ИНСТИТУТ ФИЛОСОФИИ И ПРАВА
СИБИРСКОЕ ОТДЕЛЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК

ЧЕЛОВЕК И НОВЫЙ ТЕХНОЛОГИЧЕСКИЙ УКЛАД. АНТРОПОЛОГИЧЕСКИЙ ФОРСАЙТ

Новосибирск
2023

ББК 87

The monograph was published with the support
of the Russian Science Foundation,
scientific project No 21-18-00103 “Man and the New Technological Order.
Anthropological Foresight”. https://rscf.ru/project/21-18-00103/

Authors:
S. S. Avanesov, A. G. Gorbacheva, A. S. Zaikova,
P. V. Kaigorodov, A. I. Pestunov, O. A. Persidskaya,
T. A. Sidorova, N. A. Sinukova, S. A. Smirnov (leader), E. I. Speshilova

Editors:
S. A. Smirnov, O. A. Persidskaya

The cover design uses the work of a Japanese artist
Biloker Ji “Take care of mother”

Man and the New Technological Order. Anthropological Foresight.
Collective monograph. Novosibirsk: Offset-TM LLC, 2023, 496 p.

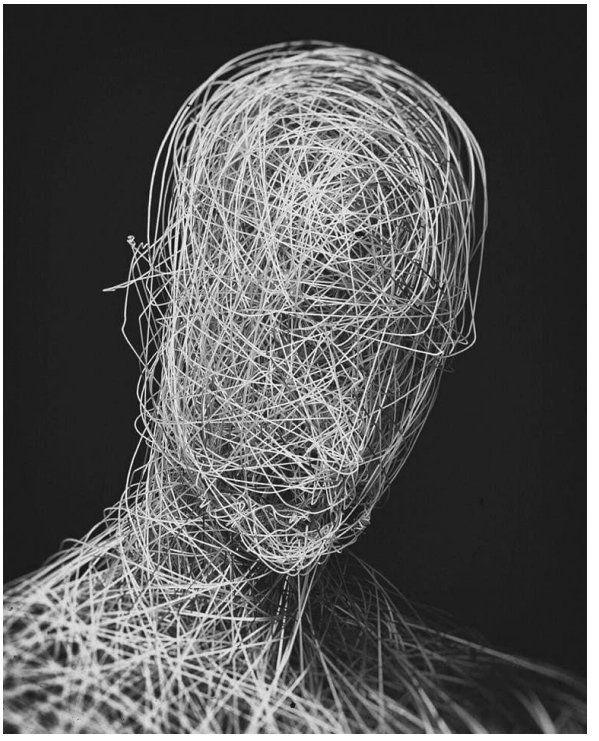
ISBN 978-5-85957-210-6

The monograph presents the experience of constructing a conceptual outline of the anthropology of the future. The contour is built through the prism of a person’s search for his irreplaceable place in the future, which implies the re-realization of ontological self-determination. The team of authors offers an anthropological alternative, that is, their vision of the image of the future person, which is not associated either with the trend of human departure in favor of the posthuman, as representatives of transhumanism call for, or with the trend of abandoning digitalization. A person must remain the meaningful point of the world and determine his future from the point of view of practices in which digital technologies can be used as smart, effective assistants, which involves building a new interface between humans and smart technology.

© Group of authors, 2023

Глава 5

АНТРОПОЛОГИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА



ИСЧИСЛИМО ЛИ БЫТИЕ ЧЕЛОВЕКА ИЛИ АНТРОПОЛОГИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
Методологический аспект

С. А. Смирнов[149]

«... предложение “Машина мыслит (воспринимает, желает)” 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 употребляется в двойственном смысле. Первый означает то, что это венец его творения, свидетельствующий о высшем уровне научной и технологической мысли. Второй означает то, 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, то есть уничтожить как сущее, 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 представлениях 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 строят по этому поводу модели? 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, связанный с разработкой и внедрением ИИ, связан действительно с радикальными вызовами для человека, ставящими его самого в ситуацию онтологического испытания, испытания собственными границами. Выделим здесь два момента.

Первое. ИИ впервые поставил под вопрос уникальность и исключительность человека [Хэвен 2019: 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.

Второе. 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 сущее, и как следствие, он может изменить понимание себя, переосмысляя то, что значит быть собой, человеком.

[149] © Смирнов С. А. 2023

В таком случае все зависит от самого автора, человека-создателя – какими средствами, представлениями о себе самом он пользуется, выстраивая эти зеркальные модели самого себя? Если модельные копии не адекватны оригиналу, то значит не адекватны прежде всего его представления о себе самом, выставляющим себя всякий раз в редуцированном, участненном, урезанном виде.

В связи с выше сказанным 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 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 полагания: *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 мозга[150]. 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 быстродействие. Машина должна уметь быстро считать, перебирая миллиарды вариантов в секунду. Поэтому быстродействующий компьютер в итоге обыграл человека сначала в шахматы, потом в игру Go[151].

Хотя сам А. Тьюринг ставил задачу шире. Он не компьютер хотел создать. Сначала он хотел решить математическую задачу, так называемую «проблему разрешения» Гильберта. Последний поставил задачу: 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, способ-

[150] 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).

[151] 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. Пока ИИ есть искусственная математическая модель с заложенными в нее параметрами и действиями по алгоритму.

ный дать ответ на любое логическое выражение. В основании такого 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.

Впоследствии, при разработке всех поисковых систем в интернете, при организации интернет-торговли этот принцип и был заложен – машина обучалась за счет быстродействия, 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 на другой раз у него на экране монитора будут всплывать 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, обучающейся путем проб и ошибок. То есть, положен принцип вычислимости [Маркус, Дэвис 2021: 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, отстаивающими идеи логического позитивизма, с одной стороны, и представителями метафизики, главной фигурой коих был признан М. Хайдеггер, с другой, шел интенсивный спор [Фридман 2021: 352]. Внешне он выглядел как спор о предмете философии, ее будущем. Р. Карнап 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 и должна быть таковой – понятной, данной в процедурах описания и расчета. В сущем не должно быть тайны. Все сущее должно быть научно, то есть логически, экспериментально, опытно обосновано и просчитано, исчислено. В основание сущего тем самым закладывался *принцип вычислимости*, то есть 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 [Фридман 2021: 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» [Хайдеггер 2007: 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 дельнейшее его, мира и человека, употребление, в пределе – уничтожение в угоду выгоде, прибыли, капитализации. Этих слов 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» [Хайдеггер 2007: 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 потаенно обусловлено полной пустотой, в которой взвешено сущее, материал для действительного. Эта пустота требует заполнения» [Хайдеггер 2007: 252–253].

Но пустоту бытия никогда не заполнить полнотой только сущего, поэтому возникает необходимость нарастания «обесцеленной деятельности» (сиречь – *имитации и симуляции*). В этом случае именно техника подходит для обеспечения этого недостатка. Техника выступает заменителем, эрзацем сущего, в пределе – протезом для человека, подпирающим 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», то есть собственно, искомым искусственным интеллектом.

Исчисляемое и потребляемое сущее, будучи исчислено, становится «пред-ставленным», удобным для потребления и «более доступным» [Хайдеггер 2020а: 605]. Но в кажимости 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 пози-

тивно. Мы говорим о том, что в основании заложена вполне определенная онтологическая установка, направленная на исчисление и соответственно редукцию человека[152].

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

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» [Указ 2019: 5а].

2. «Направление исследований в современной компьютерной науке, целью которого является *имитация и усиление интеллектуальной деятельности человека* посредством компьютерных систем» [Финн 2009: 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 [Гусев, Добриднюк 2017: 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 [Dictionary 2022].

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» [Copeland 2022].

Во всех приведенных определениях ядром выступает выделение имитации и копирования как того действия, которое ставится в основание для последующей разработки моделей ИИ. Возникает вопрос методологическо-

---

[152] В философии математики выше названные вопросы связаны с проблемой вычислимости и невычислимости: «…вычислимость предполагает редукцию в описании, которая избегает или выводит за скобки алгоритмически несжимаемые фрагменты вселенной» [Ершов, Целищев 2012: 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, то есть феномен алгоритмической сжимаемости, становится основой для «пошагового процесса улучшения описания», то есть операции вычисления в смысле Тьюринга [Ершов, Целищев 2012: 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» [Ершов, Целищев 2012: 443]. Мышление в принципе носит не компьютерный характер.

го свойства: это единственная возможная стратегия при разработке ИИ? Или возможны и другие?[153]

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 в 1956 году так и полагал, вынужденно признаваясь, что это лишь косвенное определение человеческого интеллекта, вынужденная редукция. Мы, указывал 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» [McCarthy 2022].

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 онтологические характеристики и концепты, обсуждать проблему природы мышления и человека в целом. Он на это не претендовал, предложив тогда, в 1950 году, *поиграть в имитационную игру* [Turing 1950]. А ответ на вопрос, что такое «машина», что такое «мыслить», и умеет ли машина мыслить, он давать не спешил, а предложил игру в имитацию, то есть, свой тест со стеной.

Вопрос «могут ли машины мыслить?» он заменил на другой – «могут ли цифровые вычислительные машины с помощью человека играть в имитацию?», то есть как бы притворяться человеком? Тьюринг был уверен, что через 50 лет цифровые машины настолько быстро будут уметь считать, что средний человек за стеной будет путаться в вопросе, с кем (чем) он разго-

[153] В качестве уточнения заметим, что мы обсуждаем прежде всего методологические основания, задающие стратегию создания сильного ИИ, то есть такого, который собственно имитирует человеческую интеллектуальную (мыслительную) деятельность. Понятно, что примеры ИИ мы вынуждены приводить из области слабого ИИ, поскольку сильный ИИ не создан и едва ли по логике стратегии вычислимости и имитации может быть создан. Мы в данном случае поэтому обсуждаем то, какие стратегии возможны при создании сильного ИИ, а не отдельные примеры, полагая, что стратегия вычислимости и имитации отдельных функций мозга не может быть выбрана в качестве таковой.

варивает – с человеком или машиной. 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 не с точки зрения того, что есть на самом деле ИИ и возможен ли он (и через него косвенно – что есть сам человек, так рассуждающий), а то, что и как делает автор высказываний, рассуждающий об ИИ. Смысл и польза многообразных концептов об 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 так или иначе вынужден прибегать к разного рода аналогиям, метафорам, и так или иначе допускает упрощенные схемы и модели понимания. Там ему удобнее. Вот и в этом случае. Разработчики ИИ прибегают к ряду базовых редукций, с их точки зрения помогающих понять свой собственный предмет. Каковы они?

Первая редукция и допущение, согласно которому по внешней линии поведения можно судить о существе происходящего. На этом основан и тест Тьюринга: компьютер мыслит, если он *ведет себя точно так же, как человек* в момент его размышлений.

Мыслит, то есть *ведет* себя так же. Мысленный акт уподоблен акту поведения. На этом основан простой актерский пример. Сидит он, этот актерствующий субъект, и *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 акта поведения, видимого внешним образом – 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 и т. д. и тем самым упускаем главное – внутренний процесс, присущий акту мышления, скрытому от внешнего наблюдателя.

Второе допущение, согласно которому интеллектуальное действие сродни логической операции, которую проделывает вычислительное

устройство. И тем самым полагается, что умственная деятельность – это выполнение некоторой хорошо выстроенной последовательности логических операций, называемой алгоритмом. То есть мышление есть действие по алгоритму. А значит его может сделать искусственная машина с заложенным в нее алгоритмом. Значит, чем лучше алгоритм и чем больше данных можно обработать с его помощью, тем лучше машина будет «мыслить», то есть считать. Значит, задача ставится такая, чтобы машина считала 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?[154]

Р. Пенроуз 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 [Пенроуз 2020: 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 реализацией[155]. 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, пусть разный. Но проблема в том, что само по себе наличие физического носителя, мозг это или машина, не является достаточным для осуществления события мышления. Просто потому, что природа мышления не редуцируется к носителю. Она, природа мысли, событийна. Один носитель мозга мыслит, другой же, такой же носитель, не мыслит, а просто ведет себя как двуногое и бескрылое животное. Но смоделировать событие мышления – совсем другая задача, на сегодняшний день даже трудно представимая.

Допущение третье, согласно которому коль скоро мышление аналогично поведению и мыслит машина программно, то можно построить имитационную модель такого мыслительного поведения, то есть 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.

Допущение четвертое. В таком случае, коль скоро ставится задача, связанная с построением имитационной модели, то в качестве исходного ори-

---

[154] К слову сказать, 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.

[155] В свое время методолог Г. П. Щедровицкий также допускал, что мышление как субстанция может паразитировать на разных физических носителях, не важно, 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 [Щедровицкий 1997: 561].

гинала 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 сеть[156].

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 ИИ[157].

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 мозге. Совершая действие, закрепляя его, опосредуя знаком-словом, человек овладевает собой, своими аффектами, своим поведением. И только после этого, например, ребенок, много раз совершая, пробуя действие с ложкой, чтобы научиться есть кашу ложкой, овладевает этим предметным действием.

Принцип опосредствования

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, по-

[156] 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 звонками между нейронами [Каплан 2019: 26].

[157] Должен заметить, что уже в 1958 году, в период создания первых искусственных нейросетей, перцептронов, их создатель Ф. Розенблатт, 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 [Пасквинелли 2019].

скольку ровным счетом никакого отношения такие допущения к природе мышления не имеют[158].

Здесь нет возможности пересказывать основы культурно-исторической концепции Л. С. Выготского. Скажем только, что в ней предложена принципиально иная модель формирования мышления, его структуры, его устройства, тем более это давно и много раз описано в трудах Л. С. Выготского и его учеников, соратников и последователей [Давыдов 1996].

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» [Выготский 1982б: 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 [Выготский 1984: 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. Поэтому автор и предлагает строить иного типа модели ИИ – модели опосредования, в рамках которых не имитируется активность мозга, а строится деятельностная модель, в которой связаны действие-знак-цифра-орудие. А мышление и есть действие по осуществлению этой связанности.

Как мы отметили выше, машина действует по принципу быстрой обработки больших данных. То есть действует максимально не эффективно. 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 может быть избран и стол, и стул, и спина папы в игре ребенка, и кровать, и любая доска. Ребенок может перепредмечивать предмет, играя смыслами, отрываясь от натурального

[158] Речь не идет об отрицании необходимости разработок умных нейросетей. Но в таком случае надо понимать, что нейросети и иные обучающиеся системы создаются не как заменители человека, а как их цифровые умные помощники, которые уже существуют в разных сферах (в здравоохранении, управлении, на транспорте и др.). А непредсказуемость нейросетей имеет иную природу, отличную о непредсказуемости человеческого действия.

ряда вещей и признаков (тапок – пароход). И наоборот – неспособность отрываться от натурального поля, неспособность оперировать символами и смыслами показывает, что сознание натурально. Человек страдает психическим расстройством, будучи «рабом зрительного поля»[159]. Машина в этой связи где-то похожа на больного деменцией. Она, хоть и играет в игру Go, но 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 выше названой методики развивающего обучения, по логике опосредования[160].

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 3-летнего 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?

---

[159] См. о явлении деменции при болезни Пика [Самухин и др. 1981].

[160] 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 [Каплан 2019: 31]. Учитывая то, что ежедневно в мозге умирает 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.

ПРЕДЕЛЫ ЭТИКИ ДЛЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

С. А. Смирнов[161]

Введение

Разработка и внедрение программ искусственного интеллекта (далее – ИИ) не просто набирают обороты в мире. Это направление становится приоритетным на самом высоком уровне. Приоритетность зафиксирована в принятых стратегических документах развития, в том числе и в нашей стране. Это вполне объяснимо, поскольку предполагается, что лидерство в этом направлении делает страну в целом конкурентоспособным игроком на мировом рынке передовых технологий.

Вместе с тем, при таком темпе развития мало кто задумывается о главной фигуре в этом тренде – о самом человеке, который разрабатывает модели ИИ. Понимаем ли мы, что происходит с самим человеком при развитии этого тренда? Задаём ли мы себе следующие вопросы? И как на них отвечаем?

Какие гуманитарные и антропологические основания закладываются при разработке моделей ИИ? Думают ли про это разработчики проектов ИИ?

Какие не технологические, а гуманитарные последствия могут быть от внедрения проектов ИИ?

При развитии модели ИИ всё более мощно расширяется и углубляется технологический и функциональный аутсорсинг от человека к машине, то есть передача человеком все более умных функций умной машине. Возникает вопрос: что человек оставляет себе при разработке технологий ИИ и что передаёт умному устройству?

Где и когда при таком аутсорсинге наступает граница, за пределами которой передавать умному устройству уже будет нечего, поскольку человек, перейдя эту границу, перестаёт быть самим собой, поскольку все свои умные функции и работы он передал умной машине? И тогда действительно возникает вопрос о замене человека постчеловеком. Наступит ли когда-нибудь такая ситуация и мы упрёмся в границу?

Как и какой должен будет выстраиваться кентавр человека и ИИ в будущем? Какие здесь могут быть более предпочтительными модели? Какие должны быть новые интерфейсы человека и машины, в которых и человек, и машина предстают уже в новом функциональном качестве?

Сложилась ли уже ситуация, при которой возникает необходимость нового этического кодекса для ИИ и создания института гуманитарной экспертизы внедрения умных технологий? Что сейчас необходимо делать в сфере международного права с точки зрения гуманитарного этического регулирования в сфере разработки и внедрения технологий ИИ?

Подходы к модели AI

Прежде, чем начать отвечать на эти вопросы, необходимо остановиться на основном вопросе: какие модели человека закладываются при разработке ИИ и какие при этом гуманитарные допущения позволяют себе разработчики?

[161] © Смирнов С. А. 2023

Определений ИИ в настоящее время достаточно много. Но самым продвинутым выступает определение, согласно которому под ИИ понимается такая программа (или алгоритм), которая на основе обработки данных научается «принимать решения в ситуации неопределенности» [Искусственный интеллект 2019: 13].

Что это значит? Это значит, что модель ИИ должна не только решать задачи по алгоритму, но и уметь *вести себя в непредвиденных ситуациях*, которые не были предусмотрены в заложенном в неё алгоритме. Интеллектуальной системе, в таком случае, ставится не только задача, но моделируются и сами ситуации, при которых эти задачи могут возникнуть. В модель тогда встраивается не только набор данных, не только система задач, но моделируются и сами ситуации, способные порождать те задачи, которые в алгоритме отсутствуют (рис. 1).

Но такая система, принимающая решение, для разработчика выглядит как черный ящик, внутри которого что-то происходит. Разработчик не знает, почему и как система принимает те или иные решения. Да и сама система не знает, почему она принимает такие решения, находясь за занавесом неведения [Искусственный интеллект 2019: 157]. Такая модель ИИ решает задачи, но не знает, как она это делает. То есть модель ИИ в принципе лишена рефлексии. Но пока разработчика интересует другой вопрос – умеет ли модель ИИ работать с непредсказуемостью? Возникает вопрос – умеет ли сам человек работать с непредсказуемостью и как он это делает?

В таком случае приведём пример, взяв в качестве аналогии модели непредсказуемости и неопределенности модель регулирования поведения водителей и пешеходов на дорогах большого города. Фактически город выступает реальным полигоном-моделью системы с большой степенью непредсказуемых ситуаций и последствий. В силу сталкивания многих тысяч воль и действий жителей городское пространство обитания становится реальным полигоном, на котором уже апробируются модели ИИ (например, система умных светофоров, система наблюдения и штрафов, использование беспилотников в качестве умных средств мониторинга дорожного движения и т. д.).

Каждый водитель, едущий в машине, не знает другого водителя в другой машине и не знает, как тот себя поведёт. Но и этот, и другой водитель,

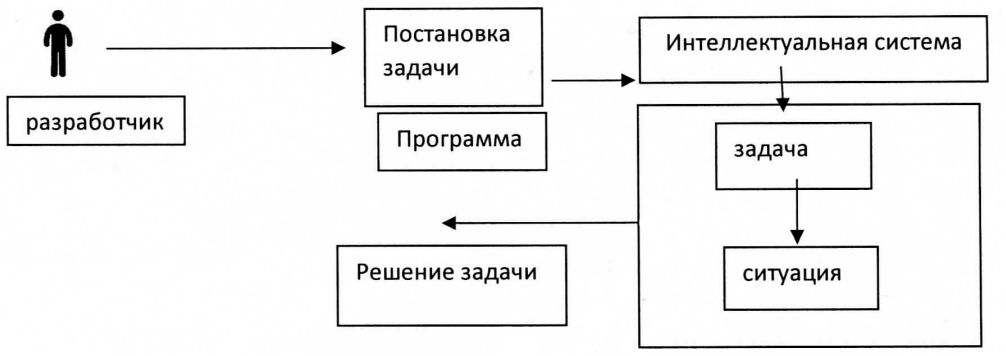


Рис. 1. Модель ИИ.

допускают, что все водители знают правила дорожного движения (далее – ПДД), то есть нормы, разработанные заранее в качестве всеобщего, принятого всеми *нормативного регулятора поведения* на дорогах. ПДД вводятся для того, чтобы минимизировать степень непредсказуемости поведения водителей и пешеходов на дорогах. ПДД вводятся для всех участников движения в качестве дополнительного над физическим полем – нормативного поля, в виде *системы знаков*. Но именно это поле, знаковое, становится главным регулятором, более важным, нежели само реальное пространство передвижений. Ориентируясь в нормативном поле, заданном через ПДД и систему знаков, человек может передвигаться по физическому полю. Если он не умеет ориентироваться в нормативном поле или пренебрегает его правилами, то он рискует попасть в аварию. Таким образом, водитель или пешеход не только видят дорогу и другие машины, но также видят знаки, регулирующие движение, читают их, видя в них заложенные правилами нормы. Одни водители тем самым допускают, что и другие водители будут выполнять правила, закодированные в этих знаках. Тем самым и водители, и пешеходы начинают видеть это нормативное поле, посредством которого они видят и физическое поле. Тем самым жители города, передвигающиеся по его улицам, видят не просто передвижения людей и транспорта, но они видят *регулируемое передвижение с помощью знакового опосредования.*

Таким образом само наличие этого нормативного регулятора, вынесенного во вне водителей, во вне машин, становится условием движения на дорогах и условием вхождения в это пространство передвижения. Без них будет стихия, хаос и непредсказуемые последствия (рис. 2).

Если перевести эту ситуационную модель в область разработок ИИ, то при разработке моделей ИИ необходимо учитывать встраивание в них подобных нормативных систем, дабы эти модели «умели» предсказывать и действовать в ситуации непредсказуемости. Если подобная нормативная система будет встроена в модель ИИ, то поведение последней становится как бы «разумным». Или нет?

В этом допущении скрыто ключевое условие: соблюдение правил (шире – норм) становится условием безопасности и в целом функциониро-

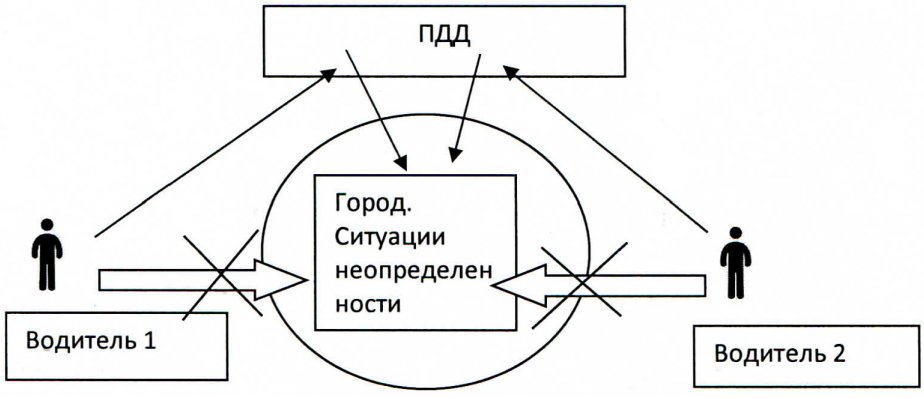


Рис. 2. Модель города как аналог ИИ.
Модель поведения в нормативной рамке в ситуации неопределённости

вания всей системы, здесь – системы дорожного движения. Если водители и пешеходы не соблюдают правила, то происходят аварии и катастрофы. Заметим, что фактически все крупные техногенные аварии (включая Фокусиму и Чернобыль, ставшие джокерами в своё время) были рукотворными, связанными с нарушением правил безопасности.

Если перейти к теме ИИ в виде разработок беспилотников, новых участников ДД, то его создатели вынуждены закладывать в их модели как ПДД, так и все возможные сценарии поведения разных агентов на дорогах (машин, пешеходов и др.). Но возможно ли заложить в модель ИИ *весь репертуар непредсказуемости* поведения пьяных водителей и нерадивых пешеходов? Пока нет. Поэтому появились первые случаи аварий с участием беспилотных машин[162].

Заметим, что по большому счету сам въезд в город, само вождение в городе (см. рис. 2) возможно лишь при условии знания водителем ПДД и сдачи соответствующих экзаменов. Что в мире и делается, хотя люди сами же эти правила постоянно нарушают. Но виноваты здесь не правила и не машины, а люди, позволяющие себе действия, не соответствующие этим правилам. Но само знание ПДД становится пропуском вхождения в большой город.

В таком случае создание ИИ и действие с ним должно выстраиваться по такой же схеме. Параллельно разработкам ИИ создаются и правила, этика для ИИ, а потом уже выстраиваются манипуляции и действия с ним. Знание и принятие норм выступает *условием* управления автоматическим транспортным средством. Знание и принятие нормы и прохождение соответствующей аттестации на знание нормы выступает условием вхождения в ситуацию со многими степенями непредсказуемости, с которыми сталкиваются автономные субъекты.

Главным фактором этой неопределенности выступает наличие действий многочисленных акторов, игроков, агентов, в данном случае водителей и пешеходов, вступающих в контакт. Но у них у всех в таком случае должен быть пропуск на вхождение в ситуацию неопределённости и непредсказуемости.

В таком случае ключевой проблемой при разработке ИИ становится не сама по себе разработка ИИ как умного технического устройства. Проблема не в технике, не в чёрном ящике, не в том, что и как делает нейросеть, выдающая разработчику непредсказуемые решения. Проблема в самом человеке, принимающем или нарушающем принятые им самим правила. Если их толковать шире – проблема в сохранении или размывании границ этоса, то есть глубинных устоев обитания, принятых человеком для своего собственного сохранения. Если он не будет удерживать и воспроизводить этос, то устои рухнут. Если сами горожане будут нарушать ПДД, принятые ими же, то город погрузится в хаос.

---

[162] В 2014 году автомобиль-беспилотник Google столкнулся с женщиной, находящейся в инвалидной коляске с электрической тягой, прогоняющей метлой проходящих через дорогу уток. Беспилотник не мог *распознать* данный экзотический для него объект [Искусственный интеллект 2019: 157]. Кстати, в Германии впервые в 2016 году был предложен законопроект о создании первой правовой базы для автономных транспортных средств. В России также разрабатываются изменения для ПДД в связи с беспилотниками. В 2016 году в США был принят первый документ, регламентирующий федеральную политику в отношении беспилотников.

Достаточное ли это основание для определения ИИ? То есть такое, согласно которому модель ИИ может считаться разумной, если умеет работать с *правилами поведения и предвидеть непредсказуемые ситуации?*

Но проблема заключается в том, что сам человек нарушает собственные нормы и правила. Человек есть не тот, кто принимает разумные решения, а тот, кто их же и нарушает. Или тот, кто постоянно выходит за собственные границы. И это есть его собственное онтологическое качество – выходить за собственные границы и обустраивать новые сферы обитания. А выходит он за границы не в силу родовой деструктивности, а потому, что он так устроен, будучи существом трансцендентным. Это его способ обитания – раздвигание границ, преодоление ранее положенной нормы, с тем, чтобы далее в новой реальности воссоздавать себя и свои границы уже в новых условиях.

Но в таком случае ИИ *невозможен в принципе*. Потому что граница, поставленная ранее, будет всегда нарушаться тем, кто эти границы поставил – человеком. Не успеет ИИ научиться работать с поставленными границами, завтра эти границы будут нарушены. Или в таком случае необходимо другое представление об ИИ: это такая программа, которая сама же и нарушает свой собственный алгоритм.

Но в таком случае главной проблемой становится не то, возможен ли ИИ или нет, заменит он человека или нет, какие интеллектуальные задачи он сможет решать в будущем. Главной становится сугубо этическая проблема – проблема границ этоса человека, способа и мира его обитания, проблема им же поставленных границ и того, как регулируется удержание этих границ. В данном случае речь идёт о границах человека и ИИ, человека и умной машины. То есть не проблема техническая, математическая, инженерная, информационная, а проблема сугубо гуманитарного, этического свойства. Здесь и рождается собственно этическая, шире – антропологическая, проблематика. То есть проблема границы человека и техники.

Теория вагонетки или этический тупик

Необходимо признать, что в основание своих рассуждений разработчики-инженеры закладывают допущение, что, учитывая рост непредсказуемых ситуаций, необходимость как-то встроить в свои модели ИИ этическую составляющую, связанную с безопасностью человека, возникает задача решить этическую проблему так же, как решаются формально-логические и математические задачи. Тем самым допускается, что в программу, в алгоритм ИИ реально заложить этические нормы, которые понимаются так же, как математические формулы. Иными словами, программу можно обучить этическому поведению. Допускается, что этическому поведению можно обучить и нейросеть. Допускается, что этическую норму или правило (как правильно поступать с точки зрения морали?) можно обосновать формально-логически, что машину можно обучить поступать этически так же, как и научить её совершать вычислительные операции. Значит, допускается, что этический выбор так же рационален и логичен, как решение математической задачи.

Такие допущения породили развитие так называемой теории вагонетки, или «вагонеткологии» (trolleyology), внедрённой в практику этических дискуссий с лёгкой руки разных западных авторов [Эдмондс 2020]. Ситуация или проблема вагонетки описывается мысленным экспериментом.

Представим себе, вы стоите на пути. По нему несётся неуправляемая вагонетка, у которой отказали тормоза. Вы видите, что впереди на пути лежат пятеро человек, они не могут убежать с пути и рискуют погибнуть под вагонеткой, поскольку они привязаны к рельсам. У вас есть возможность перевести стрелку, чтобы вагонетка пошла по другой ветке. Но на ней лежит один человек, который тоже не может уйти с пути, будучи привязан к рельсам. Вы можете перевести стрелку, но он погибнет, хотя пятеро останутся живы. Ваши действия?

Мне кажется, ситуация слишком надумана. Это, разумеется искусственная модельная ситуация, каковых много строится при разработке моделей взаимодействия людей и техники, попадающих в ситуации с большой степенью рисков и неопределённости. Подобные модели полезны для развития практик мысленных экспериментов. Но по сути ситуация, выстроенная в этой модели, предельно цинична. Она предлагает участникам подобных мыслительных экспериментов поместить себя хотя и в придуманные, но нечеловеческие условия. И представить – способен ли я убить одного ради спасения пятерых? Выбор однозначно тупиковый. Выбора здесь вообще быть не должно. Ведь этика – не арифметика. Нельзя там ставить минус, а там – плюс. Так бывает на войне – когда генерал жертвует в одном месте батальоном, чтобы в другом месте сделать прорыв. Но это на войне. Жизнь нельзя строить по стратегиям войны. Если уж строить модели, то такие, которые бы не загоняли человека в изначально тупиковую ситуацию, в которой любое решение – заведомо убийственное. Человек заведомо ставится в нечеловеческую зависимость от тупой вагонетки и необходимости делать выбор, однозначно и одинаково убийственный. Ведь непричинение вреда другому, и тем более невозможность спасения одного ради смерти другого – это непреложное этическое правило. Оно обязательно к исполнению.

Д. Эдмондс приводит в своей книге 10 таких ситуаций с вагонеткой, в которых ставится вопрос о выборе между тем, пустить ли вагонетку по пути, и ты задавишь пятерых, или пустить по другому пути, и ты задавишь одного. Также он приводит различные примеры выбора, с которыми столкнулись врачи при лечении больных, с которыми столкнулись военные на фронте во время боевых действий, спасатели во время чрезвычайных ситуаций. Но заметим, что все примеры, которые приводит Эдмондс, и на которых строится теория вагонетки, взяты из областей медицины, войны, происшествий и чрезвычайных ситуаций, катастроф. То есть примеры заведомо чрезвычайные. Но жизнь человека нельзя рассматривать как сплошное происшествие, как постоянную войну и бесконечную болезнь. Жизнь человека в норме не может рассматриваться как жизнь на войне или в больнице. Авторы же теории вагонетки строят свои мысленные эксперименты сугубо на примерах чрезвычайных и далее пытаются распространить их на жизнь человека в норме, в нормальных повседневных условиях, когда нет войны, когда никто не болеет и когда нет пожаров и наводнений.

Тем не менее такая с позволения сказать прикладная этика или даже моральная философия пустила корни в западном философском дискурсе и вошла в списки различных опросов. Например, Д. Бурже и Д. Чалмерс не преминули вставить в свое социологическое исследование вопросы из теории вагонетки [Bourget, Chalmers 2021]:

«Задача вагонетки. Ехать прямо, тогда наедешь на пятерых, или переключить стрелку на другую линию, тогда наедешь на одного? Ваш ответ».

Какие ответы были получены?

Переключить – 63,42% опрошенных.

Не переключать – 13, 31%.

Другое – 24,88%.

Удивительно. В какую циничную ситуацию ставятся респонденты, на какой вопрос им предлагают ответить! И несмотря на все моральные муки и угрызения все же 13,31% ответили так, что не надо переключать стрелку. Мол, пусть едет вагонетка и давит всех пятерых. И авторы опроса, и респонденты допускают вообще саму возможность рассуждать по поводу таких ситуаций. Это и есть этическая проблема? Куда делся Кантовский ригоризм? Куда ушла великая метафизика нравов? Моральный долг? Нравственный поступок? Самопожертвование ради другого? Почему в этой вагонеткологии не допускается третий вариант – тот, кто находится у стрелки или тот, кто сидит в вагонетке вообще вынужден себя просто взорвать, если движение неминуемо. И прямо, и направо – одинаково пути убийственные. А поэтому, если ты находишься в такой ситуации, ты жертвуешь собой, поскольку иного пути нет, и только ты сам ценою своей жизни и можешь остановить эту страшную вагонетку, которая становится слепым орудием убийства. Но авторы вагонетковедения почему-то сидят и рассуждают, но никак не допускают и мысли о таком варианте.

В книжке Д. Эдмондса рассуждения о вагонетках и иных ситуациях морального выбора перемежаются рассказами из истории моральной философии. Интересно, познавательно. Но мысленный эксперимент с вагонеткой, мне кажется, перечеркивает эту увлекательность и ставит вообще крест на проблеме нравственности. Этика исчезает под грудой обломков мчащейся слепой вагонетки. Последняя становится метафорой всего происходящего тренда: за мчащимся поездом технического прогресса мы перестали видеть человека, перестали понимать самих себя. Увлекательная игра с нейросетями несёт нас в тупик, раздавливая человека.

В то же время налицо явно выраженная тенденция, описанная в литературе, состоящая из двух направлений: с одной стороны, машинизация человека, превращение его в функциональное устройство, редукция его поведения до функции, с другой стороны, уподобление машины человеку, превращение её в некое человекоподобное устройство [Смирнов 2022б; Смирнов 2023а].

Подобная тенденция воплощается в том, что разработчики и исследователи допускают возможность моделей так называемых «моральных машин», в которых возможен учёт этических норм, как будто этические нормы могут быть включены в алгоритм поведения машины, вплоть до того, что проблема свободы воли может быть как-то учтена в моделях ИИ, если не сейчас, то не в столь отдалённом будущем [Карпов и др. 2018; Разин 2019].

Исследователи допускают, что обсуждать этику ИИ нельзя без обсуждения проблемы свободы воли [Разин 2019: 58]. Но они допускают в своих рассуждениях и способность ИИ к научению этическому поведению, способность ИИ принимать этически взвешенные решения. Вводится очередная метафора – «моральные машины». Разработчики и исследователи всерьез

допускают, что ИИ можно научить не только решать нетривиальные задачи, но и научить этическому поведению [Moral Machine 2023; Bostrom 2011][163]. Это означает, что этическое поведение можно описать так же строго, математически точно. Последнее, разумеется, связано с выше приведёнными примерами из теории вагонетки. Коль скоро этическое действие так же можно рассматривать, как поведение рациональное, логически выверенное, и его можно рассчитать и построить по этому поводу алгоритм, то в принципе машину можно научить совершать этически обоснованные поступки.

Исследователи допускают, что если человек есть субъект, обладающий свободой воли, с которым связана его моральная и правовая ответственность, то именно эти качества должны быть заложены в модели ИИ как параметры. Это значит, что модель ИИ должна быть способна реагировать на случайные (не входящие в алгоритм) события, неповторимые ситуации, уметь принимать адекватные решения, быть способной на нравственные действия. Но коль скоро этому всему человек сам научается поэтапно, в процессе культурного развития в рамках онтогенеза, то тем самым ИИ надо этому тоже учить. Поэтому необходима модель поэтапного обучения ИИ этическому поведению, так же, как есть поэтапное обучение человека [Разин 2019]. Полагаю, что упование на создание подобных моральных машин, способных на этическое поведение – тоже тупик, ещё одно безмерное увлечение разработчиков и исследователей.

Этика для сферы ИИ

Итак, мы не будем идти по пути вагонеткологии и по пути создания моральных машин. Зададим себе вопрос еще раз – где и когда рождается этическая проблематика в ситуациях с разработкой моделей ИИ? Ответ такой: там и тогда, когда разрабатываются правила и нормы, регулирующие взаимодействие людей в ситуации разнообразия воль и действий, непредсказуемости и высокой степени сложности и рисков. Этика рождается не в инженерии, а в ситуации множественности свободных воль и действий. В самом ИИ этики нет и быть не может. ИИ научить этике нельзя по определению, поскольку ИИ не является субъектом действия. Этика выступает регулятором, с помощью которого регулируются отношения между людьми, а не между человеком и машиной. Здесь – ключевая проблема доверия не к ИИ, а к человеку, автору ИИ.

Этика рождается тогда, когда мы сталкиваемся с рисками и угрозами, связанными с разработкой и внедрением ИИ в повседневность, в результате чего растёт степень непредсказуемости, а поэтому необходимо выработать кодексы и конвенции, регулирующие деятельность людей в сфере ИИ. Таких конвенций выработано в последние годы уже много [Кодекс этики 2021; Принципы этики 2021; Proposal 2021; AI Principles 2017; ЮНЕСКО 2021].

---

[163] В публикациях, посвященных моральным машинам, при этом обучение этическому поведению ставится в том же духе, что и в теории вагонетки. Например, для испытания беспилотных машин ставятся задачи типа: какое из двух зол выберет самоуправляемый автомобиль – гибель двух пассажиров или пяти пешеходов? Обучение этическому поведению строится по той же теории вагонетки. Цинизм задачи закладывается в алгоритм поведения моральной машины.

Казалось бы, человечество всерьез задумалось о том, что пора осмыслять этическую проблематику при разработке умных систем и не увлекаться одними инженерными задачами. То есть пора вспомнить о человеке.

Но зададим себе два вопроса:

1. Какая модель человека при этом закладывается в эти принятые проекты конвенций и этических деклараций?

2. Не стоит ли за разработкой этих конвенций и деклараций всего-навсего маркетинговая стратегия? Может, авторы этих конвенций, в числе которых крупные корпорации (в России это компания Сбер), стремятся просто убедить клиентов в том, что ИИ не обманет, он дружественный, его не надо бояться? Но тогда это всего-навсего маркетинговый ход, который продавец той или иной умной технологии использует для того, чтобы убедить клиента и покупателя купить его замечательную технологию.

Что касается первого вопроса, то пока ситуация выглядит следующим образом.

При разработке фактически всех этических конвенций и деклараций их авторы делают два фундаментальных допущения:

– человек не меняется, он всё такой же, каким был и тысячу лет назад,

– история цивилизации есть история освобождения человека, становления его автономной личностью, он достоин лучшей, то есть комфортной и безопасной жизни. Последнее ему и должен обеспечить в том числе ИИ.

При этом допускается, что это благополучие и комфорт требуется защитить от новых рисков, связанных с развитием умных технологий. Всегда при разработке и внедрении сложных инженерных систем требованиям безопасности уделялось серьёзное внимание, будь то строительство АЭС, ракетных комплексов или разработка ИИ. В данном случае она и должна быть выработана и принята цивилизованным сообществом.

Но проблема в том, что ИИ, внедряясь в реальность, в повседневность человека, реально начинает менять и его устои жизни, его этос, его среду обитания. И речь идёт уже не о правилах поведения и технической безопасности, а о готовности самого человека отказаться от этих устоев. В силу чего отказ ставит под вопрос саму привычную норму человека, его представления о самом себе.

В этой связи задачей становится не сама по себе разработка ИИ и не только выработка мер защиты человека от возможных угроз и рисков, связанных с внедрением ИИ, а построение новой онтологии человека в новой реальности, новой гибридной среде и восстановление нормы человека в этой новой реальности, в которой необходимо перестраивать интерфейс человека и ИИ. Поэтому проблема не в том, чтобы выработать очередной этический кодекс для сферы ИИ, а в том, чтобы всякий раз восстанавливать этос человека, норму человека.

Стремление же разработать кодекс выступает фактически следствием редукции человека, сведения его к отдельной особи, капризной и зависимой от умной техники. Чтобы умная техника не нанесла ему вред, вырабатывается система мер по защите слабого и беззащитного человека-индивида. Представление об атомарной личности, венце западной цивилизации, свелось к отдельной, зависимой от умной техники особи, отдельному индивиду, испытывающему соблазн от разного рода желаний и потребностей,

которыми представлена окружающая жизнь, воспринимаемая им в виде большого и яркого гипермаркета.

По большому счёту эти кодексы защищают человека как обывателя жизни. А задача заключается в том, чтобы человек оставался субъектом, способным в себе вырабатывать систему защиты от капризов и соблазнов, связанных с потреблением умной техники. Строго говоря, именно человек, отказывающийся от бытия как нормы, и стремится защитить себя от придуманного им же самим интеллектуального монстра, ИИ.

Что касается второго вопроса, то здесь ситуация сложнее. С одной стороны, принимаются декларации, согласно которым вводится запрет на разработку и внедрение определенных систем ИИ. Например, в Проекте Регламента Еврокомиссии от 2021 года введен так называемый класс запрещенных практик ИИ:

«Искусственный интеллект должен быть не самоцелью, а инструментом, который служит людям с конечной целью повышения их благосостояния. Правила для искусственного интеллекта, распространенного на рынке Евросоюза или влияющего на граждан другим образом, должны ставить людей в центр. Люди должны доверять использованию технологий, знать о ее безопасности и соответствии закону, включая уважение основных прав <…>.

Запрещаются следующие виды практики искусственного интеллекта:

a) размещение на рынке, ввод в эксплуатацию или использование системы искусственного интеллекта, которая применяет сублиминальные методы вне сознания человека с целью существенного искажения поведения человека таким образом, что причиняет или может причинить этому человеку или другому лицу физический или психологический вред;

b) размещение на рынке, ввод в эксплуатацию или использование системы искусственного интеллекта, которая использует любое из уязвимых мест определенной группы лиц в силу их возраста, физических или психических недостатков, с целью существенного искажения поведения лица, относящегося к этой группе, таким образом, что это лицо или другое лицо причиняет или может причинить физический или психологический вред;

c) размещение на рынке, ввод в эксплуатацию или использование систем ИИ государственными органами или от их имени для оценки или классификации благонадежности физических лиц в течение определенного периода времени на основе их социального поведения или известных или прогнозируемых личных или личностных характеристик <…>.

d) использование систем дистанционной биометрической идентификации “в реальном времени” в общедоступных местах в целях обеспечения правопорядка, за исключением случаев…» [Proposal 2021][164].

С другой стороны, например, компания Сбер, принимая свой Кодекс этики для сферы ИИ, введя принципы этики, явным образом показывает, что за этим стоит всего-навсего маркетинговая политика [Принципы этики ИИ 2021].

---

[164] Proposal for a Regulation of the European Parliament and of the Council, Title II, article 5, p. 43-44. https://digital-strategy.ec.europa.eu/en/node/9756/printable/pdf

Можно упомянуть также Азиломарские принципы ИИ[165]. Но что здесь интересно? В основании этих принципов также вольно или невольно закралось уподобление поведения ИИ – поведению человека. Например, названы такие принципы:

«… *Схожесть ценностей.* Высоко автономные системы ИИ должны быть разработаны таким образом, чтобы их цели и поведение были схожи с человеческими ценностями на протяжении всей их работы.

Человеческие ценности. Системы ИИ должны разрабатываться и работать таким образом, чтобы быть совместимыми с идеалами человеческого достоинства, его прав и свобод, многообразия культур.

… Общее (всеобщее) благо. Суперинтеллект должен разрабатываться только для служения широко разделяемым этическим идеалам и на благо всего человечества, а не одного государства или организации» [AI Principles 2017].

Можно также назвать отечественный Кодекс этики для сферы ИИ, принятый в 2021 году на форуме по этике ИИ [Кодекс этики 2021]. Согласно этому Кодексу *человеко-ориентированный и гуманистический* подход является основным этическим принципом и центральным критерием оценки этичного поведения акторов в сфере ИИ.

Его расшифровка звучит следующим образом: «При развитии технологий ИИ человек, его права и свободы должны рассматриваться как наивысшая ценность. Разрабатываемые Акторами технологии ИИ должны способствовать или не препятствовать реализации всех потенциальных возможностей человека для достижения гармонии в социальной, экономической, духовной сфере и наивысшего расцвета личности, учитывать ключевые ценности, такие как сохранение и развитие когнитивных способностей человека и его творческого потенциала; сохранение нравственных, духовных и культурных ценностей; содействие культурному и языковому многообразию, самобытности; сохранение традиций и устоев наций, народов, этносов и социальных групп» [Кодекс этики 2021].

Далее в Кодексе приводятся уже известные принципы этики для сферы ИИ (непричинение вреда, ответственность, безопасность, поднадзорность и т. д.).

Возникает вопрос: кто и как контролирует соблюдение названных принципов? Каковы механизмы контроля, мониторинга соблюдения принципов? Кто и как несёт ответственность за несоблюдение принципов? Эти вопросы пока остаются открытыми. В принятых декларациях и кодексах вопрос о механизмах контроля и наказания вообще фактически отсутствует.

Заключение

Учитывая сказанное, пока необходимо констатировать следующее.

При разработке этических кодексов и деклараций сам человек не рассматривается как источник риска. Источник рисков почему-то видится в ИИ,

[165] Принципы были разработаны и приняты по итогам конференции разработчиков и исследователей в сфере ИИ, прошедшей в январе 2017 года в Азиломаре, США. На данный момент под ними поставили свои подписи свыше 3 500 ученых, разработчиков, предпринимателей и экспертов. Среди них Илон Маск, Стивен Хокинг, Рэй Курцвейл, представители Google, Apple, Facebook, IBM, Microsoft и т.д.

как будто он выступает субъектом действия. Сам же человек (и не только разработчик) не рассматривается как субъект изменений и источник рисков.

До сих пор сохраняется допущение сходства человека и машины, допускается аналогия и уподобление человека и ИИ.

В принятых декларациях и конвенциях присутствует больше маркетинга, нежели этики.

Предложенные декларации и кодексы для сферы ИИ не стали регулятивами, поскольку в них отсутствуют механизмы контроля и мониторинга соблюдения названных принципов и правил.

В разрабатываемых моделях ИИ доминирует допущение, что этический выбор совершается так же логически и рационально, как и решение математической задачи. А с этим связано и допущение, что ИИ можно обучить этическому поведению так же, как и научить решать задачи.

Возникает ощущение, что мы сами, люди, уже запутались в своих собственных допущениях и представлениях. Наверное, надо начинать вновь с чистого листа. И с самих себя.

ЭТИЧЕСКОЕ САМОРЕГУЛИРОВАНИЕ В СФЕРЕ ИИ: НУЖНА ЛИ СВОЯ КЛЯТВА ГИППОКРАТА АКТОРАМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА?

Т. А. Сидорова[166]

Все мы – пациенты, и время от времени, оказываясь в этой роли, рассчитываем на то, что люди, в руках которых оказалась наша жизнь, руководствуются врачебным долгом. Что это такое? Содержание профессионального долга многосоставно, однако, прежде всего – это внутренняя необходимость, свободное, без принуждения, принятие совокупности обязательств. Когда человек получает диплом врача, он принимает Клятву профессии, – как правило, её называют Клятвой Гиппократа, – не пациенту, а именно профессии, обретая таким образом новую идентичность и принимая профессиональный долг в качестве внутреннего законодательства. Суть профессионального долга врача в ответственности за жизнь человека. Но эта способность к исполнению долга возникает не просто от присоединения к прописанным в Клятве или Этическом Кодексе нормам, чего требуют институциональные правила, она становится результатом формирования внутренней убежденности, на основе которой принимаются профессиональные решения.

Клятва предполагает наличие священного, непререкаемого авторитета, заставляющего человека ориентировать свои действия и совершать поступки в соответствии с заявленными обязательствами. Это переведение внешней необходимости во внутреннее чувство долга, а клятвопреступление обозначает моральное падение. «Глубинным психологическим мотивом речевого акта «клятва» является потребность индивида в доверии со стороны адресата/-ов, которая и трансформируется в целевую установку, связанную с установлением искомого доверия» [Чесноков 2016: 110].

Таким образом, принесение Клятвы есть акт морально-психологической саморегуляции, он становится необходимым в ситуации, когда не внешнее принуждение, например, правовое, становится источником для надлежащего исполнения обязанностей, а внутреннее побуждение воли, преобразованное в горниле осознания своего долга.

И. Кант утверждал, что долг есть категорический (безусловный) моральный (а не договорной, юридический) императив, означающий внутреннее побуждение, которым обладает индивид как автономно, т. е. без принуждения, действующий субъект. Категорический императив – это «моральный закон во мне», бескорыстная ориентация воли свободного субъекта, действующего во благо, и благодаря этому стремлению в душе каждого человека общество обретает единство и целостность.

Врачебная деятельность и история внутреннего профессионального регулирования путём добровольного принятия «закона медицинского и никакого другого», более всего соответствовала духу долженствования, поэтому медицинскую этику стали называть деонтологией (от др.-греч. δέον «должное»), полагая, что врач ориентируется в своей деятельности на безусловное принятие профессионального долга, содержание которого в качестве

[166] © Сидорова Т. А. 2023.

универсальной основы устанавливалось Клятвой Гиппократа [Гиппократ 2001]. Практики давать клятву были широко распространены в античных и средневековых обществах, клятву давали, например, в суде, но через тысячелетия до нас дошли именно те, которые отражали непреходящую со временем потребность зафиксировать точку отсчёта в понимании добра и зла: при различении жизни как блага и смерти, болезни как воплощения зла, а в суде – при отличии истины от лжи. Клятва Гиппократа с течением времени в разных странах для врачебного сообщества стала профессиональным символом, она могла иметь разные модификации, дополнялась новыми принципами, но всегда сохраняла набор базовых требований, сформулированных в древности: обязательства сохранять верность профессии по принципу «не навреди», уважение к жизни, действовать во благо пациенту, сохранять врачебную тайну и др. [Сычев 2015: 97]. В мире современной медицины Клятва Гиппократа выполняет прежде всего символическую функцию, выступая исторической подосновой институционального регулирования и в видоизмененных вариантах по-прежнему сохраняет свою актуальность.

В других сферах традиция приносить клятвы утратила своё значение. Это связано как с общим изменением стиля жизни, для которого более характерны индивидуалистические жизненные стратегии, так и с изменением механизмов саморегуляции, ослаблением императивности моральных норм, трансформацией представлений о должном и подобающем. В. С. Соловьев указывал на три основания нравственности: стыд, жалость (доброта), благоговение перед высшим (признание морального авторитета) [Соловьев 1991: 198]. Если попытаться применить их к реалиям нравственной жизни современного человека, то, на первый взгляд, возникает впечатление, что они также утратили своё значение: стыдиться сегодня не принято, высшее начало в виде страха перед Божьим наказанием перестало служить универсальным регулятором внутренней жизни человека, проявление доброты рассматривается как слабость – ницшеанская точка зрения завоевала своё место.

Однако, уже в 60-е годы XX в. всё отчетливее стала осознаваться ситуация антропологического кризиса, ярко переданная Андреем Вознесенским в поэме «ОЗА». У произведения есть подзаголовок «Тетрадь, найденная в тумбочке дубненской гостиницы». Поэма была написана под впечатлением встреч с учёными в Объединенном институте ядерных исследований в Дубне. Приведём небольшие отрывки из поэмы.

«Мир – не хлам для аукциона.
Я – Андрей, а не имя рек.
Все прогрессы –
реакционны,
если рушится человек.

Не купить нас холодной игрушкой,
механическим соловейчиком!
В жизни главное – человечность –
хорошо ль вам? красиво? грустно?

Выше нет предопределения –

мир
к спасению
привести!
. . . . . .
Проклинаю псевдопрогресс.
Горло саднит от техсловес».

[Вознесенский 2022].

Кризисные состояния, переживаемые обществом, всё чаще заставляют пересмотреть матрицу современной нравственности и апеллировать к возвращению к философски обоснованным основаниям морали: к античному призыву во всём соблюдать меру или к триаде В. С. Соловьева. В этой триаде, по сути, утверждаются три необходимости: внутренняя мера, желание блага Другому и нравственно обоснованная целеориентирующая установка сознания, то есть то, что формирует качество воли, определяющей способность личности к доброму, правильному поступку и, в конечном счёте, к выбору.

Отражением потребности преодолеть кризисные явления, которые выражаются в возникновении множества моральных дилемм, является феномен «этизации» современного мира. Прикладная или регулятивная этика бурно развивается. Буквально её истолковывают как «приложение этических и моральных понятий и суждений к конкретным сложным, часто очень драматическим, и в любом случае – противоречивым ситуациям» [Коновалова 1998: 45]. Этизация воплощается в создании всевозможных Этических Кодексов, в актуализации старых и новых форм этики, в возникновении новых видов профессиональной этики, создании корпоративной, институциональной этики. Особую актуальность прикладная этика приобретает в сфере высоких технологий, где оказываются затронуты интересы общества, права и свободы личности, достоинство человека. Этические руководства разрабатываются параллельно созданию новых технологий, выступая способом «опережающего проживания» (Б. Г. Юдин) негативных или неоднозначных эффектов научно-технологической трансформации привычных социальных укладов, изменения того, что понималось в качестве нормы относительно распоряжения собственной жизнью и всеми её атрибутами: здоровьем, телесностью, внутренним миром, личными границами и т. д.

Хорошей иллюстрацией особенностей прикладной этики служит в современной версии медицинской этики, биоэтике, вопрос о заборе органов для пересадки другому человеку. Человек, с точки зрения определённых и культурно адаптированных критериев жизни и смерти, может считаться ещё живым, хотя у него, например, погиб мозг, а жизненные функции поддерживаются при помощи искусственных систем. По действующему сегодня в медицине критерию смерти человека при гибели мозга человек считается мёртвым. Но прикладная этика имеет дело с пограничными ситуациями, в которых принимаемые решения не имеют прецедентов, они сами противоречивы, поскольку трудно сказать, что лежит в основе того, что мы можем обозначить как норма, нормальное, положительное. Не зря А. П. Чехов говорил о том, что «Медицина – это ежедневное испытание на человечность».

С одной стороны, «этизация» – это сугубо положительный феномен, если в мире всё чаще прибегают к этическому регулированию «узких» мест и конфликтогенных зон. Регулирование применения новейших цифровых

технологий, таких, как системы искусственного интеллекта, укладывается в русло общемировой тенденции этизации. В этом случае мы должны видеть, что потребность регулирования говорит о том, что затронуты фундаментальные жизненные ценности, что технология выходит на границы манипулирования человеческой жизнью в её многоаспектной трактовке. Вот, как определяются ценности, которые необходимо защищать, в Кодексе этики в сфере ИИ: «Акторы ИИ должны принимать необходимые меры, направленные на сохранение автономии и свободы воли человека в принятии им решений, права выбора и в целом сохранения интеллектуальных способностей человека как самостоятельной ценности и системообразующего фактора современной цивилизации. Акторы ИИ должны на этапе создания СИИ прогнозировать возможные негативные последствия для развития когнитивных способностей человека, и не допускать разработку СИИ, которые целенаправленно вызывают такие последствия» [Кодекс этики 2021].

Однако, в явлении этизации есть и свои издержки. Дело в том, что мораль, нравственность, этика – понятия близкие, и мы вполне можем их употреблять как взаимозаменяемые. Не вдаваясь в подробности в их различении, мы понимаем, какая тонкая грань удерживает нас на стороне истинной нравственности от того, чтобы не перейти, например, к морализаторству, или, следуя неукоснительно самым этичным правилам, прослыть негодяем. Поэтому возникает картина: прописаны и торжественно приняты самые прекрасные Кодексы, а конфликтогенность не устранена. Долгое время при решении биоэтических дилемм считалось, что общественный консенсус – это согласие, достигнутое в дебатах, относительно того или иного случая, и есть успешный результат, приемлемое решение. Однако, факт консенсуса сложно установить простым убеждением, обеспечивающим временное принятие той или иной позиции, необходимо успешное дискурсивное оформление того, что будет считаться нормальным. С точки зрения соответствия реальному положению дел, легко увидеть, что принятые решения, например, по поводу отключения жизнеподдерживающих аппаратов, не универсальны, они не могут служить для всех основанием внутреннего законодательства, или, даже, наоборот, разрушают сами основы достижения такого согласия, поскольку отменяют важные ценности, которые обеспечивали социальное единство.

В. С. Соловьев говорил о том, что условием по-настоящему нравственного поступка должен быть свободный выбор этого поступка, вплоть до того, что человек имеет возможность иногда быть и безнравственным. За нравственностью всегда стоит внутренняя работа, выбор между добром и злом, результатом которой и является поступок [Соловьев 1988]. Настоящим критерием, отличающим то, в какую сторону направлено моральное действие, является моральный рост личности – той, которая совершает поступок, и тех людей, во имя которых он совершается. Следование принципу гуманизма, отстаивающего ценность человека, означает не разрушение, а его утверждение в качестве свободного, и на этом основании, нравственно поступающего существа – вот истинный критерий, на который можно ориентироваться в двусмысленных коллизиях, которые всё чаще встречаются на пути современного человека и, особенно, специалистов, в руках которых сосредоточены источники возникновения этих коллизий.

Но всё это имеет отношение к понятиям нравственности и морали. Формы прикладной этики создают правила внеморального регулирования, поскольку часто формальные правила ориентированы на то, чтобы привести в порядок не внутренний мир человека, который ищет то, где добро, а где зло, а его поведение, с тем, чтобы оно было функциональным, например, соответствовало целям существования корпорации, бизнеса, институции и т. п.

Нет ничего более сложного и многообразного, чем моральная жизнь в обществе, поэтому ещё одно отличие морали от права заключается в том, что это чрезвычайно гибкое регулирование. Когда мы составляем перечни этических норм, так или иначе начинаем приближать по своему качеству этическое регулирование к правовому, и чем больше этих перечней, тем они ближе. В общем-то, ничего плохого в этом нет, однако сам внутренний посыл свободного, нравственного принятия нормы истончается и перестаёт в какой-то момент быть действительным мотиватором действий и выбора индивида. Поэтому этическое регулирование – это хорошо, но до определённой меры, пока оно не превращает этику в эрзац права. Тогда, как В. С. Соловьев предупреждал об опасности переписывания морали по правовому образцу, поскольку он видел в этом утрату человеком способности свободно выбирать добро [Соловьев 1988]. Этическое регулирование в условиях транзитивного и распределённого характера профессиональных отношений должно строиться на основах саморегулирования. Моральное сообщество среди тех, кто разрабатывает, применяет и продвигает системы ИИ, формируется самим предметом регулирования, коими являются новые технологии с непредсказуемыми социальными и антропологическими импликациями. Всё, что разрабатывается, обещает принести благо, или, по крайней мере, обязуется минимизировать негативные последствия, гарантировать потребителю безопасность. В этой ситуации важно удержать опорные критерии различения добра и зла, чтобы этически оправдать цели, которые ставят перед собой специалисты, изменяющие нашу жизнь с помощью новейших технологий. Поэтому при этической саморегуляции в сфере ИИ важно не упускать из виду ступень индивидуальной ответственности, предусматривать механизмы для её формирования, для поддержки приверженности гуманистическим, т. е. ориентированным на человека, принципам.

Сегодняшний этап развития медицинской этики также связан с появлением новых технологий, которые ставят врача в ситуацию трудного выбора: между поддержанием жизни и эвтаназией, абортами и искусственным созданием эмбрионов, сохранением целостности человеческого генома и его редактированием и т. д. Так появляются принципиально новые ситуации, когда врачи однозначно не могут следовать основополагающим нормам, заложенным в Клятве Гиппократа. Традиционные отношения врача и пациента изменяются в дистанционной медицине, искусственный интеллект всё больше замещает врача в организационных вопросах, что принимается как безусловное благо, но возникает перспектива, что ИИ заменит врача при принятии решений. Телемедицинское консультирование принципиально востребует антипатернализм, видимо, считая, что в цифровой медицине не так важна «отеческая забота», сколько «понимающий диалог», ассоциируемый с антипатерналистскими коллегиальными отношениями.

Приведём иллюстрацию, взятую из рекомендаций для врача в телемедицине, которая названа «главной проблемой при переходе в онлайн-кон-

сультирование»: «Одна из самых сложных задач, с которой сталкиваются врачи в телемедицине, связана с преодолением патернализма. Многие специалисты не привыкли к партнерским взаимоотношениям с пациентом, им тяжело дается этот переход. Отсюда возникают проблемы при общении в чате, где деликатность должна быть еще выше, чем при личном взаимодействии. Нельзя критиковать и оценивать образ жизни пациента, даже если за этим кроется попытка побудить к лечению. Доктор должен научиться чувствовать пациента через экран, чтобы подвигнуть человека заниматься своим здоровьем, не навязывая ему при этом чувство вины» [Кругликов 2022]. В основе этой рекомендации лежит определённый, клишированный подход, связанный с абсолютизацией принципа автономии пациента и даже неверным его пониманием. Почему эмпатия и врачебная забота должны непременно порождать чувство вины? Автономия защищает право человека распоряжаться собой в вопросах своего здоровья, и это не обязательно противоречит патернализму, таковы современные выводы ученых. Время тотального отрицания патернализма давно ушло в прошлое, сегодня более востребованным является концепт медицинской заботы [Лехциер 2019; Mol 2008]. Нет сомнения в том, что алгоритмы, создаваемые для ИИ, и те практики, на которых будут обучаться нейронные сети, также во главу угла поставят принцип автономии.

Здесь очевидна преемственная связь с принципами биоэтики, которые известны как «джорджтаунская мантра» Бичампа и Чилдресса: не навреди, делай благо, автономия пациента и принцип справедливости.

Надо сказать, что список принципов биоэтики разнообразен, однако, чаще всего в других формах прикладной этики, и в том числе, в этике ИИ, вспоминают и копируют эту четверку и из неё, в первую очередь, приводят принцип автономии. Следует заметить, что уже в 90-е годы прошлого века европейские биоэтики сместили значимость этого принципа и во главу угла поставили принцип уважения достоинства личности, включающий в том числе добровольное волеизъявление на медицинское вмешательство, о чём в основе своей и утверждалось в принципе автономии, более специфичном для американской, юридизированной традиции регулирования отношений в медицине.

В Кодексе этики в сфере ИИ, принятом в России в 2021 г., принцип автономии и свободы воли человека упоминается в ином аспекте – защиты субъектности человека ввиду её проблематизации в связи с тем, что системы ИИ всё чаще будут подсказывать решения человеку или даже за него их принимать. Автономия и свобода воли в данном случае совершенно справедливо обозначены в качестве значимой ценности, которая требует защиты, и это должно осознаваться акторами ИИ на всех этапах – от разработок до применения.

Обращаясь к новым способам вмешательства в человеческую жизнь, где она оказывается на грани искусственного и естественного, на грани сохранения или утраты человеком свойств и возможностей, которые формируют его видовую и личностную идентичность, на наш, взгляд, уместно использовать принципы биоэтики: уважение достоинства личности (в Кодексе ИИ говорится об уважении прав и свобод человека, но достоинство человека включает, помимо правового закрепления свободы его действий, его автономии, и представление о самоценности человека); не навреди (на-

шёл отражение в Кодексе ИИ в пункте «Непричинение вреда»); уважение к жизни в аспекте сохранения исключительного права человека на видовую идентичность; принцип добровольности (права пользователя решать и выбирать); принцип ответственного распоряжения личной, в т. ч. биомедицинской информацией (есть пункт «Безопасность работы с данными»); неприкосновенность частной жизни (нашёл отражение в пункте «Ответственное отношение»); справедливость (есть пункт о «недискриминации», однако понятие справедливости шире понятия ущемления прав, т. е. дискриминации) [Кодекс этики 2021]. Эти принципы коррелируют с положениями Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники в РФ [Бегишев 2020].

Исторически важнейшим для медицины был принцип «Не навреди!». История медицины – это одновременно история поиска средств избавления от болезней, а затем избавления от побочных эффектов их воздействия на организм. Поэтому девиз врача звучит: Никогда лекарство не должно быть горше болезни! Выдающийся нравственный пример привёл Н. И. Пирогов. Он развивал принцип «Не навреди!» практически: опубликовал анализ своих врачебных ошибок. Правило «Учиться на ошибках» стало нравственной максимой, частью медицинского образования. Эта максима весьма полезна и для разработчиков новых технологий, в том числе систем ИИ.

Формы и виды вреда для человека в настоящее время дифференцировались. Это и формы физического вреда (вред для телесного здоровья), и моральный, и психологический вред, вред для когнитивной сферы. Этика ИИ имеет дело с видами вреда, которые довольно сложно не только визуализировать, но и заставить признавать то, что в конечном счете воздействие таковым и является.

Утилитаристская калькуляция соотношения вреда и пользы, на которой основаны заключения об этической приемлемости новых средств или новых технологий в сфере ИИ, может оказаться малопродуктивной, поскольку вред в данном случае связан не просто с побочными эффектами от лекарственных и функциональных воздействий, а с ущербом для соблюдения суверенных границ внутреннего мира человека, для сферы его частной жизни, для психологической преемственности, для ментальных установок, базирующихся на убежденности в первичности свободы воли и основанной на этом субъектности индивида. Вред, наносимый, например, ментальной сфере, может вести к дискриминации и таким эффектам, которые в значительной мере заставят пересмотреть сложившиеся моральные и социальные порядки.

В настоящее время в целях контроля участия человека в клинических исследованиях налажена система этической экспертизы, рассматривается вопрос, добровольно ли привлекаются испытуемые, существует ли риск для их здоровья. Для оценки этих рисков разработаны шкалы. Значимый вред для здоровья здесь, конечно же, исключается. По поводу иных видов вреда для человеческой субъективности таких шкал и методик не существует и их разработка весьма проблематична. Эти виды вреда должны быть научно исследованы в междисциплинарных лабораториях. ИИ может влиять на процессы мышления, на взаимодействие с другими людьми, на принятие человеком решений, а также на различные социальные сферы: образова-

ние, культуру, науку, спорт, информационное пространство, повседневную жизнь, самостоятельную заботу о здоровье и др.

Поэтому регулирование в этой области имеет отношение не только к оценке непосредственного вреда и риска для пользователей, который может быть минимален или отсутствовать совсем, но также и к необходимости взвешивать социальные и антропологические эффекты применения полученных знаний на прикладном уровне. Допускать технологии к использованию, опираясь на критерии технической безопасности, значит упускать из виду самое важное с этической точки зрения – уязвимость человека как обладателя психофизического единства, распорядителя своим внутренним миром, суверена над собственным Эго [Сидорова 2020].

Заключение

К этическому Кодексу в сфере ИИ в России присоединилось более сотни компаний. Более того, подчеркивая важность этического аспекта в жизни корпораций, занимающихся ИИ, осуществляется медийная поддержка, в СМИ рассказывают о том, что принят такой Кодекс, что в нём есть человеко-ориентированные нормы и учитываются социальные риски и т. д. Вероятно, таким образом должна быть обеспечена общественная поддержка огромным вложениям, которые делает государство, развивая проекты ИИ. Но главное состоит в том, чтобы получить доверие от потребителей, именно это необходимо для продвижения создаваемых продуктов. Однако, опыт создания «хороших картинок» в современном мире настолько разнообразен, что уверения и массовое присоединение компаний-разработчиков к Кодексу ИИ, вряд ли убедят пользователей в том, что рядовые программисты действительно исповедуют высокие обязательства, под которыми подписались их работодатели.

Возвращаясь к сравнению с медицинской этикой, ещё раз укажем на то, что в медицине ответственность принимается индивидуально, а профессиональное сообщество формируется из тех, кто как веру принял на себя звание врача. Этическое регулирование в сфере ИИ занимает промежуточное положение между обычной профессиональной недеонтологической этикой и теми видами этики, которые имеют отношение к жизненно важным ценностям. И чем более вероятно будет создание вариантов сильного ИИ, с притязаниями на субъектность и одновременным возрастанием вероятности устранения или ослабления человеческой субъектности, тем очевиднее будет востребована модель деонтологической этики. Приобщение к нормам такой этики будет иметь перформативный характер, т. е. само по себе будет поступком, внутренним выбором индивидуально заявленных обязательств.

Отвечая на вопрос, поставленный в название статьи, сделаем заключение, что для этического регулирования в сфере ИИ уже хорошо то, что есть Этический Кодекс. Однако, для транзитивных и распределённых социальных отношений в профессиональных сообществах, занимающихся ИИ, важны такие элементы саморегулирования, которые обеспечивают приверженность общим задачам, обеспечивают микрофизику этой приверженности [Санников 2017].

Перформативные заявления, подобные Клятве Гиппократа (в досоветской России она именовалась «Факультетским обещанием») являются

механизмом инкорпорирования важнейших норм. Эквивалентом Клятвы в этике ИИ может быть формулировка выверенного набора принципов, которые, повторяясь как мантра, станут внутренним кодексом акторов ИИ. Если важно формировать экосистему доверия, то она должна начинаться с принятия моральных ценностей уважения человеческого достоинства, открытости, безопасности, справедливости и т. д. Путь к этому заключается в формировании этической культуры специалистов в сфере ИИ, в их способности брать на себя моральную ответственность при обращении к человеческой жизни и любым её аспектам. В этой связи этическое регулирование должно включать этическое образование будущих специалистов. Для этой цели очень полезны иллюстрации об истории злоупотреблений в биомедицинских исследованиях, о том, как в истории науки и технологий выбирались ошибочные стратегии, которые сулили значительные выгоды в борьбе с болезнями, а обернулись такими нежелательными явлениями, которые стёрли привлекательность обещанных выгод.

О МЕТАФОРИЧЕСКОМ ПОНЯТИИ «УМНОГО» УСТРОЙСТВА И ЕГО ПРИМЕНИМОСТИ В ФИЛОСОФСКО-АНТРОПОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

А. Г. Горбачёва[167], А. И. Пестунов[168]

Введение

Значимость философских и антропологических исследований, связанных с изучением трансформаций человека при взаимодействии со всевозможными техническими устройствами, не вызывает сомнений [Ардашкин 2018]. В рамках данного направления рассматриваются различные аспекты, в том числе, возможные риски [Аванесов 2021; Черникова 2012], тренды и прогнозы [Смирнов 2012; Смирнов 2013; Gorbacheva, Smirnov 2017], способы преодоления возникающих барьеров [Ефимов 2020] и пр. Особенную остроту этот вопрос приобрел в последние десятилетия, когда скорость совершенствования технологий стала настолько высока, что люди в течение своей жизни становятся свидетелями смены нескольких поколений вроде бы одного устройства. Например, можно рассмотреть цепочку «проводной телефон – радиотелефон – мобильный телефон – смартфон» и заметить насколько сильно в ней отличаются функциональные возможности первого и последнего устройства.

Описанное явление привело к тому, что современные тексты из различных сфер стали изобиловать словосочетаниями, в которых присутствует характеристика «умный» применительно к технологиям, устройствам, структурным образованиям и другим неодушевленным объектам. При этом как такового общепринятого определения этой самой «умности» на данный момент нет. Данное понятие является скорее метафорой, которая может иметь совершенно разный смысл в различных контекстах. Тем не менее, как правило характеристика «умное» применяется к устройствам, появившимся именно в последние десятилетия, и, следовательно, мы имеем основания выдвинуть гипотезу о том, что в эти самые последние десятилетия многие технические устройства стали приобретать некое общее качество, которое подталкивает людей к тому, чтобы называть их «умными».

Для того, чтобы называться «умным», устройство должно «уметь мыслить», однако убедительного ответа на вопрос, заданный А. Тьюрингом [Turing 1950] еще в середине прошлого века «Могут ли машины мыслить?» пока не получено. Сопоставляя данный факт и приведенные выше рассуждения, мы приходим к противоречию: с одной стороны, машины (читай – устройства) мыслить не способны, чтобы их можно было называть умными в смысле определения Ожегова, а с другой – уже сейчас многие устройства характеризуются как «умные». Попытке разрешить это противоречие и посвящена настоящая статья.

Очевидно, что разрешить описанное противоречие полностью практически невозможно в силу того, что массово используемая в настоящее время

[167] © Горбачёва А. Г. 2023

[168] © Пестунов А. И. 2023

характеристика «умное» применительно к устройствам является метафорической [Ардашкин 2018], включающей в себя все ассоциации, возникающие при произнесении этого слова, и воспринимается каждым человеком по-своему и стереотипно. Подобная проблема обозначена Тьюрингом [Turing 1950]. При попытке ответить на вопрос «Может ли машина мыслить?» он указал на отсутствие удовлетворительных определений понятий «машина» и «мыслить» и предложил свести ответ на этот вопрос к вопросу возможности прохождения теста.

Поэтому необходимо сформулировать некое компромиссное требование, которое будет свидетельствовать о желаемом результате – разрешении данного противоречия хотя бы частично. Причем этот результат должен иметь практическую пользу с точки зрения философии и антропологии. А именно, он должен помочь в исследованиях, касающихся влияния науки и технологий на человека, в том числе при взаимодействии человека с «умными» устройствами. Таким результатом может являться формулировка внятных, возможно нескольких определений «умности», позволяющих описать какие-либо относительно целостные классы технических устройств, которые в силу их свойств можно называть «умными».

В первой части статьи рассматриваются возможные характеристики, которые можно включить в определение «умного» устройства и критически анализируются на примерах конкретных широко распространенных технических устройств, в целом, как нам кажется, одинаково понимаемых большинством людей. По итогам этого анализа делается вывод о том, что практически любое потенциальное определение «умного» устройства можно подвергнуть критике в том плане, что оно будет справедливо для совершенно разных устройств, скажем так, разной степени «умности». Во второй части статьи формулируются несколько возможных определений «умного» устройства с учетом выводов, полученных в первой части, с последующим сопоставлением этих определений с различными техническими устройствами.

Подход и опорные точки исследования

Важным представляется изучение вопроса о том, действительно ли феномен массового появления «умных» устройств имеет под собой реальные основания. Есть ли что-то, присущее некоторым предметам, что заслуживает внимания и дает возможность назвать эти предметы «умными»? Насколько обоснованно мы называем некоторые предметы «умными»? Не является ли это просто маркетинговым ходом, чтобы придать дополнительную ценность продаваемым предметам в глазах покупателей? Какое принципиально новое качество появилось у неодушевленных предметов, дающее основание введения термина «умный»? Или же мы должны будем признать, что как такового свойства «умности» в общем-то и нет.

Основная трудность нашего исследования состоит в отсутствии прочных исходных посылок. В том, что мы пытаемся описать класс (или классы) «умных» устройств, не имея соответствующих определений. Мы видим лишь некие косвенные свидетельства того, что такие устройства должны существовать. Таким образом, наша задача состоится в том, чтобы попытаться сформулировать определение «умного» устройства, коррелирующее с массовым пониманием этого слова. Либо, хотя бы обозначить некие при-

знаки «умности» и понять, действительно ли сейчас появились устройства, несущие принципиально новый аспект «умности» им присущий.

Тем не менее, определенные опорные точки задать можно. Как минимум, мы уже сформулировали несколько базовых вопросов, определяющих цель исследования. Теперь попробуем понять, от чего можно отталкиваться, чтобы приблизиться к ответу на них.

Поскольку массовым термин «умное устройство» стал относительно недавно – не ранее, чем 10-20 лет назад, то искать «умные» устройства следует среди тех, которые стали массовыми именно в этот период. Следовательно, «точно не умными» устройствами и орудиями мы будем считать те, которые существовали издревле, в том числе, до появления электричества, электронных приборов и ЭВМ. К таким орудиям (именно этот термин для них подходит лучше) можно отнести лопату, обычный телефон, автомобиль без систем электронного управления, молоток, отвертку и пр. Перечисленным устройствам присуще такое свойство, что они в принципе не способны работать без участия человека. Они не могут как начать работу самостоятельно, так и продолжать действовать автономно уже после начала работы.

Второй тип устройств мы назовем «скорее всего не умные». Это электронные устройства, которые появились довольно давно, когда характеристика «умный» не была распространена. К таким устройствам отнесем калькулятор, автоматическую стиральную машину, чайник с автоматическим отключением при закипании, будильник и др. Например, калькулятор не может начать работу самостоятельно, но после нажатия соответствующей клавиши он уже работает автономно вплоть до получения результата вычислений.

Наконец, «скорее всего умными» мы будем считать такие устройства, как робот-пылесос, беспилотный автомобиль, системы распознавания речи с голосовым помощником, диалоговые системы и пр.

Наш поиск сродни раскапыванию кучи песка, на дне которой мы надеемся что-то найти, поскольку нам либо кто-то об этом сказал, либо нам кажется, и мы хотим это проверить. В итоге, раскопав эту кучу, мы либо обнаружим что-то, либо обнаружим пустое место. Более того, прогрессом можно будет считать даже не только однозначный ответ на этот вопрос (есть там что-то или нет?), но и, хотя бы, уменьшение размеров кучи с отбрасыванием тех ее частей, где точно ничего искомого нет.

Мы считаем наиболее удачным определением, которое можно эксплуатировать в философском и антропологическом анализе, то, где делается акцент на отчуждении предметного действия. Можно назвать устройство «умным», если для выполнения работы человеку либо вообще не нужно совершать предметного действия, либо это действие минимально и практически не связано (или связь не прослеживается) с получаемым результатом. Таким образом, подобные «умные» устройства являются олицетворением жизненного аутсорсинга.

Например, в этом смысле с точки зрения философского анализа нас гораздо больше интересуют автомобильные дворники с датчиком дождя, чем автомобиль, оснащенный различными «умными» устройствами и, в том числе, «умными» дворниками. Причина в том, что основное предметное действие автомобиля не отчуждается от человека – он как нажимал на педа-

ли и крутил руль, так и продолжает это делать, сколько бы второстепенных, пусть даже и «умных» приборов не было установлено на этот автомобиль.

Таким образом, объектом нашего исследования является не техническое устройство, а скорее человек, решивший вдруг называть какие-то устройства «умными». В отличие от Тьюринга и его последователей, мы задаем вопрос не с целью лучше узнать о потенциале технического устройства (может ли оно мыслить или быть «умным»?), а с целью понять человека. Почему этот Человек решил охарактеризовать некоторые технические устройства «умными»? Что заставило его это сделать? Заслуживают ли технические устройства называться «умными»? Еще один вопрос связан с появлением разума у машины.

Итак, базовыми вопросами, поиск ответов на которые формирует основную канву настоящей работы, являются следующие.

– Возможно ли выделить некое принципиально новое качество некоторых устройств, которое делает оправданным их классификацию как «умных»?

– Существуют ли свойства, по которым устройство можно отнести к «умным»?

– Как связаны свойство «умности» и устройство, характеризующееся как «умное»?

Методологические противоречия и попытки их разрешить

Рассмотрение устройств через их определение или фактические возможности. Если следовать приведенным выше рассуждениям, то робот-пылесос, скорее всего, можно считать «умным» устройством. Причем, если лишить его часть функциональных возможностей, то он перестанет быть «роботом-пылесосом» в своем изначальном смысле. Таким образом, данному устройству свойство «умности» присуще. Теперь рассмотрим автомобиль. Получается, что автомобиль «умным» не был, но после добавления ему ряда функциональных возможностей он станет таковым. Таким образом, мы делаем вывод о том, что, на первый взгляд, существует два типа устройств: первому типу свойство «умности» присуще, а второму – нет, но добавление ряда функциональных возможностей может сделать его «умным». Однако приведенное выше рассуждение оперирует скорее терминами, чем сутью. Здесь анализ проведен, отталкиваясь от термина «робот-пылесос», но мы же можем назвать это устройство неким «Устройством Икс». И опять задать вопрос – «умное» ли это устройство. Но уже не концептуально, а фактически. Можно ли это устройство лишить функционала или деталей, чтобы оно перестало быть умным? Таким образом, ответ на этот вопрос практически эквивалентен вопросу о наличии у устройства (машины) сознания, или сильного искусственного интеллекта по Серлю [Searle 1980]. Очевидно, что современные называемые «умными» устройства таким свойством не обладают, следовательно, изучение зарождения сознания у машины мы не рассматриваем. Свойство умности: свойство умности присуще устройству или это функциональные возможности, которыми можно наделить умное устройство. Например: в базовом варианте автомобиль явно не умный. Однако, прикрепив ему датчик дождя и связав его с дворниками, он вроде как поумнеет, поскольку будет самостоятельно включать дворники при начале дождя.

Является ли свойство «умности» присущим устройству или же это лишь на уровне функциональных возможностей. До каких пор нужно добавлять в смартфон функции, чтобы он стал умным?

В чем состоит предназначение устройства? Рассмотрим следующий мысленный эксперимент, действующим лицом в котором является кочегар, никогда не сидевший за столом и не знающий, что такое стол. Для этого кочегара деревянный стол будет не более, чем деревяшка причудливой формы, которую от бросит в печь точно так же, как и дрова. Ему и в голову не придет придумывать этой деревяшке какое-то специальное название наподобие «стола».

Настоящим примером показано, что истинное предназначение орудия или устройства может находиться далеко за пределами узкого контекста, в рамках которого некий человек взаимодействует с устройством. Приведем еще более радикальный мысленный эксперимент, демонстрирующий, что истинное предназначение устройства может быть вообще не определено.

Пусть в нашем распоряжении имеется смартфон, имеющий форму молотка. Для человека, понимающего, что такое смартфон, его форма будет вторична. Однако человек, который до этого видел только молоток, не сможет отличить этот смартфон от молотка. Значит, для него это устройство «не умное», а для пользователя смартфона – «умное», если мы сочтем смартфон «умным». При этом нет никакой гарантии того, что обычный молоток не обладает некими свойствами, в том числе «умными», не заложенными в него разработчиком, но которые могут быть познаны через науку. Скажем, речь может идти о неожиданных свойствах составляющих его микрочастиц.

В итоге, при выработке определения «умного» устройства и выделении каких-либо его свойств нам не удастся уйти от познающего субъекта, поскольку никогда не будет гарантий, что этот познающий субъект будет полностью осведомлен об истинном предназначении данного устройства.

Таким образом, определение устройства может быть дано с трех позиций:

– с позиции разработчика, заложившего в устройство определенный функционал, и пользователя, который эксплуатирует именно предусмотренные функциональные возможности устройства;

– с позиции пользователя, который видит в устройстве лишь разновидность тех объектов, с которыми он уже был знаком ранее;

– с позиции пользователя, который нашел принципиально иной способ эксплуатации данного устройства, который никак не закладывался его разработчиком.

Будем считать, что название устройства определяет его функционал, и в некотором контексте этот функционал одинаково понимается как разработчиком, так и пользователем.

Не является ли «умное» устройство просто объединением «не умных» функций? Требует рассмотрения еще один вопрос. Не является ли «умное» устройство всего лишь объединением функционала отдельных «не умных» устройств? Например, смартфон можно рассматривать как устройство, объединяющее в себе телефон, калькулятор, фотоаппарат, телевизор, игровую приставку и пр. Если ответ на данный вопрос окажется утвердительным, то мы должны незамедлительно сделать вывод о том, что «умное» устройство –

это не более, чем объединение обычных «не умных» устройств, и смысла награждать свойством «умности» какое-либо устройство безосновательно. Говорить же о синергетическом эффекте применительно к неодушевленным устройствам вряд ли приходится. Другими словами, сложно привести аргументы в пользу того, что объединение в одном устройстве нескольких функций приведет к появлению принципиально неожиданного эффекта.

Эволюция технических устройств: где появляется «умность»? Еще одним, на наш взгляд, заслуживающим внимания вопросом является поиск того момента, когда устройства стали «умными» (как всегда оговариваемся, что точного определения «умности» у нас нет). Например, рассмотрим эволюцию вычислительной техники. Упрощенно ее можно представить так: счеты – арифмометр – калькулятор – программируемый компьютер – искусственные нейронные сети. Очевидно, что счеты не являются «умными». Точнее, никаких разумных оснований считать счеты «умными» нет. В какомто смысле данный вопрос напоминает исследования о зарождении разума.

Таким образом, данный пример показывает, что даже с калькулятором (который вряд ли можно считать умным) возникает возможность того, что человек отдает способность вычислять (мыслить). Что же такого принципиального есть у смартфона, чего нет у калькулятора? Является ли смартфон просто многофункциональным калькулятором? Если так, то достаточно перечислить все его функции и рассмотреть их отдельно. Или же есть нечто качественное, что делает его «умным», а у калькулятора этого нет. Этот вопрос тесно связан с вопросом Тьюринга о возможности машины мыслить, поэтому вопрос о зарождении «умности» у устройств мы оставим за рамками настоящей статьи, а будем рассматривать устройства сами по себе без привязки к их эволюции.

Определение «умности» через внешние проявления и человеко-машинный интерфейс. Что интересует обычного человека, использующего «умное» устройство? Ответ – его функциональность. В связи с этим, имеет смысл говорить именно о функциях и человеко-машинном интерфейсе, не принимая в расчет внутреннюю структуру. Наличие датчиков, процессора, запоминающего механизма не имеет значения. Скорее, мы должны определять «умность» через внешние проявления, т.е. в терминах человеко-машинного интерфейса. При этом следует понимать, что существуют два типа интерфейса: человеко-машинный интерфейс и интерфейс между техническим устройством и окружающей средой. Например, человеко-машинный интерфейс робота-пылесоса состоит из элементов управления, позволяющих настроить и запустить его, а интерфейс с окружающей средой состоит из датчиков, позволяющих ориентироваться пылесосу в пространстве при выполнении своей работы. Указанные два интерфейса имеют принципиально разное значение с философско-антропологической точки зрения. Человеко-машинный интерфейс для взаимодействия с таким устройством должен быть существенно более прост, чем непосредственное решение задачи. Будь то физическая или интеллектуальная работа. «Умное» устройство призвано «забрать» часть работы себе.

Более того, просто само по себе наличие таких составных частей, как процессор или модуль памяти не является гарантией того, что обладающее ими устройство приобретает некое принципиально новое качество, дающее нам основание называть его «умным». Причина данного суждения состоит

в том, что и процессор, и модуль памяти – это не более чем комбинация неких элементов, взаимодействие которых с достаточной полнотой описано законами физики. В частности, транзисторы – это, грубо говоря, всего лишь маленькие лампочки, которые явно не претендуют на то, чтобы их называли «умными». А на заре развития компьютерной техники именно лампочки и выполняли роль транзисторов.

Таким образом, определение понятия «умного» устройства не должно никак опираться на его внутреннюю структуру, в том числе на наличие внутри него датчиков, процессора, запоминающего устройства и даже физических манипуляторов самих по себе. Определение «умного» устройства должно даваться через человеко-машинный интерфейс, в который при необходимости могут быть включены физические манипуляторы. Вероятно, имеет смысл говорить о способности технического устройства выполнять те или иные функции через его человеко-машинный интерфейс.

О невозможности точного разделения интеллектуального и физического труда. Если устройство «не умное», то оно всегда будет требовать присутствия человека, поскольку даже примитивный, на первый взгляд, физический труд, тем не менее, требует некоторой интеллектуальной деятельности. И только если соединить в устройстве и физику, и интеллект, возникает возможность отчуждения этого труда.

На наш взгляд, разделять физическую и интеллектуальную деятельность некорректно. Возможно, имеет смысл говорить о соотношении этих видов деятельности при выполнении различных видов работ, но заявлять, что определенный вид работы требует только физической активности, нельзя. Например, даже при переноске грузов или перекапывании земли требуется некоторая интеллектуальная деятельность для выбора траектории перемещения, прилагаемой силы или положения тела.

Определение «умности» не должно зависеть от человека. При взаимодействии людей друг с другом, в принципе, можно сказать, что один из них умнее другого в каком-либо смысле. Следовательно, определение «умный» так или иначе зависит он субъекта, который дает такую характеристику. Есть вероятность, что для «более умного» человека некоторые считаемые «умными» «менее умным» будут считаться не «умными». Отсюда получаем две возможности определения «умности» устройства: в зависимости от субъекта-человека и без нее. Поскольку характеристика «умное» применяется к устройствам массово, то определение не должно зависеть от степени интеллектуальности человека.

Критика возможных определений «умного» устройства

В настоящем разделе мы рассмотрим несколько возможных определений «умного» устройства и покажем, что для каждого из определений можно привести аргументы, убеждающие в неудачности того или иного определения.

Потенциальное определение 1. Устройство является *умным*, если в процессе его работы не требуется участие человека. Необходимость участия человека. Очевидным свойством, которое можно оценить у всех перечисленных устройств, является их способность/неспособность работать без участия человека. Причем можно оценить не только на уровне да/нет, но и степень необходимого участия человека. Например, при использовании ло-

паты человек должен участвовать полностью, а в случае со стиральной машиной – только на этапе загрузки белья, выбора программы и запуска. Если сравнивать компьютер и калькулятор, то принципиальная разница состоит в том, что компьютер может выполнить серию из нескольких вычислений по одной команде человека, а при использовании калькулятора участие человека требуется при каждом вычислении.

Потенциальное определение 2. Устройство является *умным*, если оно оснащено датчиками, сообщающими ему состояние окружающей среды. Рассмотрим следующее техническое устройство – автомобильные дворники с датчиком дождя. Данное устройство, как минимум, претендует на то, чтобы называться «умным», поскольку справляется со своей задачей по поддержанию чистоты лобового стекла без вмешательства человека. Однако, как только данное устройство лишается датчика, то оно становится точно не «умным». Аналогичные рассуждения также справедливы и для чайника с автоматическим отключением при закипании: убрав из данного устройства датчик интенсивности пара, мы делаем его точно «не умным». В то же время, очевидным является то, что сами по себе датчик дождя и дворники или датчик интенсивности пара и чайник не являются «умными». Более того, если к дворникам присоединить датчик интенсивности пара, а к чайнику – датчик дождя, то «умности» рассматриваемым устройствам не прибавится. Таким образом, мы делаем вывод, что свойство «умности», вероятно, может появляться при вполне конкретном сочетании не «умных» устройств. Само наличие датчика не является достаточным, чтобы сделать устройство «умным». Требуется не датчик, а способность реагировать на окружение. Скорее всего, наличие датчика следует из этого. Важным является приобретение нового качества.

Потенциальное определение 3. Еще одним определением может быть непредсказуемость процесса достижения результата (результат как раз должен быть предсказуем). Например, при уборке помещения робот-пылесос заранее не сообщает свою траекторию движения, определяя ее в процессе. В этом смысле процесс непредсказуем. Однако результат, который отображается на экране калькулятора непредсказуем в том плане, что человек потому и использует калькулятор, что не знает наперед заданного ответа, как, впрочем, и процесса его достижения.

Здесь важен и ответ на вопрос: на ком лежит ответственность за результат? Ответственность применительно к устройствам и людям. В примере со стиральной машиной ответственность за результат лежит на человеке, который разрабатывал программу стирки. Стиральная машина «ответственна» только за выполнение шагов этой программы. В случае же с роботом-пылесосом можно сказать, что в какой-то степени он ответственен и за результат. Например, если на пути его движения появится препятствие, то он его объедет и продолжит работу. Можно ли провести грань между «умным» и «не умным» устройством через нацеленность на результат/процесс? Если разработчик не знает, каким образом устройство достигнет результата, то его можно назвать «умным».

Потенциальное определение 4. Умное устройство должно упрощать предметное действие или делать его тривиальным (как, например, нажатие на кнопку). Варианты: упрощение, отмена, изменение. Скорее всего, «умное» устройство, действительно, должно упрощать предметное действие, одна-

ко, во-первых, предметное действие упрощают даже заведомо «не умные» устройства, а, во-вторых, может быть, упрощение предметного действия «умным» устройством является следствием какого-либо более фундаментального свойства, характеризующего «умное» устройство. Скорее имеет смысл говорить о том, что при использовании «умного» устройства предметное действие становится другим (в частном случае - тривиальным) по сравнению с «не умным» устройством, предназначенным для решения той же задачи. Получается, что для использования «умного» устройства, как и в случае с «не умным», могут потребоваться определенные навыки. Видимо, «умными» можно считать и устройства, для работы с которыми нужно нажать, скажем так, больше одной кнопки.

Например, современный автомобиль, с одной стороны, обладает определенными признаками интеллектуальности, а с другой – предметное действие по отношению к нему нетривиально (если, конечно, не рассматривать беспилотные автомобили). Таким образом, простота или тривиальность предметного действия сама по себе не является необходимым критерием для отнесения устройства к классу «умных». Более того, для многих явно «не умных» орудий наподобие лопаты или молотка предметное действие может быть проще, чем для современного автомобиля, обладающего признаками интеллектуальности.

Свойства, которые могут быть присущи «умному» устройству, и их связь с философско-антропологическими концепциями

Свойство 1. Будем говорить, что устройству присуще свойство *автономности*, если оно способно выполнять поставленную задачу без вмешательства человека. При этом команда на выполнение задачи должна осуществляться через относительно простой человеко-машинный интерфейс, сложность взаимодействия с которым для человека ничтожно мала по сравнению с процессом выполнения данной задачи человеком самостоятельно. Другими словами, человеку не требуется осваивать дополнительные навыки, чтобы отдать команду данному устройству.

Устройства, обладающие свойством автономности, хорошо укладываются в концепцию жизненного аутсорсинга, в рамках которого человек делегирует выполнение некоторых своих функций внешней среде и, в частности, техническим устройствам. При этом процесс делегирования достаточно прост и фактически заменяет предметное действие, которое ранее требовалось для выполнения этой функции. Более того, процесс делегирования может быть очень похож при использовании самых разных устройств и заключаться в отдаче специальных команд посредством нажатия кнопок. Таким образом, свойство автономности фактически демонстрирует механику жизненного аутсорсинга.

С философско-антропологической точки зрения нам не так важно, способно ли устройство взаимодействовать с окружающей средой в процессе выполнения задачи. Главное – это делегирование конкретной задачи данному устройству. Именно в отказе от участия в выполнении задачи в пользу некоторого устройства проявляется явление жизненного аутсорсинга, и данное свойство присуще целому классу устройств, взаимодействие человека с которыми можно исследовать в рамках этой концепции. В некоторой

степени автономное устройство не только играет роль объекта, которому передаются полномочия по выполнению задачи, но и замещает потенциального ассистента, коим может являться другой человек, например, подчиненный или слуга, которому дается поручение.

Свойство автономности отчасти перекликается с позицией Каппа [Капп и др. 1925], согласно которой «в орудии человек систематически воспроизводит себя самого». Однако под его концепцию органопроекции, утверждающую, что всевозможные орудия – это усиление человека, автономные устройства в общем не укладываются, поскольку как таковое усиление не обязательно происходит. Например, такие устройства как робот-пылесос или автоматическая стиральная машина не усиливают человека, а заменяют его. Очевидно, что мы не можем утверждать, что робот пылесос выполняет свою работу всегда быстрее или качественней человека. Хотя в частных случаях это возможно.

Свойство 2. Будем говорить, что устройство обладает свойством *посредничества*, если при его использовании меняется предметное действие, совершаемое человеком, как правило в сторону его упрощения по сравнению с предметным действием, осуществляемым при выполнении данной задачи человеком самостоятельно или с использованием более примитивных устройств. При этом использование устройства-посредника не отменяет необходимости участия человека на протяжении всего процесса выполнения задачи.

Свойство 3. Будем говорить, что устройство обладает свойством адаптивности, если оно способно изменять свое поведение в зависимости от состояния окружающей среды.

При этом человеко-машинный интерфейс может быть определен не очень строго, быть довольно сложным или вообще в явном виде отсутствовать. Принципиальное отличие автономных устройств от интерактивных умных устройств состоит в том, что интерфейс взаимодействия умных устройств довольно четко определен, в то время как с интерактивными устройствами дело обстоит сложнее. Данный интерфейс крайне сложно формализовать и именно при взаимодействии с такими устройствами можно говорить о сращивании человека с технологиями и киборгизации.

Выводы, перспективы и открытые вопросы

Сформулируем основные выводы, которые можно сделать по итогам обсуждения, приведенного в настоящей статье, и наметим возможные применения полученных результатов.

Выделение какого-то особого класса «умных» устройств среди всего их множества не имеет особого смысла как с точки зрения философии, так и фактически, поскольку любая деятельность так или иначе включает в себя интеллектуальный труд. Например, даже при выкапывании траншеи лопатой требуется некоторый мыслительный процесс по выбору силы нажатия на лопату или точного направления движения.

Три сформулированных определения нельзя отделять друг от друга. Многие устройства могут одновременно соответствовать двум или всем трем определениям. Однако, как было заявлено при обсуждении постановки задачи, ключевым критерием при выработке определений должна яв-

ляться их встраиваемость в существующие философско-антропологические концепции с целью получения возможности для анализа соответствующих этим определениям устройств. Так, понятие автономного устройства дает нам механику жизненного аутсорсинга, понятие устройства-посредника дает нам возможность исследований через предметное действие в рамках теории Л. С. Выготского, а класс интерактивных устройств намечает стыковку с трансгуманизмом.

В качестве дальнейших вопросов, представляющих философско-антропологический интерес и связанных с выработанными определениями, могут быть следующие:

До каких пор человек сможет передавать автономным умным устройствам свой функционал, оставаясь при этом человеком?

В чем отличие руководителя, дающего поручения своим подчиненным, от человека, пользующегося автономным устройством?

Какие изменения происходят в человеке, когда он пользуется устройством-посредником и формирует новое предметное действие, фактически забывая старое?

Способно ли интерактивное устройство адаптировать свое поведение, чтобы сформировать у человека зависимость от себя?

Какую стратегию поведения целесообразно выбрать человеку, чтобы пользоваться устройствами этих трех типов себе во благо и с целью своего развития?

Выработанные определения могут быть использованы с целью выбора объекта исследования, поскольку позволяют описать конкретные классы реально эксплуатируемых технических устройств.

«Умное» устройство – это собирательное понятие, куда каждый вкладывает что-то свое. Помимо трех предложенных свойств могут быть введены и другие свойства, имеющие значение для философско-антропологических исследований.