

# 10, октябрь 2015

УДК 004.896

### **Философские проблемы искусственного интеллекта**

*Лавриненко А.Ю., студент  
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,  
кафедра «Специальное машиностроение»*

*Научный руководитель: Губанов Н.Н., к.ф.н., доцент  
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,  
кафедра «Философия»  
[dekan.fsgn@bmstu.ru](mailto:dekan.fsgn@bmstu.ru)*

«Умные» технические системы уже давно проникли в нашу жизнь и прочно укоренились в ней. Сложно представить себе решение многих вопросов без консультаций с «поисковиком» вроде Google, а люди, «разговаривающие» со своими гаджетами, давно не вызывают удивленных взглядов на улице. Мы разработали машины, способные понимать конкретные приказы и выполнять заложенные в них действия, в какой-то мере даже отвечать на нашем языке, создав подобие диалога. Однако, это все примеры т.н. «примитивного» интеллекта. Возможность же создания искусственного интеллекта, способного быть с человеком на равных – вот то, о чем не прекращаются жаркие споры. Проблема обостряется ещё и тем, что современными учёными отмечается всё более возрастающая – почти в экспоненциальной зависимости – роль субъективной реальности (интеллекта) в жизни общества [2; 3; 10].

Само понятие «искусственный интеллект» находится ныне на стыке наук, и каждая рассматривает его по-своему. Такая ситуация порождает широкий круг проблем: от дискуссий о возможности функционирования «умных машин» до вопроса необходимости или даже правомерности их создания. Но, прежде чем начать разбираться в этих вопросах, рассмотрим, что же такое интеллект. Существует множество дефиниций этого понятия в каждой науке; из них я хотел бы выделить три основных характеристики, определяющих интеллект: способность к обучению, способность к абстрактному мышлению, способность к адаптации и выбору решения поставленной задачи. Три атрибута, которыми полностью обладает только человек. Их мы должны как бы «передать» машине, чтобы признать (сделать) ее разумной. Важные аспекты первой характеристики,

связанной с влиянием обучения на формирование интеллекта, подробно рассматриваются в работах отечественных исследователей [7; 8; 9].

Считается, что интеллект находится у человека в головном мозге, содержащем около  $2^{40}$  основных узлов нейронов, которых соединяют около  $2^{50}$  связей синапсов [1]. Это число представляется невероятно огромным, но современная техника стремительно приближается по своим вычислительным возможностям к указанным параметрам. В наши дни искусственные нейронные сети контролируют сложнейшие системы управления и слежения, проявляют способности в области распознавания изображения вплоть до возможности создания «интеллектуальных» автопилотов. Те же шахматы, с древности считавшиеся прерогативой узкой группы людей-интеллектуалов, были освоены компьютерами и результаты «игры» машин значительно лучше человеческих (они неоднократно обыгрывали бывшего чемпиона мира Г.Каспарова и др.). В таких условиях приобретает особую значимость рассмотрение основных философских вопросов, связанных с искусственным интеллектом и порождаемой им особой «искусственной жизнью». При этом, очевидно глубокое взаимопроникновение философских проблем искусственного интеллекта и искусственной жизни и философских проблем мышления и жизни вообще. Многие современные западные и отечественные философы глубоко исследуют данные проблемы. При этом на первое место выходит обсуждение мышления и самосознания.

В качестве аргумента, говорящего о невозможности создания искусственного интеллекта, часто приводят такой факт: если сама природа мышления человека до сих пор остается загадкой, то как можно создать то, чего мы не понимаем? В общем смысле, мышление – это способность человека рассуждать, размышлять, отражая объективную действительность в неких представлениях, понятиях, суждениях и организованных на их базе теориях и учениях. Можно ли воспроизвести этот процесс с помощью технических средств? Есть множество теорий, как подтверждающих это, так и опровергающих.

Над вопросом «Может ли Машина размышлять?» впервые задумался еще Алан Тьюринг в далеком 1950 г. В разных публикациях ученый выдвигал гипотезы как подтверждающие эту возможность, так и оппозиционные взгляды. По его мнению, самым важным для искусственного интеллекта является способность к коммуникации с человеком [26]. На этой концепции и был основан тест Тьюринга, смысл которого заключается в следующем. В разных комнатах помещаются человек и машины. Третий человек, который будет являться «судьей» эксперимента, общается с ними посредством

текста. Если арбитр не может однозначно определить, с кем он общается – тест для машины считается пройденным.

Конечно, тест Тьюринга затрагивает лишь часть проблем, связанных с искусственным интеллектом, например, он не учитывает работу с визуальной информацией, где компьютеры наиболее слабы. А ведь 70% информации для размышлений человек получает именно зрительно, значит и машина, ему подобная, не должна уступать индивиду и в этом параметре. Однако сам тест признан эталоном, и еще не был пройден ни одной машиной.

Описанное выше – краткий обзор технологических проблем искусственного интеллекта, а ведь есть еще множество социально-этических, над решением которых исследователи бьются ещё с момента появления самой идеи создания искусственного интеллекта.

Допустим, люди создадут в будущем некую систему с искусственным интеллектом. Очевидно, что сразу же над этой системой начнёт проводиться множество экспериментов, целью которых будет изучение принципов и законов мышления человека, изучение творческой компоненты интеллекта и прочие вопросы, ответы на которые можно будет найти опытным путем. Но ведь мы создаем систему, равную себе, обладающую мыслями, чувствами, ощущениями. Спустя время она осознает свое положение и тут возникает существенный морально этический вопрос – имеем ли мы право подвергать экспериментальным «пыткам» порожденным нами интеллект? Именно пыткам, ведь мы не являемся абсолютно компетентными в области мышления, и наше изучение в конце концов сведется к своеобразному «тыканью палкой» и наблюдению за результатом. А результатом таких опытов («тычков») могут быть страдания «подопытного». Чтобы вернуть машину в изначальное (прединтеллектуальное) состояние потребуется перегрузка, что, скорее всего, будет равносильно смерти для искусственного интеллекта.

Допустимо ли убивать ИИ, «издеваться» над ним, когда он обладает таким же набором чувств, что и прототип, т.е. мы с вами? Стоит ли того те тайны, что могут быть открыты в результате этих исследований? Если воспринимать искусственный интеллект равным себе – ответ «недопустим» напрашивается сразу, однако, я уверен, что в мире всегда найдутся заказчики, финансирующие подобные исследования, и ученые, готовые их проводить.

Еще одна проблема, которая крайне волнует учёных при дальнейшем исследовании ИИ, – это проблема безопасности. На полках книжных магазинов и в интернете можно

найти десятки произведений научной (и не очень) фантастики про вышедший из под контроля машинный разум. Однако подобная перспектива вполне возможна на практике, если заранее не принять соответствующие меры. Проблемам обеспечения солидарности между взаимодействующими субъектами, обладающими своими особыми ментальными характеристиками, а также методикам понимания «Другого» посвящен ряд работ [4; 5]. Одна из самых эффективных и общепризнанных мер была предложена писателем-фантастом и ученым Айзеком Азимовым – три закона робототехники, а именно [27]:

1) Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку причинили вред;

2) Робот должен повиноваться командам, которые дает ему человек, кроме тех случаев, когда эти команды противоречат первому закону;

3) Робот должен заботиться о своей безопасности, насколько это не противоречит остальным законам.

Это – очень разумные законы. Основная проблема заключена в том, что данные нормы сформулированы на «человеческом» языке, их достаточно сложно перевести в форму машинных алгоритмов. Попробуйте (хотя бы гипотетически) объяснить машине, что такое «причинить вред», а что нет, или попробуйте объяснить, в каких ситуациях машина должна вмешиваться, а когда это не имеет смысла.

Даже если предположить, что мы смогли все «растолковать» машине, остается вопрос интерпретации. К каким логическим выкладкам придет машина, размышляя над уже обсуждаемым словом «вред». Не решит ли она, что само существование людей есть вред, ведь мы уничтожаем планету, убиваем друг друга, убиваем себя алкоголем, наркотиками, сигаретами? Не решит ли она, что самый простой выход – уничтожить нас всех? Конечно, можно возразить, что машине можно разъяснить многие частные случаи, но тогда это уже будут не исходные 3 закона, а гораздо, гораздо больше.

Другой вопрос, который напрашивается – как поведет себя машина, если перед ней будет выбор, кого спасти из двух людей. Особенно интересна ее реакция, в случае, когда машина не обладает значительным количеством информации об обоих людях.

Получается, нет надёжной системы безопасности, позволяющей сосуществовать человеку и машине? Если рассматривать концепцию ИИ, как прототипа человека, можно рассмотреть следующий вариант.

Да, мы не до сих пор не знаем, за что отвечает каждый нейрон в мозгу у человека. Однако группы нейронов, отвечающие за определённые эмоции, нам прекрасно известны. Были проведены эксперименты, когда с помощью раздражения определённых зон

нейронов получали желанные эмоции у человека [28]. Соответственно, можно повторить подобное и на машинах, взяв давно используемый в технике принцип «хорошо – плохо». В аналитической интерпретации это сведется к минимумам и максимумам определённых функций. Если к тому же предусмотреть возможные отклонения психики у людей, с которыми машина будет взаимодействовать, вроде психоза или депрессии, получается достаточно надёжная система. Ещё одно направление, в котором видится возможность прояснения данной проблемы – это моделирование объективно-нереальных, но субъективно-переживаемых ситуаций [6; 11; 21; 23; 24].

Следующая проблема получила название «проблема Творца». Действительно, создавая машину, способную мыслить, человек уподобляет себя Богу, заимствуя у него функцию: создавать самого себя (точнее, копию по «образу и подобию»). Отсюда, кстати, вытекает ещё одна значительная проблема – теологическая. «Мышление является неотъемлемой частью бессмертной души человека. А души Бог подарил только людям. Поэтому машины не могут мыслить», – так звучит один известный контраргумент против создания искусственного интеллекта. Однако уже известный нам Алан Тьюринг утверждал, что этот аргумент ошибочный, поскольку, если Бог существует, то для него не должно быть проблемой заключить душу в другую ёмкость, нежели человек. По мнению ученого, создание достаточно сложной машины, способной мыслить, эквивалентно рождению ребенка, то есть созданию индивидуума, в которого Всевышний должен имплантировать душу [26].

Развитие ИИ породит и ряд других проблем. Во-первых, потребует иного, чем современное, наукоемкого, лишённого ручного, механического труда, производства [16; 18]. Во-вторых, приведет к переоценке содержания понятий «прогресс» и «регресс» относительно общества [19; 20]. В-третьих, вызовет необходимость существенного повышения уровня образования управленческой элиты, ее выхода за пределы «узкоспециальных» знаний [25]. В-четвертых, породит в социуме новые формы (виды) ритуальной деятельности [12;13;14;15]. В-пятых, усилит вариативность развития общества, а, возможно, проведет новую грань социальной дифференциации между носителями традиционного и искусственного интеллектов, что уже сегодня проявляется в рекламе [17]. Подобный процесс роста проблемности общественных отношений благодаря ИИ, будет, на мой взгляд, только усиливаться.

В заключение можно сказать, что хотя формирование искусственного интеллекта сейчас пока технически невозможно, но уже подробно проработаны отдельные его аспекты, и, если верить оптимистичным прогнозам, создание первой «умной» машины

уже не за горами. Но, обдумав даже те несколько основных проблем, приведенных выше, возникает вопрос: «А надо ли вообще создавать искусственный интеллект? Может, стоит собрать все исследования по этому вопросу, положить в некий «дальний ящик стола», заперев вместе с этим и все последующие проблемы»? Весьма сомнительно, что человечество отложит свои давние мечты, испугавшись возникающих проблем. Рано или поздно, по воле большинства людей или против него, «умные» машины будут созданы. И лучше, если мы подойдем к этому моменту, тщательно проработав все возможные трудности, с которыми нам придется столкнуться. Такая ситуация делает исследование философских аспектов создания искусственного интеллекта чрезвычайно актуальным в ближайшее время.

### Список литературы

1. Головной мозг человека // Энциклопедия «Кругосвет». Режим доступа: [http://www.krugosvet.ru/enc/nauka\\_i\\_tehnika/biologiya/GOLOVNO\\_MOZG\\_CHELOVEK\\_A.html?page=0,0](http://www.krugosvet.ru/enc/nauka_i_tehnika/biologiya/GOLOVNO_MOZG_CHELOVEK_A.html?page=0,0) (дата обращения: 20.02.2015).
2. Губанов Н.Н. Менталитет и его функционирование в обществе // Философия и общество. 2006. № 4. С. 125–141.
3. Губанов Н.Н. Роль менталитета в развитии общества // Вестник Тюменского государственного университета. 2007. № 1. С. 99–105.
4. Губанов Н.И., Губанов Н.Н. Особенности познавательной деятельности в социально-гуманитарных науках // Философия и общество. 2010. № 2. С. 90–104.
5. Губанов Н.И., Губанов Н.Н. О ментальных основаниях общественной солидарности // Социум и власть. 2012. № 1. С. 98–102.
6. Губанов Н.И., Губанов Н.Н. Перспективы использования объективно-нереальных ситуаций // Вестник Ишимского государственного педагогического института им. П.П. Ершова. 2013. № 3 (9). С. 18–23.
7. Губанов Н.Н. Становление университетской традиции в эпоху Высокого Средневековья // Социология образования. 2014. № 1. С. 56–69.
8. Губанов Н.Н. Вызов Аполлона как стимул развития образования // Alma mater (Вестник высшей школы). 2014. № 5. С. 19–23.
9. Губанов Н.И., Губанов Н.Н. Роль образования в формировании глобалистского менталитета // Alma mater (Вестник высшей школы). 2014. № 11. С. 11–17.

10. Губанов Н.И., Губанов Н.Н. Менталитет в системе движущих сил социального развития // Историческая психология и социология истории. 2014. Т. 7. № 2. С. 149–163.
11. Губанов Н.И., Губанов Н.Н. Субъективная реальность и пространство // Вопросы философии. 2015. № 3. С. 45–54.
12. Лисина Е.А. НОМО RITUALIS. Миф, история, современность. МО-Ногинск: АНАЛИТИКА РОДИС, 2011. 282 с.
13. Лисина Е.А. Роль ритуального в культуре и социуме // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. Тамбов: Грамота, 2011. № 8. С. 142-146.
14. Лисина Е.А. К вопросу о механизмах перехода: социально-философский аспект // Вестник Оренбургского государственного университета. № 7. С. 115-120.
15. Лисина Е.А. Матричная природа ритуального действия (на примере китайской цивилизации) // Вестник ОГУ. Оренбург. 2005. № 2. С. 27-32.
16. Малькова Т.П. Источники развития науки как социального института // Этносоциум и межнациональная культура. 2013. № 1 (55). С. 47-52.
17. Нифаева О.В., Нехамкин А.Н. Социальная реклама как фактор экономического развития // Мировая экономика и международные отношения. 2013. № 5. С. 48-55.
18. Нехамкин А.Н. Научное производство: механизм развития // Экономист. 1996. № 9. С. 52-58.
19. Нехамкин В.А. Теория общественного прогресса: достижения и пределы // Вестник Российской академии наук. 2013. Т. 83. № 8. С. 711-719.
20. Нехамкин В.А., Полякова И.П. Антипрогрессистские теории социально-исторической динамики // Вестник Российской академии наук. 2014. Т. 84. № 7. С. 610 - 617.
21. Нехамкин В.А. Контрфактические исторические исследования и синергетика // Общественные науки и современность. 2006. № 5. С. 120-130.
22. Нехамкин В.А. Контрфактические исследования в историческом познании: генезис, методология. М.: МАКС Пресс, 2006. 176 с.
23. Нехамкин В.А. Контрфактические исторические исследования в системе научного познания // Общественные науки и современность. 2007. № 5. С. 131-140.
24. Нехамкин В.А. От альтернативной истории к контрфактическому моделированию // Человек. 2005. № 6. С. 56-62.
25. Нехамкин В.А. Роль образования руководителя государства в процессе исторического развития // Социология образования. 2012. № 2. С. 82-97.

26. Тьюринг А.Ю. Может ли машина мыслить? М.: Государственное издательство физико-математической литературы, 1960. 68 с.
27. Три закона робототехники // ProRobot Роботы и робототехника Режим доступа: <http://www.prorobot.ru/slovarik/robotics-zakon.php> (дата обращения 21.02.2015).
28. Ученые научились влиять на принятие решений человека с помощью нейробиологических методов // Экзотерические новости. Режим доступа: <http://www.prorobot.ru/slovarik/robotics-zakon.php> (дата обращения 21.02.2015).