенроуз, Р. (2020), *Новый ум короля: О компьютерах, мышлении и новых законах физики*, пер. с англ. Малышенко, В. О., М.: УРСС: ЛЕНАНД.
15. Щедровицкий, Г. П. (1997), *Философия. Наука. Методология*, М.: Школа культурной политики.
16. Каплан, А. Я. (2019), Мы — это больше, чем мозг: в поисках субъективного начала, *Труды кафедры богословия*, № 2(4), с. 25–34.
17. Пасквинелли, М. (2019), Машины, формирующие(ся) в логику: нейронные сети и искаженная автоматизация интеллекта в качестве статистического вывода, *Новое литературное обозрение*, № 4, с. 153–168.
18. Смирнов, С. А. (2016), Событийность мысли (к вопросу об онтологии событийности), *Вопросы философии*, № 8, с. 103–114.
19. Давыдов, В. В. (1996), *Теория развивающего обучения*, М.: ИНТОР.
20. Выготский, Л. С. (1982), *Собрание сочинений: в 6 т., т. 2*, М.: Педагогика.
21. Выготский, Л. С. (1984), *Собрание сочинений: в 6 т., т. 6*, М.: Педагогика.
22. Самухин, Н. В., Биренбаум, Г. В. и Выготский, Л. С. (1981), К вопросу о деменции при болезни Пика, в: *Хрестоматия по патопсихологии*, М.: МГУ, с. 114–149.

Статья поступила в редакцию 27 ноября 2022 г.;
рекомендована к печати 15 апреля 2023 г.

Контактная информация:

Смирнов Сергей Алевтинович — д-р филос. наук, вед. науч. сотр.; smirnoff1955@yandex.ru

Is Human Being Computable, or Anthropology of Artificial Intelligence. Methodological Aspect*

S. A. Smirnov

Institute of Philosophy and Law of the Siberian Branch of the Russian Academy of Sciences,
8, ul. Nikolayeva, Novosibirsk, 630090, Russian Federation

For citation: Smirnov S. A. Is Human Being Computable, or Anthropology of Artificial Intelligence. Methodological Aspect. *Vestnik of Saint Petersburg University. Philosophy and Conflict Studies*, 2023, vol. 39, issue 3, pp. 478–491. <https://doi.org/10.21638/spbu17.2023.306> (In Russian)

The article analyzes the anthropological paradigms of the mainstream approaches to AI development. The article concludes that the principle of computability (quantifiability) first introduced in the early stages of AI development (namely by A. Turing) is still paramount. The article showcases the way this principle is rooted in what Heidegger described as treating humans and the state of being itself as a consumable and ultimately destructible commodity. One tries to quantify the state of being in order to consume it. As a consequence, humanity approaches the creation of a “thinking machine” as an exercise in self-imitation. The article highlights the deficiencies of creating AI via copying human patterns. The article also suggests a viable alternative — a mediatory method, developed by culturally-historical psychology. The latter perceives the thought process not as a brain activity but as a objective act, mediated through sings and speech. This way of thinking leads to a new strategy of AI development — as a smart helping hand for humans. This, however, requires an overhaul in medical, educational, governmental and many other social practices. The author concludes that any AI development is bound to start with a social engineering process, renovating the principles, governing day to day lives. This is why the development of AI is not a purely technological task. It affects a deeper level of human existence, unfit to be interpreted in numerical terms.

Keywords: artificial intelligence, human, anthropological basis, the mediation principle, the quantifiability principle.

References

1. Haven D. et al. (2019), *Machine that Think*. Everything You Need to Know about the Coming Age of Artificial Intelligence, trans. by Saifutdinova, O. D., Moscow: AST Publ. (In Russian)
2. Markus, G. and Davis, E. (2021), *Rebooting AI*. Building Artificial Intelligence we can trust, trans. by Skvortzov, B., Moscow: Al'pinaPRO Publ. (In Russian)
3. Friedman, M. (2021), *Parting of the Ways: Carnap, Cassirer and Heidegger*, trans. by Tselishev, V. V., Moscow: Kanon+ ROOI “Reabilitatsiia” Publ. (In Russian)
4. Heidegger, M. (2007), *Was ist Metaphysik?*, trans. by Bibikhin, V. V., Moscow: Akademicheskii projekt Publ. (In Russian)
5. Heidegger, M. (2020), *Beiträge zur Philosophie (vom Ereignis)*, trans. by Sagetdinov, B., Moscow: Institut Gaidara Publ. (In Russian)
6. Ershov, Ju. L. and Tselishhev, V. V. (2012) *Algorithms and Computability in Human Knowledge*, Novosibirsk: Sibirskoe otdelenie RAN Publ. (In Russian)
7. Decree of the President of the Russian Federation “On the development of artificial intelligence in the Russian Federation”, 19.10.2019, no. 490. Available at: <http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=102608394> (accessed: 24.02.2022). (In Russian)
8. Finn, V.K. (2009), Artificial Intelligence, in: *Entsiklopediia epistemologii i filosofii nauki*, Moscow: Kanon+, ROOI “Reabilitatsiia” Publ., pp. 316–318. (In Russian)
9. Gusev, A.V. and Dobridnjuk, S.L. (2017), Artificial Intelligence in medicine and healthcare, *Informatsionnoe obshchestvo*, no. 4–5, pp. 78–93. (In Russian)

* The work was written within the framework of the grant project supported by the Russian Science Foundation project no. 21-18-00103. <https://rscf.ru/project/21-18-00103/>

10. Oxford English Dictionary, (2022). Available at: <https://www.oxfordlearnersdictionaries.com> (accessed: 24.02.2022).
11. Copeland, B.J. (2022), Artificial Intelligence, in: *Encyclopedia Britannica*, (2022). Available at: <https://www.britannica.com/technology/artificial-intelligence> (accessed: 24.02.2022).
12. McCarthy, J. (2022), What is artificial intelligence? Available at: <http://www-formal.stanford.edu/jmc/whatisai/whatisai.html> (accessed: 24.02.2022).
13. Turing, A. (1950), Computing Machinery and Intelligence, *Mind*, vol. 59, pp. 433–460.
14. Penrouz, R. (2020), *Emperor's New Mind: Concerning Computer, Mind and the Laws of Physics*, trans. by Malyshenko, V.O., Moscow: URSS: LENAND Publ. (In Russian)
15. Shchedrovicky, G.P. (1997), *Philosophy. Science. Methodology*, Moscow: Shkola kul'turnoi politiki Publ. (In Russian)
16. Kaplan, A. Ja. (2019), We are more than a brain: in search of a subjective beginning, *Trudy kafedry bogosloviia*, no. 2(4), pp. 25–34. (In Russian)
17. Paskvinelli, M. (2019), Machines that Morph Logic: Neural Network and the Distorted Automation of Intelligence as Statistical Inference, *Novoe literaturnoe obozrenie*, no. 4, pp. 153–168. (In Russian)
18. Smirnov, S. A. (2016), Eventfulness of thought. On the question of the ontology of eventfulness, *Voprosy filosofii*, no. 8, pp. 103–114. (In Russian)
19. Davydov, V. V. (1996), *Developmental Learning Theory*. Moscow: INTOR Publ. (In Russian)
20. Vygotskij, L. S. (1982), *Collected Works*, vol. 2. Moscow: Pedagogika Publ. (In Russian)
21. Vygotskij, L. S. (1984), *Collected Works*, vol. 6, Moscow: Pedagogika Publ. (In Russian)
22. Samuhin, N. V., Birenbaum, G. V. and Vygotskij, L. S. (1981), To the question of dementia in Pick's disease, in: *Khrestomatiia po patopsikhologii*, MSU Publ., pp. 114–149. (In Russian)

Received: November 27, 2022

Accepted: April 15, 2023

Author's information:

Sergey A. Smirnov — Dr. Sci. in Philosophy, Leading Researcher; smirnoff1955@yandex.ru