

УДК 577.3

## К ВОПРОСУ О РАЗЛИЧИИ МЕЖДУ ИСКУССТВЕННЫМ И ЕСТЕСТВЕННЫМ ИНТЕЛЛЕКТОМ

© 2020 г. В.А. Намиот

*Научно-исследовательский институт ядерной физики имени Д.В. Скобельцына Московского государственного университета имени М.В. Ломоносова, 119991, Москва, Ленинские горы, 1*

*E-mail: vnamiot@gmail.com*

Поступила в редакцию 11.10.2019 г.

После доработки 11.10.2019 г.

Принята к публикации 18.10.2019 г.

Обсуждается вопрос о том, может ли, хотя бы в принципе, кибернетическая система, функционирующая на основе алгоритмов, моделировать поведение естественного интеллекта. Ответить на этот вопрос математически строго пока что невозможно. Тем не менее в работе приводятся достаточно убедительные соображения, показывающие, что любые существующие в настоящее время кибернетические системы в принципе не могут моделировать поведение естественного интеллекта, причем изменить эту ситуацию с помощью каких-либо программных ухищрений также невозможно. Для этого нужны принципиально новые подходы, некоторые из которых также обсуждаются в настоящей статье.

*Ключевые слова: естественный и искусственный интеллект, универсальная машина Тьюринга, оракул.*

DOI: 10.31857/S0006302920010238

Прежде всего, следует отметить, что самое очевидное различие между искусственным и естественным интеллектом, которое сразу же бросается в глаза (искусственный интеллект реализуется в настоящее время в электронных вычислительных устройствах, например, компьютерах или нейросетях, в то время как естественный интеллект является атрибутом биологических систем), само по себе не является принципиальным. Устройства, о которых можно сказать, что они вроде бы демонстрируют наличие интеллекта, могут быть изготовлены из самых разных материалов, включая биологические. Это не имеет особого значения в том случае, если в основе их функционирования лежат одни и те же принципы работы [1]. Поэтому чтобы ответить на вопрос, поставленный в заглавии, нужно в первую очередь разобраться с тем, может ли компьютерное устройство, хотя бы в принципе, моделировать происходящее в биосистемах и соответственно «вести себя» как они. Если компьютерное устройство в состоянии это делать, то все имеющиеся различия между искусственным и естественным интеллектом заведомо не являются принципиальными.

Но прежде чем отвечать на этот вопрос, предварительно нужно сформулировать критерий, выполнение которого позволяет признать подоб-

ное моделирование успешным. Например, в качестве подобного критерия предполагалось проведение некоего испытания (своего рода экзамена), прохождение которого позволяет признать, что искусственный интеллект не уступает естественному. В качестве такого испытания предлагалась, в частности, следующая процедура. Берутся две полностью закрытые кабины, в каждой из которых имеется телефон, в одну из них помещается человек, а в другую — искусственный интеллект. Эксперт (человек) может по телефону связываться с любой из этих кабин и разговаривать с тем (или с чем), находящимся там. И если в результате этих разговоров эксперт окажется не в состоянии разобраться с тем, кто в какой кабине находится, то следует признать, что принципиальной разницы между искусственным и естественным интеллектом нет.

Однако к настоящему времени выяснилось, что все существенно сложнее — прохождение подобного критерия еще ни в коей мере не гарантирует, что искусственный интеллект окажется в силах решать те задачи, которые доступны естественному интеллекту. Можно написать относительно простую, в каком-то смысле даже примитивную программу, заведомо не способную разобраться в сколько-нибудь сложных ин-

теллектуальных вопросах, которая, тем не менее, сможет пройти вышеуказанный критерий. Для этого достаточно составить программу таким образом, чтобы она могла поддерживать разговор до тех пор, пока он касается нескольких предварительно определенных тем, а если эксперт захочет выйти за их рамки, то применить какой-либо из также заранее запрограммированных приемов, позволяющих или «перевести» разговор на что-то другое, или же вернуться к предыдущему. При этом эксперт может подумать, что разговаривающий с ним груб, капризен, отличается плохим характером, возможно, даже просто глуп, но он не усомнится, что это все-таки человек.

Вернемся вновь к вопросу о критерии. Приведенное выше рассуждение показывает, что если критерий основан на каком-то экзамене, оценке эксперта и тому подобным процедурам, то его ни в коей мере нельзя признать сколько-нибудь надежным. Критерий должен быть объективным. Так, если бы удалось доказать, что существуют задачи, которые естественный интеллект решает, в то время как искусственный интеллект принципиально решить не в состоянии, то это бы означало, что два данных вида интеллекта заведомо функционируют по-разному. Следовательно, они не могут быть эквивалентны один другому.

Начнем с искусственного интеллекта. Обсудим самые основы существующих систем такого интеллекта.

Любая существующая на настоящий момент кибернетическая система, в том числе и такая, о которой говорят, что она способна «обучаться», работает на основе заложенного в нее алгоритма. Существует по крайней мере несколько математически строгих определений понятия «алгоритм», причем в математической логике доказываем, что все эти определения эквивалентны друг другу. На практике, например, при доказательстве теорем, очень удобным является определение алгоритма, данное А. Тьюрингом. Это определение использует понятие так называемой «универсальной машины Тьюринга» [2], которая при соответствующей введенной в нее программе оказывается в состоянии моделировать работу любого другого существующего в настоящее время логического устройства.

Машина Тьюринга, о которой здесь идет речь, устроена в достаточной мере просто. В ней имеется бесконечно длинная в обе стороны лента, разделенная на ячейки, и так называемая головка, представляющая собой конечный автомат, способный перемещаться в обоих направлениях вдоль ленты. В ячейках ленты могут записываться определенные символы (причем число этих сим-

волов конечно), головка же может эти символы считывать, после чего менять свое состояние, стирать из ячейки прежний символ и записывать в ней новый, а затем перемещаться вдоль ленты на какое-то определенное расстояние и повторять всю вышеуказанную процедуру. Число таких повторений в принципе не ограничено. На ленте выделены специальные ячейки, в которых записывается программа, входная информация, а также результат вычислений. Как было показано А. Тьюрингом, среди всевозможных подобных машин с различными головками и т.п. имеются, как уже говорилось выше, универсальные машины, способные моделировать работу любых других таких машин.

Потребность в математически строгом определении алгоритма чаще всего возникает даже не тогда, когда нужно сконструировать какой-то алгоритм, а тогда, когда нужно доказать, что в принципе не может существовать алгоритма, способного решить поставленную задачу. Действительно, если алгоритм существует, то для доказательства его существования достаточно просто его построить и предъявить. При этом обычно, вне зависимости от наличия или отсутствия математически строгого определения алгоритма, никаких трудностей с тем, чтобы понять, является ли полученный результат алгоритмом или же нет, просто не возникает. Но если алгоритма в принципе не может существовать, то подобным образом мы ничего доказать не сможем. Ведь то, что мы не смогли построить искомым алгоритм, может означать не то, что его в принципе не существует, а только то, что мы в силу каких-то причин оказались не в состоянии его найти. Но вот если удастся доказать, что в принципе не может существовать такой программы, которая позволила бы универсальной машине Тьюринга решить поставленную задачу, то тем самым мы строго доказываем, что и соответствующего алгоритма также не существует. Наиболее известным примером подобной теоремы о «не существовании алгоритма» является теорема об «остановке» [2], т. е. о том, что в принципе не может существовать алгоритма, способного определить, остановится ли хоть когда-нибудь «универсальная машина Тьюринга», реализующая какую-либо программу, или же будет вычислять результат «вечно».

Если бы удалось установить, что естественный интеллект в состоянии найти решение такой задачи, для которой можно доказать, что алгоритма, способного ее решить, не может существовать в принципе, то это означало бы, что между естественным интеллектом и моделирующим его поведение искусственным интеллектом существует принципиальное различие. И это различие, как

ни стараться, невозможно убрать никакими программными ухищрениями...

Но поскольку мы не знаем, как «в действительности» функционирует естественный интеллект, мы оказываемся не в состоянии определить его возможности только на основе теоретических соображений. Экспериментальные же подходы также не могут дать материала для сколько-нибудь надежных утверждений и тоже уязвимы для критики...

Тем не менее, если естественный интеллект возник и сохранился в ходе эволюции, то можно утверждать, что он важен для выживания самых разных видов. В тоже время можно привести ряд аргументов, позволяющих утверждать, что интеллект, основанный только на алгоритмах, не мог бы способствовать выживанию. Действительно, системы, основанные на использовании алгоритмов, с самого начала должны быть достаточно сложными: если же усложнение происходит непосредственно в процессе развития, то на начальных этапах, пока система еще не отлажена, неизбежно будет появляться очень большое количество ошибок. Если биосистема будет управляться подобным «интеллектом», то это не только не будет способствовать ее выживанию, но, наоборот, достаточно быстро приведет к гибели. Основываясь на подобных соображениях, можно утверждать, что задача о выживании и эволюции биосистем, так или иначе решаемая естественным интеллектом, искусственным интеллектом решена быть не может...

Но возникает вопрос, может ли вообще существовать какая-либо информационная система, чьи возможности превосходят возможности универсальной машины Тьюринга? В какой-то (но далеко не в полной мере) этот вопрос ставится и изучается только в математической логике. В частности, там рассматриваются различные способы добавить нечто такое к универсальной машине Тьюринга, что позволит увеличить (если, конечно, это удастся) ее возможности. Например, к машине Тьюринга можно добавить датчик случайных чисел (машина Тьюринга в состоянии вычислять псевдослучайные числа, но, строго говоря, «истинно случайных» чисел она генерировать не может). Но даже с такой добавкой решать алгоритмически неразрешимые задачи машина Тьюринга по-прежнему не способна.

Однако если рассуждать чисто абстрактно, к машине Тьюринга в принципе можно подключить и более интересные устройства, чем датчик случайных чисел. В математической логике, например, рассматривается устройство, называемое «оракул» [3]. Оно представляет собой систему, которой можно задать вопрос, и она на этот вопрос ответит. При этом вероятность того, что от-

вет окажется ошибочным, или просто равна нулю, или же, если она все-таки отлична от нуля, крайне мала. Используя подобный «оракул», машина Тьюринга уже оказывается в состоянии решать алгоритмически неразрешимые задачи [3].

Само по себе существование «оракула» (если бы, конечно, он существовал в реальности, а не являлся чисто абстрактным объектом, изучаемым только в рамках математической логики) означает гораздо большее, чем просто возможность решать те или иные алгоритмически неразрешимые математические задачи. Например, в биологии имеется ряд фундаментальных вопросов, в частности, связанных с эволюцией биосферы, на которые там до сих пор нет убедительных ответов. Если же существует возможность обратиться к «оракулу» и «получить у него подсказку, позволяющую найти хорошее (с эволюционной точки зрения) решение», то многие, если не все из этих вопросов, могли бы быть «сняты»...

Однако математическая логика не ставит (да она и не должна ставить) вопроса о том, как может быть устроен подобный «оракул» и можно ли, а если можно, то каким именно способом, его изготовить или же чем-либо заменить. На настоящее время ответ на этот вопрос остается открытым. Соответственно, по-прежнему остается неизвестным и ответ на вопрос о том, в чем же конкретно, т. е. на уровне устройства и принципов работы, состоит различие между искусственным и естественным интеллектом.

## КОНФЛИКТ ИНТЕРЕСОВ

Автор заявляет об отсутствии конфликта интересов.

## СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания каких-либо исследований с использованием людей и животных в качестве объектов.

## СПИСОК ЛИТЕРАТУРЫ

1. Н. Винер, *Творец и робот* (Прогресс, М., 1966).
2. М. Минский, *Вычисления и автоматы* (Мир, М., 1971).
3. С. Ааронсон, *Квантовые вычисления со времен Демокрита* (АНФ, М., 2018).

## **On the Question Regarding the Difference between Artificial and Natural Intelligence**

**V.A. Namiot**

*Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University,  
Leninskie Gory 1, Moscow, 119991 Russia*

There is some debate as to whether a cybernetic system built with the use of algorithms is theoretically able to model the behavior of natural intelligence systems. A valid answer to this question, from a mathematical viewpoint, can not be given. Meanwhile, in this paper we provide rather convincing reasoning that there is still no cybernetic systems potentially which could model the behavior of natural intelligence systems. It is also impossible even with the help of any program tricks meaning that fundamentally new approaches some of which are discussed in this paper are needed.

*Keywords: natural and artificial intelligence, universal Turing machine, Oracle*