УДК 1/14. 004

ФИЛОСОФИЯ СОЗНАНИЯ: ПРЕОДОЛЕНИЕ ПРЕДЕЛА НА ПУТИ РЕАЛИЗАЦИИ ПРОГРАММЫ СИЛЬНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Стойков В.П., Михайлова Т.Л.

ФГБОУ ВО «Нижегородский государственный технический университет им. Р.Е. Алексеева», Нижний Новгород, Россия (603950, Нижний Новгород, ГСП-41, ул. Минина, 24), e-mail: tmichailova2012@yandex.ru

В статье предлагается обсуждение возможностей конструирования сильного искусственного интеллекта. Для этого приводятся аргументы Д.Серля и Д. Чалмерса, позволяющие провести границу как основание дифференциации искусственного интеллекта и человеческого сознания. Выделены критерии как маркер этой границы. Артикулируется проблема предела. Постулируется, вслед за Д. Чалмерсом, положение о том, что проблема предела – это проблема наведения моста между областью абстрактного и конкретного. Суть проблемы состоит в невозможности проверки наличия ментального сознания в самом искусственном интеллекте. Сможет ли искусственный интеллект обладать человеческими чувствами, доступна ли ему интуиция – все эти открытые вопросы есть приглашение к дискуссии по междисциплинарным проблемам. *Итог* статьи – обозначение теоретической платформы для артикуляции проблемы реализации сильной версии искусственного интеллекта.

Ключевые слова: искусственный интеллект, сознание, программный код, когнитивная система, конструирование, ментальное восприятие, каузальное взаимодействие, казуальная динамика, граница, имплементация.

PHILOSOPHY OF MIND: OVERCOMING THE LIMIT ON THE WAY TO IMPLEMENT THE PROGRAM OF STRONG ARTIFICIAL INTELLIGENCE

Stoikov V.P., Mikhailova T. L.

Nizhny Novgorod State Technical University n.a. R.E. Alekseev, Nizhny Novgorod, Russia (603950, Nizhny Novgorod, GSP-41, street Minina, 24), e-mail: tmichailova2012@yandex.ru

The paper discusses the possibilities of constructing a strong artificial intelligence. The authors look into the works of John Searle and David Chalmers whose arguments enable them to draw a border as the basis of differentiation between the artificial intelligence and human consciousness. The criteria for marking this border are examined. The problem of limit is analysed. Based on Chalmers' philosophy, the authors suggest that a problem of limit is a problem of establishing a bridge between the area of the abstract and the area of the concrete. The key part of the problem lies in the fact that it is impossible to verify the presence of mental consciousness in the artificial intelligence. Can artificial intelligence possess human feelings or intuition? This remains an open questions and an invitation to the discussion on a number of interdisciplinary problems. The paper results in identifying a theoretical platform for discussing a problem of realization of strong artificial intelligence.

Key words: artificial intelligence, consciousness, programming code, cognitive system, design, mental perception, causal interaction, casual dynamics, the boundary implementation.

С середины XX века особую популярность начали набирать идеи создания сильного искусственного интеллекта (ИИ) [1]. Человечество преодолевает очередную преграду – конструирование сложного механизма (разума) при помощи *программного кода* [6]. Сознание человека прошло множество преград, постоянно сталкиваясь с все новыми препятствиями. Способны ли люди оживить металлическую оболочку или это очередная

временная игрушка человека XXI века – вопрос, тождественный постановке философской проблемы, связующей множество специальных прикладных вопросов.

Прежде чем заниматься созданием разума для ИИ, попытаемся сформулировать, *что же представляет сознание человека.* Испокон веков, каждый день оно, сознание, формировалось – от прыгающих по деревьям обезьян до вступивших в космос покорителей. Мысли людей разнообразились – от поиска еды и ежедневного выживания в условиях жестокой окружающей среды до духовных потребностей и прогресса. Сознание прошло множество преград на пути воспроизводства собственного самосознания. Начинаются преграды прямо с рождения человека. Если ребенок растет изолированным от общества и достижений человечества, не имея возможности контактировать с себе подобными, человеком в привычном понимании этого слова, ему не стать. Известны случаи детей, так называемых феральных людей, выращенных животными (среди волков, обезьян), дети перенимают их образ жизни, становясь одними из них. Если их попытаться погрузить в обычную для человека среду обитания, то нередко это заканчивается летальным исходом. Что же тогда представляет сознание человека, если тысячелетия эволюции и наследования ДНК, не могут защитить его с самого начала антросоциогенеза, от возможности стать подобным тем, в чьей среде воспитания он формируется. Каким же тогда может быть исход, если ребенок будет расти в абсолютной изоляции от общения с разумными существами? Сможет ли искусственный интеллект в подобных случаях вести себя как человеческий организм, и кем он будет себя считать, если в аналогичных условиях лишить его какой-либо базы данных, наградив лишь способностью к анализу ситуации, в которую он попал. Будет ли робот, находящийся в социальной среде людей, имеющий гуманоидные черты, отождествлять себя с человеком.

Следующим немаловажный барьер человеческого сознания – *язык*, на котором мы с вами общаемся. Все, что есть вокруг нас, все, что было изобретено человеческим разумом, имеет свое название, и эти названия нередко устанавливают ограничения в полете мысли. Назвать вещь – это значит выделить ее из окружающей среды [3]. Поэтому есть лишь строго определенный набор слов, из которого можно произвести идею. Бывают случаи, когда люди, знающие несколько языков, не могут перевести предложение с одного из языка на другой, просто из-за того, что в одном из них нет слов, способных точно передать смысл сказанного. Сможет ли искусственный интеллект, способный оперировать всеми языками, имеющимися в этот момент на планете, размышлять лучше людей. Обладая более высоким порогом мысли, нежели человек, сможет ли искусственный интеллект изобрести язык, способный элиминировать эти рамки вообще или он произведет иной способ общения, изначально не имеющий барьеров и препятствий.

Немаловажной частью в развитии человеческой истории была религия, ставшая исходной клеточкой культуры. Вследствие отсутствия возможности обоснования природных явлений с научной стороны, людьми было придумано множество богов, религий и подобных ритуальных вещей, а также незнание и страх того, что же будет с человеком после конца его жизни, – способствовали «изобретению» таких понятий, как душа и загробная жизнь. Именно благодаря этим феноменам, человек пролонгировал себя в мир трансцендентного, подавляя вышеуказанные страхи и преодолевая неполноту своего бытия. Как искусственный интеллект будет воспринимать эти духовные явления, ведь для него момент смерти и нажатие человеком на кнопку выключения, практически тождественны. Если принимать во внимание сознание искусственного интеллекта как систему мониторинга состояния машины, систему, создающую различные задачи для анализа ситуации и решения проблем, то может ли тогда сознание человека быть системой, подобно машине, просто принимающей начальные условия своего нахождения, остальное – лишь закономерная реакция частного организма, на которую воздействуют эти аргументы.

Наш организм, кроме *способности мыслить*, может также *чувствовать*, при этом каждая из способностей независима, будучи тесно связанной, друг с другом, однако чувства играют немаловажное влияние на субъективное восприятие, окружающего мира. Можно ли наделить искусственный интеллект способностями, подобными человеческим способностям? Быть может, найдут что-то, заменяющее их. Являются ли пять чувств человека нашими очередными рамками и ограничениями или же наоборот способствуют к прогрессу. Если искусственному интеллекту дать способность обладать чувствами, то могут ли у него появиться новые виды чувств, недоступных человеческому организму.

Есть два типичных возражения против искусственного интеллекта: а) невозможность воспроизведения поведения вычислительных машин по образу когнитивных систем; б) невозможность воспроизведения компьютерами интуиции человека. *Внешние возражения* наталкиваются на трудности, связанные с успехами вычислительных симуляций физических процессов в целом [5, с. 390]. За счет правильно выстроенных алгоритмов компьютерной программы, можно создать машину, способную симулировать подобие человеческого поведения. Собственно, такая машина уже есть на сегодняшний день это – робот София.

Большее распространение получили *возражения*, называемые *внутренними* [5, с. 390]. Согласно этим возражения машина смогла бы сделать лишь вид того, что обладает сознательностью, не имея при этом никакого *ментального восприятия*. Так, робот София, симулирующая модель живого человека, сконструирована так, что может саморазвиваться

за счет получения новой информации из вне. Однако дискуссии с ней проводятся по заранее определенным темам, и ответы в этих дискуссиях вероятнее всего она дает, анализируя и выделяя из всех возможных вариантов те, которые уже встречались в каких-либо книгах, статьях, сочинениях, других подобных контентах. Это доказывает факт того, что некоторые её ответы являются вообще бессмысленными.

Однако если построить сознательную машину, используя вычислительные системы, нельзя, то, каким образом построен человеческий мозг, раз он может создавать такую вещь, как сознание. Ведь в мозгу, по сути, находятся нейроны, выполняющие определенные физические операции, наделенные некой логикой; следовательно, можно сделать чипы, заменяющие данный нейрон и выполняющие все его функции. «*Стандартная теория вычисления* имеет дело исключительно с *абстрактными объектами*: машинами Тьюринга, программами на Паскале, конечными автоматами и т.п. Это – математические сущности. *Когнитивные* же *системы*, существующие в реальном мире, – *это конкретные объекты*, имеющие физическое воплощение и *каузальное взаимодействие* с другими объектами физического мира. Но зачастую мы хотим использовать теорию вычисления для получения выводов о конкретных объектах в реальном мире. Для этого нам нужен *мост между областью абстрактного и областью конкретного*» [5, с. 392].

Понятие имплементации Дэвида Чалмерса – это такого рода мост между «вычислителями» и физическими системами. В своей книге Чалмерс реализует *понятие имплементации* через *систему конечных автоматов* (КА). КА имеют набор данных на *входе*, конечный набор внутренних состояний, конечный набор данных на *выходе,* а также наборы отношений перехода от одного состояния к другому. Такую систему, можно было бы применить к нейронам в голове, построив за счет этого *математическую модель* человеческого мозга. Так, можно наблюдать, что при применении вычислительных описаний к физическим системам, они, по сути, предоставляют нам формальное описание *каузальной организации системы.*

Заманчиво представлять компьютер просто как устройство с входом и выходом, а черный ящик – это место лишь для формальных математических манипуляций. Подобный взгляд на вещи, однако, игнорирует тот ключевой факт, что внутри компьютера – так же как и внутри мозга – имеется богатая *казуальная динамика*. В самом деле, в обычном компьютере, понейронно имплементирующем симуляцию моего мозга, есть место для реальной каузальности между напряжениями различных цепей, в точности, отражающей *паттерны каузальности между нейронами.* Каждому нейрону будет соответствовать участок памяти, репрезентирующий этот нейрон, и каждый из этих частей будет

физически реализован напряжением в каком-то физическом виде. За возникающий сознательный опыт отвечают именно эти каузальные паттерны электрических цепей, так же как за их возникновение отвечают *каузальные паттерны нейронов мозга* [5, с. 400]. Можно ли сегодня сконструировать *машину, обладающую ментальным сознанием*, с привлечением вычислительной системы, аналогичной *казуальному восприятию,* ибо «порождение разума обязано различать *причинно-следственные связи*, подобно тому, как это делает человеческий мозг» [2, с. 594]. Однозначно ответить на этот вопрос нельзя. Чтобы попытаться понять, сможет ли вычислительная машина иметь способность к сознанию, предположим, что мозг человека, будет постепенно заменяться чипами вычислительной системой.

По Дж. Серлу, есть ТРИ варианта исхода такого конструирования.

Исход первый – постепенная замена мозга кремниевыми чипами приводит к ощущению потери контроля над своим телом. Ученые, которые имплантировали бы в его мозг машину, могли даже не заметить потерю контроля человека над телом, ведь перед ними будет машина, способная на реакции человека, однако *из-за отсутствия способности к казуальному мышлению*, она так и осталось бы обычной машиной.

Второй исход – замена нейронов в мозге человека – это дублирование физических и ментальных свойств, тогда с высокой вероятностью можно было бы утверждать, что сильный ИИ, имеет место к жизни.

Третий вариант исхода – имплантация чипов, сохраняющая человеческое сознание при полной парализации организма [4].

Выдвигая знаменитый аргумент против сильного искусственного интеллекта, Дж. Серл в 1980 г. доказывал, что любая программа может быть имплементирована без порождения ментального. Доказательство построено на демонстрации того, что он считает контрприёмом для тезиса о сильном искусственном интеллекте. Китайская комната, внутри которой субъект, манипулируя символами, симулирует человека, понимающего китайский язык. Китайская комната задумана как иллюстрация системы, имплементирующей программу – какой бы они ни была – при отсутствии релевантности сознательного опыта [5, с. 400].

Однако, даже имея абстрактную систему, создающую подобие когнитивной, а также металлический мозг, проблема наведения моста между областью абстрактного и областью конкретного порождает *проблему предела*, суть которого в невозможности проверить наличие *ментального сознания* в самом искусственном интеллекте.

Машины окружают нас везде и уже давно стали верными работниками людей. Создание осознанного искусственного интеллекта, вероятнее всего, тоже будет направлено на

улучшение наших жизней. Однако захочет ли сам ИИ, работать на людей, ведь, по сути, в момент, когда появится первый на нашей планете самостоятельный, независимый от человека ИИ, это будет момент свержения человека с верхушки иерархии пищевой цепи. Ученые предпримут способы к ограничению мысли ИИ в сторону самой мысли о том, что он может навредить людям, однако это самое ограничение, приведет к тому что, ИИ не сможет мыслить критично, или, по крайней мере, мыслить критично во всем. Так как эти ограничения будут придуманы людьми, то вероятнее всего найдутся люди, которые эти самые ограничения, сломают: либо во благо науке и развития, либо ли во вред людям. Выбор остается за человеком. Пока бесспорно одно. Территория дискуссии о сильном искусственном интеллекте, возможностях его конструирования, функционирования и прогнозирования – это территория, где живут *междисциплинарные проблемы*, на которой *философия сознания* выполняет *критическую функцию*, будучи своеобразной измерительной линейкой. Возможно, в недалеком будущем эта линейка позволит преодолеть предел между сильной версией искусственного интеллекта и человеческим сознанием. И этим преодолением обозначит новый предел, новый горизонт, без которого теряет смысл наука как самое интригующее приключение человеческого разума.

Список литературы

[1]. Галстян, Р.В. От технологической сингулярности – к будущим сценариям развития искусственного интеллекта / Р.В. Галстян, Т.Л. Михайлова // Международный студенческий научный вестник. 2017. № 4-3. С. 370-374.

[2]. Масленников, А.Г. Может ли искусственный интеллект эволюционировать, или об аутопоэзисе как границе живого и неживого / А.Г. Масленников, М.А. Лемешевский, Т.Л. Михайлова // Международный студенческий научный вестник. 2017. № 4-4. С. 593-595.

[3]. Михайлова, Т.Л. Вещь как текст: безмолвие вещи VS забвение мира // Антропологическая аналитика: сборник научных трудов. Нижегород. гос. техн. ун-т им. Р.Е. Алексеева. – Нижний Новгород, 2015. С. 86-94.

[4]. Серл, Дж. Открывая сознание заново; пер. с англ. А. Ф.Грязнова. – М.: Идея-Прогресс, 2002. – 256 с.

[5].Чалмерс, Д. Сознающий ум. В поисках фундаментальной теории; пер. с англ. – М: УРСС: Книжный дом «ЛИБРОКОМ», 2013. – С. 512.

[6]. Чернобаев, И.Д. «Чистый код как искусство», или о глубинных тайнах коммуникации / И.Д. Чернобаев, Т.Л. Михайлова // Международный студенческий научный вестник. 2016. № 3-4. С. 597-600.