«Вычислительные машины и разум»

Алан Мэтисон Тьюринг

Обзорная статья

Вступление

В своей знаменитой статье Алан Тьюринг решил задать прямой вопрос: «Могут ли вычислительные машины мыслить?». Причем вопрос, его постановка, не удовлетворяют автора с самого начала. В самом деле, что есть разум, какова его локализация, и как определить сам процесс мышления? Что считать сознанием и мыслями? Может быть, человек просто не готов «поверить» в мыслительный процесс машина или нам «удобно» считать себя венцом творения, единственным разумным и сознательным существом.

На эти вопросы в своей статье Тьюринг и ищет ответы, предлагая читателю, в некоторой мере, самому для себя решить, возможны ли мыслительные процессы в вычислительной машине, способна ли она к таким же умозаключениям. Наконец, сможет ли машина сыграть в имитацию человека.

Игра в имитацию

Для проверки гипотезы о способности вычислительной машины мыслить Тьюринг предлагает эксперимент, в котором стороннему исследователю (C) предлагается беседовать посредством телетайпа с двумя другими игроками – мужчиной (A) и женщиной (B). Исследователь знает их под обозначениями X и Y, и в конце игры должен однозначно сказать, что X это A, а Y это B, или наоборот. Он вправе задавать вопросы вида: «Попрошу X сказать, короткие или длинные у него и нее волосы». При этом цель игрока A побудить исследователя C к неверным умозаключениям. Целью же игрока B является помощь исследователю, то есть лучше стратегией игрока B – давать правдивые ответы. Но при этом никто не мешает игроку A вставлять свои комментарии, которые будут «путать» исследователя.

Теперь, если в данной игре заменить игрока A машиной, будет ли исследователь ошибаться так же часто, как в случае реальных мужчины и женщины?

Именно таким вопросом Тьюринг заменяет первоначальную проблему.

Критика новой постановки проблемы

В следующем параграфе Тьюринг рассматривает саму корректность постановки сформулированного вопроса. Для начала требуется отбросить сугубо физическую сторону вопроса – абсолютно несущественно, как выглядит машина, является ли она имитацией человека или нет. Важен именно интеллектуальный аспект. Сама «постановка вопроса в новой форме отражает это условие, напрямую запрещая касаться других участников игры, видеть их или вслушиваться в их голоса».

Главным аспектом в критике предлагаемой игры, полагает Тьюринг, может являться лишь априорное превосходство машины над человеком: «Если человек попытается выдать себя за машину, не подлежит сомнению, что эта попытка окажется неудачной. Все мгновенно станет понятно по замедленности расчетов и по ошибкам в арифметических вычислениях. Кроме того, машины, возможно, обладают чем-то, что можно описать как мышление, но это мышление принципиально отличается от человеческого. Данное возражение видится весьма убедительным, но мы можем постулировать, что, если будет когда-либо сконструирована машина для удовлетворительной игры в имитацию, указанное выражение нет необходимости учитывать».

Машины, задействованные в игре

Далее Тьюринг желает показать, какие вычислительные машины должны будут применены в предложенной игре, выделяя следующие принципы:

1) Вполне естественно ожидать, что в наших машинах будут использованы все новейшие достижения инженерной мысли.
2) Мы согласны допустить, что некий инженер или команда инженеров сконструирует машину, которая будет работать, но принципы функционирования которой ее конструкторы не смогут удовлетворительно описать, поскольку при конструировании они применяли методы, по большей части экспериментальные.
3) Мы желаем исключить из числа машин людей, рожденных естественным способом.

Однако использование безусловно всех достижений инженерной мфсли не должно приводить к подмене понятий, то есть, например, не нужен для игры в имитацию биокомпьютер, имитирующий биологические процессы. Нужен именно вычислитель, который был бы способен сымитровать мыслительную деятельность.

«Возникает побуждение отвергнуть условие о применимости в игре всех без исключения инженерных достижений. И тем охотнее мы сделаем это, чем скорее осознаем, что текущий интерес к мыслящим машинам возник, в частности, благодаря особой разновидности машин, обычно именуемых “электронными компьютерами” или “цифровыми компьютерами”. <…> Мы не ставим себе целью выяснить, все ли цифровые компьютеры смогут играть в нашу игру, или понять, справятся ли с игрой в имитацию нынешние цифровые компьютеры. Мы хотим установить, возможно ли корректное участие в игре в имитацию неких воображаемых компьютеров».

Цифровые компьютеры

После введения термина «digital computer» важно уточнить, из каких элементов состоит такой компьютер и какие функции человеческого разума они имитируют; почему в нем должна быть применена именно цифровая логика и цифровой способ обработки информации.

Цифровой компьютер обычно состоит из следующих блоков:

1) *Запоминающее устройство* – это хранилище информации, соответствующее запасам бумаги для человека-вычислителя. <…> Если же человек выполняет некоторые расчеты в уме, часть запоминающего устройства будет соответствовать человеческой памяти.
2) *Исполнительный блок* – устройство, выполняющее разнообразные индивидуальные операции в рамках общего вычисления. Характер этих одиночных операций варьируется от машины к машине.
3) *Контролирующий блок* – его обязанность заключается в том, чтобы эти операции (команды) выполнялись в правильном порядке. Конструируется таким образом, чтобы обеспечить соблюдение этого условия. <…> Помимо того, контролирующий блок должен позволять проверку неких условий для повтора последовательности операций снова и снова.

Тьюринг постулирует процесс выбора человеком того или иного действия как некий свод правил, которым тот руководствуется. Составление такого «руководства к действию» не более чем «полезная фикция», так как человек попросту помнит, что нужно делать. Что касается машины, то для получения результата неких действий или вычисления необходимо составить свод команд, описывающих такое вычисление. «Составление таких сводов обычно называют *программированием*».

Использование именно цифровых компьютеров обусловлено также его сходством с теми нейропроцессами, которые происходят в человеческом разуме. «Современные цифровые компьютеры являются электрическими аппаратами и нервную систему человека также можно назвать электрическим аппаратом. <…> Подобное применение электричества имеет не только теоретическое значение. Разумеется, электричество появляется там, где происходит быстрая передача сигналов, поэтому неудивительно, что мы сталкиваемся с ним в обоих перечисленных случаях. Однако в нервной системе человека химические реакции важны ничуть не меньше, чем электрические. <…>

Обозначенное выше сходство по признаку использования электричества предполагает лишь поверхностную аналогию. Если мы хотим найти

глубинное сходство, нам следует изучить математические модели функционирования цифровых компьютеров и нервной системы человека.

Универсальность цифровых компьютеров

Фактически, показанные Тьюрингом цифровые компьютеры являются «машинами с дискретными состояниями», которые функционируют, перемещаясь внезапными «прыжками» или «рывками» от одного конкретного состояния к другому. Строго говоря, существуют они чисто теоретически, «поскольку в действительности всякое движется непрерывно». Однако стоит вспомнить одного из основоположников квантовой теории физики, Макса Планка, который первым употребил термин *квант* как «некой неделимой порции физической величины», то есть, формально, выдвинул идею дискретности физических величин, противопоставив, таким образом, это понятие непрерывности.

То есть все-таки существуют такие типы машин, которые можно считать машинами с дискретными состояниями. В самом деле они оказываются весьма пригодными для построенной игры, так как «такие состояния достаточно отличаются друг от друга, и потому мы вправе игнорировать случаи ошибочного принятия одного состояния за другое.

На примере осветительной сети Тьюринг показывает, что в полной мере такую систему возможно описать ее начальным состоянием и входным сигналом, для чего можно составить таблицу. При этом несущественными представляются промежуточные положения переключателя – существенными же являются его фиксированные положения, при которых происходит загорание лампочки. Аналогичную таблицу возможно построить и для выходных сигналов. Таким образом, подобная система оказывается типичной для машин с дискретными состояниями.

«На основе заданного начального состояния машины и заданных входных сигналов всегда можно предсказать все будущие состояния. Это напоминает утверждение Лапласа о том, что из полного описания состояния Вселенной в конкретный момент времени на основе положений и скоростей всех частиц возможно предсказать все ее будущие состояния». Однако, как мы понимаем, такая сложная и многогранная система как Вселенная не может быть вполне дискретна, так как малые отклонения при начальных условиях могут обернуться колоссальными расхождениями впоследствии. Поэтому предсказание будущего состояния Вселенной дело, близкое, скорее, к утопии.

Важно понимать, что для рассматриваемых нами машин такие явления отсутствуют. «Даже когда мы рассматриваем реально существующие машины вместо идеализированных, допустимо точное знание состояния машины в некий момент времени позволяет допустимо точно предугадать ее состояния на любое количество шагов впоследствии. <…> При наличии таблицы (свода команд), соответствующей некоей машине с дискретными состояниями,

можно предсказать, что будет делать эта машина. Нет причин, по которым такие вычисления невозможно было бы выполнить при помощи цифрового компьютера. Если вычисления можно производить на сравнительно высокой скорости, цифровой компьютер окажется способным имитировать деятельность любой машины с дискретными состояниями. Тогда в игре в имитацию могли бы участвовать машина с дискретными состояниями (в качестве игрока B) и имитирующий ее цифровой компьютер «в качестве игрока A), и исследователь не смог бы отличить их друг от друга. <…>

Это особое свойство цифровых компьютеров – способность имитировать деятельность любой машины с дискретными состояниями – как раз подразумевают под утверждением, что такие машины являются универсальными».

Тогда отпадает и необходимость конструирования новых машин под каждую конкретную задачу. Достаточно должным образом запрограммировать цифровой компьютер. Отсюда, в некотором роде, вытекает эквивалентность всех таких компьютеров.

С учетом универсальности цифровых компьютеров наш генеральный вопрос сводится к следующей посылке: «Возьмем в качестве примера конкретный цифровой компьютер C. Справедливо ли будет сказать, что, модифицируя этот компьютер с таким расчетом, чтобы увеличивались емкость его памяти и быстродействие, и загружая в него соответствующую программу, можно добиться того, чтобы C удовлетворительно заменил игрока A в нашей игре в имитацию, а роль B исполнял при этом человек?».

Противоположные точки зрения на основной вопрос

После того, как все исходные положения сформулированы, можно перейти к обсуждению означенного вопроса. Важно заметить, что на момент написания статьи (1950 год) Тьюринг рисковал с изначальной постановкой вопроса «Могут ли машины мыслить» быть неверно понятым, а сама формулировка вопроса могла быть истолкована по-разному. Здесь он высказывает пророческую мысль о том, что в следующем столетии такой вопрос не будет вызывать подобных проблем. Нет ничего постыдного, говорит Тьюринг, в следовании предположениям и гипотезам, поэтому не имея сейчас достаточной вычислительной мощности компьютеров, тем не менее, рассуждения о том, что подобные мощности будут достигнуты в будущем лишь открывают новые горизонты исследований для ученых.

Отсюда на любую гипотезу находятся точки зрения, прямо или косвенно ей противоположные, опровергающие. Именно их Тьюриг и рассматривает, давая краткий ответ. Далее рассмотрим суть каждого такого возражения и кратко приведем суть ответного положения.

(1) Теологическое возражение

Суть: Мышление представляет собой свойство бессмертной души человека. Господь наделил бессмертной душой каждого мужчину и каждую женщину, но не дал души животным или машинам. Следовательно, никакое животное или машина не способны мыслить.

Ответ: Я счел бы данное утверждение более убедительным, если бы животные были включены в один класс с людьми, так как, на мой взгляд, налицо более существенное различие между типичным одушевленным и типичным неодушевленным предметами, чем между человеком и другими животными. Произвольный характер ортодоксальной точки зрения становится очевиднее, кода мы попробуем оценить ее в сопоставлении с доктринами иных вероисповеданий. Как, например, воспримут христиане мусульманское утверждение о том, что женщины лишены души? <…> Мне представляется, что приведенный довод подразумевает серьезное ограничение всесилия Всемогущего. Из этого довода следует, что на свете есть нечто, Ему неподвластное, скажем, сделать единицу равной двум; но разве кто-либо из верующих откажется признать, что Он вправе наделить душой слона, если решит, что это необходимо? Мы могли бы ожидать, что Он использует Свое всесилие только в сочетании с мутациями, которые усовершенствовали бы слоновий мозг в той степени, какая требуется для принятия души. Точно такое же обоснование применимо к машинам. <…> Пытаясь конструировать наши мыслящие машины, мы ни в коем случае не стремимся непочтительно узурпировать Божью способность сотворения душ – не больше, чем все люди

при воспроизводстве потомства; скорее, в обоих случаях мы выступаем орудиями и проводниками Его воли и создаем прибежища для душ, которые Он творит.

(2) «Голова в песке»

Суть: Последствия появления мыслящих машин могут оказаться слишком жуткими. Будем надеяться и верить, что такого никогда не произойдет.

Ответ: Нам нравится верить в то, что человек в некоем трудноуловимом отношении превосходит остальные творения Божьи. Еще лучше, когда возможно показать, что он превосходит прочих по необходимости, ибо тогда для нет ни малейшей опасности утратить свое доминирующее положение. <…> Указанное ощущение присуще, как мне кажется, многим интеллектуалам, поскольку они ценят силу мышления выше, чем все остальные, и более склонны основывать свою веру в превосходство человека над этой силой.

Я не считаю этот довод достаточно существенным для того, чтобы предлагать его опровержение. Здесь куда уместнее было бы утешение: не поискать ли его в учении о переселении душ?

(3) Математическое возражение

Суть: Ряд положений математической логики можно использовать для того, чтобы показать, что существуют ограничения на способности и возможности машин с дискретными состояниями. Наиболее известным из них является теорема Геделя, которая гласит: в любой достаточно мощной логической системе можно сформулировать утверждения, которые не могут быть доказаны или опровергнуты внутри данной системы, если сама система непротиворечива. <…> Если описанная машина демонстрирует определенное сравнительно простое отношение к машине, которой мы задаем вопросы, то можно показать, что либо ее ответ будет неправильным, либо не поступит вовсе. Это математический результат, и утверждается, будто он доказывает ущербность машин в сопоставлении с человеческим интеллектом.

Ответ: Пускай достоверно известны ограничения на возможности какой-либо конкретной машины, лишь утверждается, без всяких доказательств, что подобные ограничения не распространяются на человеческий интеллект. <…> Всякий раз, когда одной из этих машин задают соответствующий важный вопрос и она дает конкретный ответ, мы твердо знаем, что данный ответ должен быть ошибочным, и это знание наделяет нас определенным чувством превосходства. Не является ли указанное чувство иллюзорным? <…> Мы

слишком часто сами даем неверные ответы на вопросы, чтобы удовлетворение, возникающее у нас при наглядных доказательствах ущербности машин, было оправданным. Кроме того, наше превосходство мы демонстрируем лишь над машиной, над которой одержали свой довольно скромный триумф. <…> Люди могут быть умнее любой машины, но найдутся и более умные машины, чем вот эта конкретная, и так далее.

(4) С точки зрения сознания

Суть: Возражение отлично сформулировано профессором Джефферсоном (Листеровская речь 1949 года): «Лишь когда машина сможет написать сонет или сочинить концерт, на основании мыслей и испытанных чувств, а не в результате случайного выпадения символов, мы способны согласиться с тем, что машина сопоставима с человеческим мозгом, способна не только писать, но и сознавать, что именно пишет. Никакой механизм не в состоянии чувствовать (не просто искусственно сигнализировать при помощи несложных устройств) удовольствие от успехов, горевать от перегрева клапанов, тешить себя лестью, страдать из-за ошибок, восхищаться противоположным полом, сердиться или грустить, когда он не добивается желаемого».

Ответ: Согласно крайней форме этого взгляда, единственный способ убедиться в том, что машина мыслит, состоит в том, чтобы стать машиной и ощутить себя мыслящим. Тогда стало бы возможным поделиться своими чувствами со всем миром, но, конечно, никто бы – вполне обоснованно – не обратил на них никакого внимания. Исходя из такой точки зрения, мы делаем вывод, что единственный способ узнать о способности другого человека к мышлению – стать этим конкретным человеком. По сути, перед нами точка зрения солипсиста. Быть может, то сугубо логическое представление о мире, но оно затрудняет общение и обмен идеями. А обязан считать, что «А мыслит, а В не мыслит», тогда как В думает, что «В мыслит, а А не мыслит». Вместо того чтобы постоянно спорить по этому поводу, из вежливости принято предполагать, что все люди мыслят.

(5) На основании различных форм недееспособности машин

Суть: Я согласен, что возможно заставить машины выполнять все, что вы перечислили, но вы никогда не сможете заставить их делать X. Под X понимаются разнообразнейшие виды деятельности.

Ответ: Мне представляется, что данное возражение порождается смешением двух разновидностей ошибок: ошибок функционирования и ошибок вывода. Ошибки функционирования вызваны какими-то механическими или электрическими неисправностями, из-за которых машина ведет себя иначе, чем предполагалось. В философских дискуссиях допустимо

игнорировать возможность возникновения таких ошибок; поэтому в подобных дискуссиях обсуждаются «абстрактные машины». Они представляют собой математические модели, а они по определению неспособны совершать ошибки функционирования. В этом смысле справедливо утверждение, что «машины никогда не ошибаются». Ошибки же совершаются лишь тогда, когда выходному сигналу устройства придается некий смысл. <…>

Утверждение о том, что машина не в состоянии осознавать себя, обретает смысл лишь в том случае, если возможно показать, что машина мыслит и ее мысли имеют предметное содержание. Выражение «предметные содержание машинных операций», по-видимому, не является бессмысленным, по крайней мере, для людей, которые занимаются машинными вычислениями. Если, например, машина пыталась найти решение уравнения $x^2 - 40x - 11 = 0$, возникает искушение описать это уравнение как часть машинных операций в конкретный момент времени. В этом отношении машина, безусловно, старается постичь самое себя. Ей можно поручить создание собственной программы для предсказания последствий внесения изменений в ее конструкцию. Наблюдая за собственным поведением, машина может модифицировать свою программу для более эффективного достижения какой-либо цели. Отмечу, что все это – возможности ближайшего будущего, а не утопические мечты.

Возражение по поводу того, что машина не способна к многообразию поведения – всего-навсего способ объяснить, что ей не хватает емкости памяти.

(6) Возражение Леди Лавлейс

Суть: Аналитический вычислитель не притязает на то, чтобы *создавать нечто новое*. Зато он может *делать все*, что мы *сумеем* ему предписать. Это не означает, что невозможно сконструировать электронное устройство, способное «думать за себя», или в котором, прибегая к биологической терминологии, возникали бы условные рефлексы, открывающие возможности для «обучения». Осуществимо ли подобное в принципе – вот стимулирующий и побуждающий к творчеству вопрос, поставленный недавними достижениями инженерии. Но не создается впечатления, что машины, построенные или спроектированные до настоящего времени, обладают этим свойством.

Ответ: Кто может быть уверен в том, что «оригинальная работа», которую он проделал, не проросла из семени, посаженного образованием, или не явилась следствием применения общеизвестных принципов? Корректнее всего, пожалуй, было бы сказать, что машина никогда и ничем не сможет

удивить человека. Это утверждение служит прямым вызовом, и мы можем принять его без колебаний. Меня самого машины удивляют очень часто. Во многом это объясняется тем, что я не могу точно рассчитать, чего от них ожидать, а еще потом, что я пусть и прибегая к расчетам, выполняю вычисления поспешно и рискуя ошибиться. Представление о том, что машины не могут удивить человека, проистекает, по-моему, из заблуждения, которому в особенности подвержены философы и математики. Речь о предположении, что, едва некий факт сделался достоянием разума, все его следствия точно так же переходят в достояние разума. Это предположение весьма полезно при многих обстоятельствах, но слишком часто забывается, что по сути оно ложно. Естественным следствием из него является мнение, что будто бы нет ничего особенного в умении делать выводы из накопленных данных и общих принципов.

(7) На основании непрерывности нервной системы

Суть: Малая ошибка в информации относительно силы нервного импульса, действующего на нейрон, способна существенно повлиять на силу импульса на выдохе. С учетом этого можно выдвинуть предположение о том, что нельзя ожидать имитации деятельности нервной системы от машины с дискретными состояниями.

Ответ: Машина с дискретными состояниями должна отличаться от машины непрерывного действия. Если мы придерживаемся условий нашей игры в имитацию, исследователь не сможет воспользоваться этим различием. <…> Тут отлично подойдет в качестве примера дифференциальный анализатор. (Это машина определенного типа, не относящаяся к числу машин с дискретными состояниями и используемая для некоторых видов вычислений). Отдельные дифференциальные анализаторы выдают ответы в печатной форме и поэтому подходят для участия в нашей игре. Цифровой компьютер вряд ли способен точно предсказать, какие решения будет предлагать дифференциальный анализатор, но он вполне может сам выдать правильный ответ. Например, если требуется найти значение числа π (фактически приблизительно 3,1416), было бы разумно выбирать случайным образом между значениями 3,12; 3,13; 3,14; 3,15; 3,16 с вероятностями выбора, соответственно, 0,05; 0,15; 0,55; 0,19; 0,06. В таких условиях для исследователя было бы крайне сложно отличить дифференциальный анализатор от цифрового компьютера.

(8) На основании неформальности поведения

Суть: Невозможно создать набор правил, описывающих действия человека в любых вообразимых обстоятельствах.

Ответ: Из этого возражения формируется недостоверное утверждение: «Будь у каждого человека определенный набор правил действия, которыми он руководствовался бы в жизни, такой человек был бы не лучше машины. Но подобных правил нет, поэтому люди не могут быть машинами». <…>

На мой взгляд, налицо известная путаница между «правилами поведения» и «законами поведения», которая препятствует решению задачи. Под «правилами поведения» я имею в виду предписания вида «Остановитесь, если горит красный свет»; эти предписания служат основой наших действий и осознаются нами. Под «законами поведения» я понимаю законы природы в применении к человеческому телу, «если человека ущипнуть, он вскрикнет». Если заменить выражение «правила действия, которыми человек руководствуется в жизни» на выражение «законы поведения, управляющие человеческой жизнью» в приведенном выше рассуждении, пропасть нераспределенности среднего термина уже не покажется непреодолимой. Ведь мы считаем подчиненность законам поведения признаком машины (пускай необязательно машины с дискретными состояниями) и убеждены в том, что, наоборот, быть такой машиной значит подчиняться указанным законам. При этом мы затрудняемся убедить себя в отсутствии полного свода законов поведения, хотя уверены в отсутствии полного свода правил действия.

(9) На основании экстрасенсорного восприятия (ЭСВ)

Суть: Многие научные теории, похоже, остаются применимыми на практике, выдержав столкновение с ЭСВ; что в действительности можно обойтись без ЭСВ, благополучно забыв о его существовании. Впрочем, это довольно сомнительное утешение, ведь имеется опасение, что мышление принадлежит к числу тех явлений, которые относятся к ЭСВ (телепатия, ясновидение, прорицание, телекинез).

Ответ: Давайте играть в имитацию, и участниками нашей игры будут человек, способный воспринимать телепатические воздействия, и цифровой компьютер. Исследователь вправе задавать такие вопросы, как: «Какой масти карта в моей правой руке?». Человек, благодаря телепатии или ясновидению, отвечает правильно в 130 случаях из 400. Машина же только догадывается случайным образом и, возможно, угадывает в 104 случаях, поэтому исследователь корректно отождествляет игроков. Здесь открывается интересная возможность. Предположим, что цифровой компьютер содержит генератор случайных чисел. Тогда естественно будет применять при ответах этот генератор. Но ведь генератор случайных чисел окажется под воздействием психокинетической силы исследователя. Возможно, такое направленное воздействие побудить машину угадывать правильные варианты ответа чаще, чем можно было бы ожидать по расчетам вероятности, и поэтому исследователь окажется не в состоянии произвести корректное

отождествление игроков. С другой стороны, он может сам угадать, кто человек, а кто машина, не задавая никаких вопросов, исключительно посредством ясновидения. С ЭСВ возможно что угодно.

Если допустить существование телепатии, наш критерий необходимо уточнить. Ситуацию можно считать аналогичной той, какая могла бы сложиться, если бы исследователь разговаривал сам с собой, а один из игроков-соперников его подслушивал, прижавшись ухом к стене. Помещение игроков в «комнату, защищенную от чтения мыслей» удовлетворит всем требованиям нашей игры.

Обучаемые машины

Покончив с предвосхищением возможных возражений, Тьюринг переходит к главной, как нам представляется, сути разума, интеллекта, мыслящего – способности к обучению, развитию, если хотите, эволюции своего собственного ума. Если уж машина в его статье показана как «нечто мыслящее», то она просто обязана уметь обучаться, развиваться, словом, проходить те этапы «взросления», которые проходит каждый человек. Приведем главные выдержки из этих размышлений.

Большинство умов видятся нам «субкритическими». Идея, проникшая в подобный ум, будет в среднем выдавать меньше одной ответной идеи. Лишь небольшую долю умов следует признать «сверхкритической». Идея, проникшая в такой ум, способна породить целую «теорию», включающую в себя вторичные, третичные и более отдаленные идеи. Что касается животных, их умы, судя по всему, явно субкритические. Следуя нашей аналогии, мы спрашиваем: «Может ли машинный разум быть сверхкритическим?». <…>

Необходимо дождаться конца столетия и провести предложенный эксперимент. Но ведь некоторые мысли можно сформулировать уже сейчас. Какие шаги следует предпринять сегодня, если допустить, что эксперимент будет успешным?

Проблема заключается, в основном, в правильном программировании. <…> емкость «хранилища» в 10^7 единиц представляется мне вполне практичным решением даже по нынешним меркам. <…> Элементы современных машин, которые можно трактовать как аналоги нервных клеток, работают в тысячу раз быстрее последних. Этого должно быть достаточно для создания «запаса надежности», который компенсирует потери скорости, характерные для многих процессов. <…>

Пытаясь подражать разуму взрослого человека, мы обязаны постоянно уделять внимание тому процессу, который позволил этому разуму обрести нынешнее состояние. Можно выделить три этапа этого процесса:

(а) начальное состояние разума, скажем, при рождении;
(б) образование, полученное человеком;
(в) иной полученный опыт, который следует отличать от образования.

<…> Наше рассуждение исходит из того, что механизм детского разума чрезвычайно прост, в потому легко поддается программированию. При этом можно допустить, что объем знаний, которые нужны для обучения машины, в первом приближении будет таким же, как и объем знаний для ребенка, получающего образование.

Таким образом, мы разделили нашу проблему на две части: «программа-ребенок» и образовательный (воспитательный) процесс. Эти две части остаются тесно взаимосвязанными. Нельзя рассчитывать на то, что удачная «машина-ребенок» получится у нас с первой попытки. Нужно экспериментировать с обучением одной машины и оценить, насколько хорошо она обучается. Затем можно обратиться к другой машине и оценить, обучается она лучше или хуже первой, и так далее. Существует очевидная связь между этим процессом и эволюцией, которая проявляет себя в следующих отождествлениях:

Структура «машины-ребенка» = Наследственность

Изменения = Мутации

Естественный отбор = Суждение экспериментатора

<…> Конечно, образовательный процесс для машины должен отличаться от обучения детей. К примеру, машина не имеет ног, поэтому ее нельзя попросить выйти наружу и принести угля для отопления. Еще, как правило, у машины нет глаз. И потом, при всех достижениях инженерии, которые позволили бы преодолеть недостатки, мы не сможем отправить машину в школу просто так, поскольку дети-одноклассники будут над нею потешаться. Перед посещением школы машина должна пройти некое начальное обучение. Нам не стоит чрезмерно беспокоиться о ногах, глазах и так далее. Пример мисс Хелен Келлер [1]показывает, что образование возможно получать при условии, что коммуникация в обоих направлениях между учителем и учеником осуществляется посредством каких-либо каналов связи.

Обычно учебный процесс подразумевает для широкой публики различные наказания и поощрения. Некоторые простые «детские» машины могут быть сконструированы или запрограммированы по такому же принципу. Машину можно сконструировать так, чтобы события, которые непосредственно предшествуют поступлению сигналов о наказании, возникали бы реже, тогда как поступление сигналов о поощрении увеличивало бы вероятность повторения событий, которые привели к их поступлению. Эти сигналы, разумеется, не подразумевают никаких чувств со стороны машины. <…>

Использование наказаний и поощрений в лучшем случае может быть только частью учебного процесса. Грубо говоря, если у учителя нет других средств коммуникации с учеником, то объем информации, который последний может усвоить, не превышает общего объема полученных поощрений и

[1] Слепоглухонемая американка, сумевшая получить высшее образование, впоследствии автор ряда книг. – прим. пер.

наказаний. <…> Необходимы некоторые другие, «внеэмоциональные» каналы связи. Если такие каналы доступны, можно с использованием наказаний и поощрений обучать машину подчиняться приказам, отданным на каком-либо языке, например, на символическом. Эти приказы должны передаваться по «неэмоциональным» каналам. Употребление такого языка значительно снизит количество наказаний и поощрений. <…>

Память устройства будет в значительной степени занята определениями и высказываниями, причем последние будут иметь различный вид, например, включать в себя установленные факты, гипотезы, математически доказанные теоремы, утверждения авторитетных личностей, выражения, имеющие логическую форму высказывания, но не претендующие на верность. Некоторые из этих положений можно охарактеризовать как «императивы». Машину следует сконструировать или запрограммировать таким образом, что, как только она классифицирует императив как «установленный», подходящее действие будет выполняться автоматически. <…>

Императивы, которые могут выполняться машиной, лишенной конечностей, должны относиться к области *интеллектуальной деятельности*. <…> наиболее важными среди таких императивов будут те, которые регулируют порядок применения правил соответствующей логической системы. <…> Выбор конкретной операции подразумевает отличие блестящего мыслителя от мастера тривиальных суждений, а не различие между тем, кто изрекает истину, и тем, кто ошибается. <…>

Некоторые из высказываний могут быть «одобрены авторитетным лицом», тогда как другие могут формироваться самой машиной, скажем, на основании научной индукции. <…>

Как должны изменяться в этом случае правила функционирования машин? Ведь эти правила должны досконально описывать, как машина будет реагировать на команды, независимо от ее предыстории и от изменений, которым она подверглась. Таким образом, правила оказываются вынесенными за временные рамки. Это совершенно верно. Объяснение парадокса заключается в том, что правила, которые изменяются в процессе обучения, имеют, скорее, преходящую ценность и притязают лишь на эфемерность истины. <…>

Важной особенностью обучаемой машины является то обстоятельство, что ее учитель зачастую оказывается во многом не осведомлен о том, что происходит внутри устройства, хотя он способен в какой-то степени предсказывать поведение своего ученика. <…>

Мнение о том, что «машина способна не делать только то, что мы сможем ей приказать», в этом случае выглядит довольно странно. Большинство

программ, которые мы можем загрузить в машину, вызовут такое ее поведение, которое мы будем не в состоянии понять вообще или которое мы посчитаем совершенно случайным. Интеллектуальное поведение, по-видимому, предполагает отказ от строгого соблюдения «дисциплины», обязательного при выполнении вычислений, и в обретении некоторой свободы действий, каковая не приведет, впрочем, ни к беспорядочному поведению, ни к бессмысленным повторам операционных циклов. <…>

Процессы обучения машины, о чем нужно помнить, не обеспечивают стопроцентную гарантию результата, иначе от усвоенных знаний и навыков невозможно было бы избавиться.

Пожалуй, обоснованно наделить обучаемую машину случайным элементом. Такой элемент весьма полезен, когда мы ищем решение какой-либо задачи. Предположим, например, что нам нужно найти в ряду от 50 до 200 число, равное квадрату суммы его цифр; мы могли бы начать с 51, перейти к 52 и двигаться дальше до получения нужного числа. Или же мы могли бы перебирать цифры наугад, пока не получим желаемый результат. Преимущество данного метода состоит в том, что нет необходимости хранить в памяти уже проверенные варианты, а недостаток заключается в том, что некоторые цифры могут проверяться повторно, однако это не слишком существенно при условии, что задача имеет несколько решений. <…> Процесс обучения можно рассматривать как поиск той формы поведения, которая вызывала бы одобрение учителя (или удовлетворяла бы какому-то иному критерию). Поскольку существует, вероятно, очень большое количество решений, отвечающих нашим требованиям, случайный выбор кажется предпочтительнее систематического. <…>

Мы можем надеяться, что машины в конечном счете станут конкурировать с людьми во всех чисто интеллектуальных областях. Но какие из этих областей наиболее пригодны, чтобы начать именно с них? <…>

Мы способны заглянуть в будущее лишь на небольшое расстояние, но очевидно, что сделать предстоит очень многое.

Заключение

Вместе с Аланом Тьюрингом мы имели возможность задаться вопросом, способна ли машина мыслить и какова природа подобного мышления? Можем ли мы достоверно рассуждать о мыслительном процессе, если и сами до конца не можем определить или описать его. Более того, не получив знания о локализации разума, сможем ли мы вообще когда-нибудь приблизиться к достоверной классификации «разумное-неразумное».

Тьюринг показывает, что мы должны освободиться от догматики. Человеку должно признавать себя венцом творения, вершиной пищевой цепочки, царем природы и мира. Однако такое положение заводит человека в определенные рамки самоосознания и самоидентификации, порождает порочный скептицизм и недоверие. Заставляет думать «как удобнее», а не стараться приближаться к тому «как на самом деле».

Ребенок, рожденный свободным от убеждений, имеющим чистый ум, взрослея, попадает в эти рамки, будучи воспитанный в человеческом обществе. От таких социальных тяжеловесных кандалов избавиться довольно сложно, но хочется тогда задать вопрос сугубо риторический, будем ли мы так непревзойденны, отождествлены со Всемогущим и Всесильным, если не сможем справиться с самими собой?

Если взглянуть на поставленный в 1950 году вопрос глазами современного человека, находящегося на отдалении в 70 лет, ну что, теперь-то машины мыслят? Можем ли мы сыграть в имитацию и получить удовлетворительные результаты?

Да, мы знаем о триумфе машинного вычисления при игре в шахматы, когда знаменитый Чемпион мира гроссмейстер Гарри Каспаров потерпел поражение от компьютера шесть партий подряд (речь идет о суперкомпьютере Deep Blue компании IBM, разработанном в конце 90-ых и «игравшего» с Каспаровым 11 мая 1997 года).

Мы также знаем о самых современных разработках в области создания искусственного интеллекта и часто слышим о том, что какие-то устройства или «роботы» уже им наделены.

Нам такие высказывания представляются не более чем красивой маркетинговой компанией. Важнейшими свойствами интеллекта являются самосознание, самоидентификация, различная система чувств (хотя последняя и относится больше к эмоциональной составляющей человека и не является чисто интеллектуальной). Сегодня же те фрагменты искусственного интеллекта, как то нейронная сеть или самообучающиеся алгоритмы являют собой не более чем сложный набор математических моделей. Вопрос, каким

образом, фактически, формула может сама себя осознать пока что скромно повисает в воздухе. Мы призываем не переходить на сторону строгого отрицания того, что некая фраза, записанная на формальном языке математики, неспособна находиться в отношении с самой собой (иначе как объяснить явление рекурсии или рекуррентных соотношений, которым в том числе подчинена и природа?).

Метафизическая природа самой математики и ее языка приближает нас, в самой своей сути, к природе и нашего такого же метафизического разума. Быть может, именно творческая проблема математики, как то «математизация выбора» и являет собой последний генеральный вопрос, который позволяет математике оставаться наукой. Быть может, именно этот вопрос и являет собой очередную форму первоначального вопроса Тьюринга.

Нашей задачей является развивать интеллектуальные и инженерные достижения всего рода человеческого, находить приближения к своей собственной природе, чтобы наконец, если представится возможность, ответить на главные вопросы, кто мы такие и каково наше предназначение? Сделано уже многое, но предстоит сделать еще больше.