



Марвин Минский

ПОЧЕМУ ЛЮДИ ДУМАЮТ, ЧТО КОМПЬЮТЕРЫ НЕ МОГУТ...

О Марвине Минском и его знаменитой книге «Society of Mind» («Общество разума») мы уже писали в №№ 3-6 за 1998 год. Публикуемая статья была написана еще до выхода этой книги и переводится впервые. Английский вариант можно найти по адресу www.ai.mit.edu/people/minsky/paper/ComputersCantThink.txt. В этой статье сконцентрированы главные и не потерявшие актуальности идеи М. Минского по «компьютерной психологии».

За прошедшие годы технические возможности компьютеров возросли многократно, поэтому следует ожидать, «что теперь дойдет очередь и до практического воплощения компьютерного разума».

Большинство людей считают, что компьютеры никогда не смогут думать. То есть, думать по-настоящему. Ни сейчас, ни когда-либо в будущем. С другой стороны, большинство людей наверняка согласны с тем, что компьютеры умеют делать много такого, для чего необходимо уметь мыслить. Но тогда как же машина может казаться думающей и не быть таковой на самом деле? Оставляя в стороне вопрос о том, что такое мышление, я полагаю, что большинство из нас сказали бы: то, что делает компьютер – это просто внешняя имитация человеческого интеллекта. Компьютер был сконструирован, чтобы подчиняться некоторым простым командам, а затем был снабжен программой, составленной из этих команд. Поэтому компьютер должен подчиняться этим командам, не имея ни малейшего представления о сути происходящего.

Действительно, когда компьютеры появились впервые, большинство их создателей предназначало их для выполнения громоздких бездумных вычислений. Поэтому они и были названы компьютерами. И все-таки даже тогда некоторые

из первооткрывателей – особенно Алан Тьюринг – предвидели то, что теперь называется Искусственным Интеллектом (ИИ). Они поняли, что компьютеры смогут, вероятно, выйти за пределы арифметики и, может быть, даже имитировать процессы, происходящие в мозгу человека.

В настоящее время, когда роботы проникли повсюду – и в промышленность и в кинофильмы, большинство людей полагает, что ИИ продвинулся гораздо дальше, чем это есть на самом деле. И все же «знатоки компьютеров» утверждают, что машины никогда не будут мыслить по-настоящему. Если это так, то как они могут быть такими сообразительными и в то же время такими тупыми?

МОЖЕТ ЛИ МАШИНА БЫТЬ ТВОРЧЕСКОЙ?

Мы восхищаемся нашими Эйнштейнами и Бетховенами и задаем себе вопрос: могли бы компьютеры создать такие чудесные теории и симфонии. Большинство людей думают, что способность творить связана с каким-то особым магическим даром, который просто нельзя объяс-

нить. Если это верно, то никакой компьютер не сможет творить, так как, по их мнению, все, что делают машины, может быть объяснено.

Чтобы понять, что здесь неверно, нужно избежать одной простенькой ловушки. Мы не должны обращаться к творениям, которые наша культура рассматривает как великие, пока не поймем, каким образом обычные люди делают обычные вещи. Мы не можем рассчитывать на то, что разгадаем сразу, как великие композиторы пишут великие симфонии. Я не верю, что существует большое различие между обычным и творческим мышлением. Я не могу никого винить за то, что он не может делать все, что делают высоко творческие люди, а также и за то, что они не могут объяснить это. Я возражаю против мысли, что никто не сможет никогда понять, как происходит творчество, только потому, что мы не можем это объяснить сейчас.

Нас не должно обескураживать наше восхищение Бетховенами и Эйнштейнами. Вместо этого, нас должно раздражать наше невежество относительно того, как у нас возникают идеи, причем не только творческие. Мы так привыкли удивляться необычному, что забываем, как мало мы знаем о чуде обычного мышления. Может быть, наши суеверия относительно творчества служат каким-то другим потребностям, например, обеспечивая нас героями с особыми качествами, на фоне которых наши недостатки кажутся более простительными.

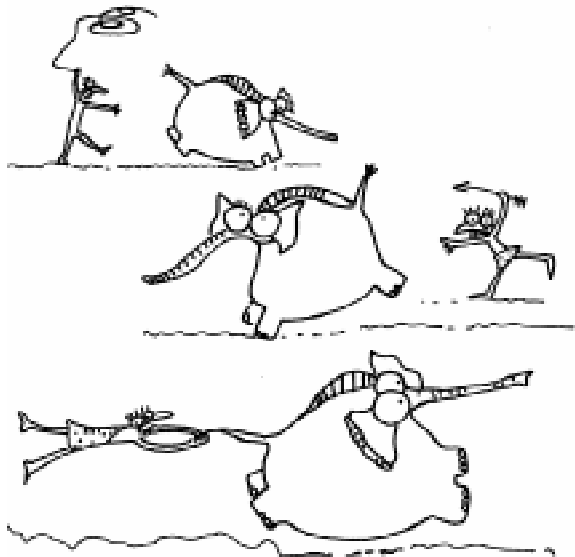
Отличаются ли выдающиеся умы от обычных чем-то особенным? Я не верю, что в гении есть что-то принципиально особенное, за исключением необычного сочетания способностей, ни одна из которых не является сама по себе особенной. Должен быть глубокий интерес к какому-то предмету, но это достаточно обычное явление. Далее, должно быть основательное владение предметом – это также



Я не верю, что в гении есть что-то принципиально особенное, за исключением необычного сочетания способностей...

бывает не столь уж редко – мы называем это мастерством. Нужна также достаточная уверенность в себе, чтобы выдержать презрение окружающих – мы это называем упорством. И, наконец, должен быть здравый смысл. В моем представлении, любой человек, понимающий обычный разговор, обладает большей частью из того, что имеют наши герои. Тогда почему обычный здравый смысл, если его лучше сбалансировать и снабдить более сильной мотивацией, не может превратить любого человека в гения?

Но тут необходимо задать вопрос, почему не каждый человек обладает такой комбинацией свойств? Прежде всего, конечно, иногда обретение нового взгляда на вещи – просто дело случая. Но затем начинают сказываться количественные различия. Одно из них – в том, как распоряжаются люди тем, что изучают. В



...не могла бы выжить никакая культура, если бы каждый человек искал свои способы мышления...

основе мастерства творческих людей лежит бессознательное умение связывать воедино многое из того, что они знают. Другое отличие состоит в том, что они во время учебы узнают намного больше и лучше. Хороший композитор свободно владеет фразировкой и умело развивает музыкальную тему, но то же самое делает любой, владеющий связной речью.

Почему некоторые люди учатся так хорошо и так много? Простейшая гипотеза состоит в том, что они наткнулись на лучшие методы обучения! Возможно, их «дар» – всего лишь умственный навык. Так же, как один ребенок учится наилучшим образом расставлять свои кубики, играя с ними, другой – может учиться, мысленно перебирая разные варианты.

Наша культура не поощряет нас анализировать обучение. Мы просто воспринимаем его как нечто происходящее с нами. Однако обучение должно само состоять из множества навыков, которые мы вырабатываем у себя. Мы начинаем с обучения некоторым из них и медленно вырабатываем остальные. Почему только некоторые продолжают учиться и совершенствовать свое умение учиться? Потому что награда приходит не сразу, усилия окупаются с большим запозданием. Когда дети играют в песочнице, их цель обыч-

но состоит в том, чтобы наполнить ведерки песком. Но если ребенок думает о большей цели, тогда скорость обучения возрастает экспоненциально! Совершенствуя свое умение учиться, мы тем самым повышаем и эффективность самой учебы, поднимая ее на новый качественный уровень. Таким образом, творческая одаренность может быть всего лишь результатом ряда случайных эпизодов из детства.

Почему же гении так редки, если у каждого есть почти все, что нужно для этого? Возможно, потому что наша эволюция относится с бездумной небрежностью к индивидууму. Я уверен, что не могла бы выжить никакая культура, если бы каждый человек искал свои способы мышления. Печально, если это так, поскольку это означает, что гены гениальности нужно не культивировать, а почаще выпалывать.

НЕРЕШЕННЫЕ ПРОБЛЕМЫ

Мы едва ли можем рассчитывать на то, что заставим машины творить чудеса, до того как добьемся, чтобы они совершали обычные разумные действия. Первые компьютерные программы мало отличались от простого списка команд и циклов команд типа «Делай это. Делай то. Делай это и то, пока не произойдет то». Большинство людей все еще пишет программы на таких языках, как BASIC или FORTRAN, которые вынуждают нас заранее представлять все, что наша программа будет делать на каждом шаге. Назовем это программированием типа «Делай сейчас».

Весьма давно исследователи ИИ открыли новые способы создания программ. Allen Newell, J.C.Shaw и Herbert A.Simon создали в конце 50-х годов систему «General Problem Solver», в которой реализованы способы описания процессов в таких терминах «Если разница между тем, что вы получили, и тем, что вы хотели получить, имеет тип D, тогда попытайтесь изменить эту разницу, используя метод M». Эти и другие идеи привели к тому, что мы называем методами типа «средство-цель» и «делай, если нужно». Такие программы автоматически применяют определенные

правила там, где это нужно, так что программист не должен думать заранее, где это произойдет. Это открыло эпоху программ, решавших задачи такими способами, которые создатели этих программ не могли заранее предугадать, поскольку программе сообщается, какие типы действий она может совершать, но неизвестно, какое из этих действий будет реализовано. Каждый знает, что если вы достаточно долго наугад выбираете различные способы решения, то в конце концов вы сможете сделать все, что угодно. Однако, если на это уйдут миллиарды лет (как у обезьяны, бьющей по клавишам пишущей машинки), то это не работа интеллекта, а просто эволюция. Новые системы работали не наугад, а использовали «советы» о том, какой путь, вероятно, более пригоден для решения той или иной проблемы. Таким образом, вместо случайного блуждания такие программы двигаются ощупью, подобно тому, как мы карабкаемся на гору в темноте, двигаясь всегда вверх по склону. Единственная проблема – не застрять на малой промежуточной вершине, так и не добравшись до главной.

С тех пор многие исследователи ИИ пытались найти более «глобальные» методы, позволяющие избежать «застревания». Однако никому не удалось открыть «абсолютно общий» способ отыскания наилучшего метода, и никто такого открытия не ожидает.



Единственная проблема — не застрять на малой промежуточной вершине...

Вместо этого, многие исследователи ИИ нацелены на программы, которые перебирают содержащиеся в их памяти модели, чтобы решить, что делать дальше. Такое программирование я назвал бы «Сделай что-нибудь разумное». Очень немногие исследователи – слишком немногие, я полагаю, – экспериментируют с программами, которые могут обучаться и рассуждать по аналогии. Эти программы смогут когда-нибудь распознавать, что именно из предыдущего опыта, содержащегося в памяти компьютера, имеет наиболее близкую аналогию к новой ситуации, и выбрать из памяти метод, который лучше всего работал в аналогичных задачах в прошлом.

МОГУТ ЛИ КОМПЬЮТЕРЫ ПОНИМАТЬ?

Можем ли мы сделать так, чтобы компьютеры понимали, что мы им говорим? В 1965 году Daniel Bobrow написал одну из первых Rule-Based Expert Systems. Она называлась STUDENT и умела решать набор школьных текстовых задач по алгебре. Например, такие:

– Расстояние между Лос-Анжелесом и Нью-Йорком 3000 миль. Если средняя скорость реактивного самолета 600 миль/час, найдите время, нужное для перелета.

– Дядя отца Билла в два раза старше отца Билла. Через два года отец Билла будет в три раза старше Билла. Их возрасты в сумме равны 92. Найдите возраст Билла.

Большинство учащихся считают такие задачи гораздо труднее, чем просто решение формальных уравнений школьной алгебры, решать которые все равно, что пользоваться поваренной книгой. Но чтобы решить неформальную текстовую задачу, вы должны выяснить, какие уравнения нужно решать, а для этого нужно сначала понять, что означают слова и предложения. Понимал ли их STUDENT? Он использовал множество приемов. Он был запрограммирован догадываться, что слово «is» обычно означает равно. Он даже не пытался выяснить, что означает «дядя отца Билла», – он только заметил, что эта фраза напоминает фразу «отец Билла». Он не знал, что слова «возраст» и «старше»

относятся ко времени, но он использовал их для обозначения чисел, подставляемых в уравнение. Используя пару сотен таких приемов для установления связей, STUDENT мог иногда получать правильные ответы.

Рискнем ли мы утверждать, что STUDENT «понимал» эти слова? Зачем об этом беспокоиться? Зачем поддаваться чувству, что необходимо определить такие старые термины, как «означать» и «понимать»? Прекрасно, когда слова помогают нам обрести хорошие идеи, и совсем наоборот, когда они запутывают нас. Вопрос нужно ставить так: игнорировал ли STUDENT «истинные значения» слов, используя свои приемы?

А может быть, то, что мы называем «значением», – просто набор подходящих приемов? Рассмотрим классический пример: что означает слово «число»? STUDENT знает кое-что из арифметики в том смысле, что он может получать утверждения типа «5 плюс 7 равно 12». Но понимает ли он, что такое число? Например, что означает «5» или «плюс» или «равно»? Что бы вы ответили, если бы я спросил у вас: «Что такое ПЯТЬ?» В начале XX-го века философы Бертран Рассел и Альфред Норт предложили новый способ определения чисел. «ПЯТЬ, – говорили они, – это множество всех множеств с пятью элементами». Такое множество содержит, в частности, множество из пяти шариковых ручек, множество из пяти котят. К несчастью, это определение приводит к причудливым противоречиям и парадоксам. Главной целью здесь было найти совершенные определения для обычных слов и идей. Но даже если эта теория (Рассела-Уайтхеда) годится для математики, необходимость обойти противоречия делает ее слишком сложной для практического использования. Деятели образования однажды пытались заставить детей пользоваться этой теорией множеств (движение Новая Математика в 60-х годах). Это еще больше отдалило тех, кто любит математику, от тех, кто ее боится.



Что такое ПЯТЬ?

Я думаю, главная причина была в том, что игнорировалось фундаментальное свойство мышления: для меня смысл любого объекта в некоторой степени зависит от моих знаний о других объектах.

Что если мы построим машины, работающие не на основе жестких определений? Не утонут ли они в парадоксах, противоречиях, двусмысленностях? Расслабьтесь! Большая часть из того, что мы знаем, переполнена противоречиями, и все-таки мы живы. Лучшее, что мы можем сделать, это проявлять разумную осторожность. Заставим же и наши машины проявлять такую же осторожность. Если же сохраняется определенная вероятность ошибки, что ж, – такова жизнь.

СМЫСЛОВЫЕ СТРУКТУРЫ

Если значение каждого объекта в нашем сознании зависит от значений других объектов, то не оказываются ли они слишком плохо определенными для того, чтобы быть объектом научного исследования? Нет, даже если значения образуют логические циклы, все равно есть предмет для научного исследования – сами эти циклы! Более ранние теории лишь пытались их скрыть, но при этом терялось все богатство нашей удивительной человеческой смысловой ткани. Сетевая структура нашего мышления, вероятно, сложнее, чем любая другая структура, когда-либо изучавшаяся наукой. Поэтому развитые тео-

рии ИИ будут, по всей вероятности, весьма сложны. Повторяем, – такова жизнь.

Вернемся к значению чисел. На этот раз для простоты рассмотрим число ТРИ. Я утверждаю, что для нас понятие ТРИ основано не на единственном базовом определении, а является сочетанием различных процессов, каждый из которых обменивается значениями с другими. Рассмотрим роли, в которых выступает число ТРИ. Во-первых, мы говорим ТРИ, когда считаем «один, два, три», указывая на различные предметы. Можно при этом считать вслух, беря каждый предмет и отодвигая его. Дети учатся делать это в уме, а если это трудно, указывают пальцем. Другой способ состоит в использовании стандартного набора из трех элементов для того, чтобы установить взаимнооднозначное соответствие между данным и стандартным наборами. Если соответствие имеется, то в данном наборе тоже ТРИ предмета. Этот стандартный набор не обязательно должен состоять из предметов, сгодятся и просто слова «один, два, три». Аналогичное рассмотрение числа ПЯТЬ дает более широкий выбор. Можно представлять себе две группы – ДВА и ТРИ или ОДИН и ЧЕТЫРЕ. Или можно представлять известные фигуры – пятиугольник, Х, самолет и т.д.

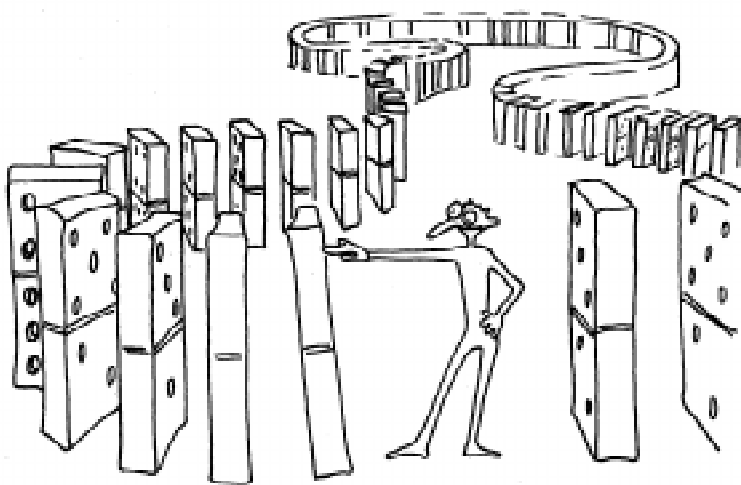
Поскольку разные способы работают в разных ситуациях, наши возможности расширяются, если мы умеем переходить

от одного способа к другому. Бессмысленно спрашивать, какой из них верен, – каждый имеет свою область применения и по-своему поддерживает другие. Каждый из них сам по себе не слишком силен, но в совокупности они образуют работоспособную универсальную систему. Вместо хрупких звеньев в цепочках определений, каждое используемое нами слово может активизировать большую смысловую структуру, содержащую различные способы рассмотрения предметов, их запоминания, сравнения и т.д. В системе знаний со множеством связей вы не можете застрять. Когда одно из значений не подходит, вы можете переключиться на другое. При строго математическом подходе, встретив малейшее препятствие, вы застреваете напрочь!

Почему же математики так привержены к структурам со слабыми связями, в которых каждый элемент должен зависеть от как можно меньшего числа других элементов? Ответ звучит иронически: они хотят застрять! Когда что-то не получается, они хотят быть первыми, кто это заметил. Лучший способ быть в этом уверенными – это устроить так, чтобы рушилось все сразу! Для них слабость связей – не недостаток, так как она помогает им найти совершенное доказательство, в котором ни один элемент не противоречил бы другим. Это хорошо для математики, в этом и состоит большая часть математики.

Однако это плохая психология. Ведь в действительности мы всегда верим во что-то, что оказывается на самом деле неверным.

Вот, я думаю, пример плохой психологии, когда учителя строят курс математики в виде длинных тонких хрупких цепочек с башнями определений вместо устойчивой структуры с перекрестными связями. Эти цепочки рвутся в своих слабых звеньях, и башни опрокидываются при ма-



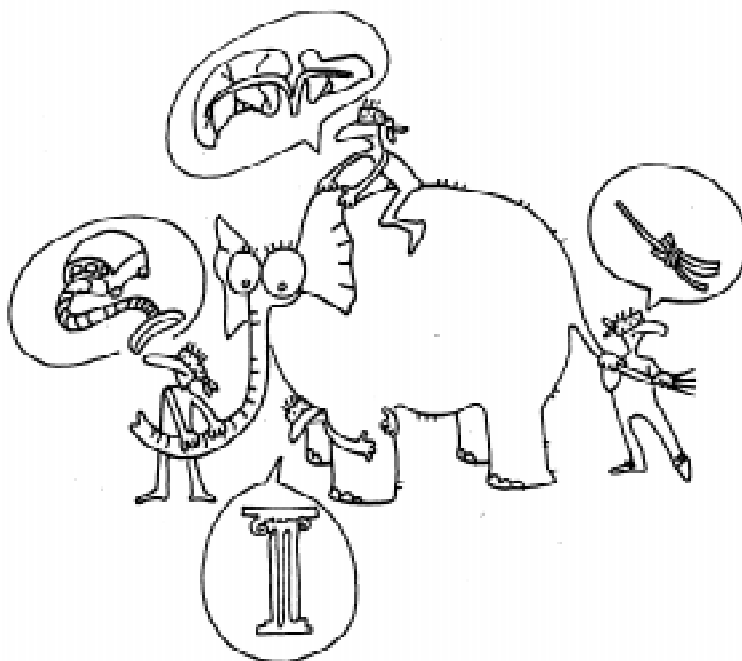
...готовы рушиться все сразу...

лейшем толчке. Это как раз то, что происходит в голове ребенка на уроке математики, когда он отвлечется на мгновение, увидев в окне красивое облако. Цели обычных людей отличаются от целей математиков и философов, которые стремятся к максимальному упрощению, сводя число связей к возможному минимуму. В реальной жизни лучшие идеи – те, которые имеют максимально возможное число взаимосвязей. Может быть, поэтому большинство детей боится математики. Мы думаем, что мы помогаем детям лучше понять предмет тем, что их действия оказываются в большинстве случаев неверными! Возможно, вместо этого, нам следует помочь им выстроить в их головах более устойчивые смысловые структуры.

ВОЗДУШНЫЕ ЗАМКИ

Значение того или иного объекта заключается в том, каким образом он связан с другими известными нам объектами. Чем больше таких связей, тем он более важен для нас. Смешно искать «истинное» значение объекта, так как если бы оно было единственным, то есть объект был бы связан только с одним другим объектом, он едва ли бы имел вообще какое-нибудь значение!

Вот почему я считаю, что мы не должны строить программы на основе ясных и простых логических определений. Такие машины никогда не смогут понимать что-либо «по-настоящему». Насыщенные многими внутренними связями структуры предоставляют нам различные пути использования знаний, так что, когда один путь не годится, можно попытаться выяснить, почему. Когда имеется много взаимосвязанных значений одного объекта, можно рассматривать его с разных точек зрения.



Когда имеется много взаимосвязанных значений одного объекта, вы можете рассматривать его с разных точек зрения...

Если одна из них приводит в тупик, можно выбрать другую. Вот что мы подразумеваем под мышлением!

Поэтому я не люблю логику и предпочитаю работать с системами определений, образующих замкнутые логические циклы. Каждое определение основано на других. Нет ничего плохого в том, что нравятся несколько разных мелодий и тем больше, чем сильнее контраст между ними. Нет ничего плохого в системе узлов или в ткани, где каждая отдельная нить помогает сцеплению остальных нитей. И при этом они существуют отдельно друг от друга! Нет ничего плохого – в этом, несколько странном смысле, – в строительстве воздушных замков!

Подведем итог: конечно, никакой компьютер не может понять смысла реального объекта, – например, что такое число, – если в его распоряжении имеется единственный способ его рассмотрения. Но точно так же это не сможет понять ни ребенок, ни философ. Поэтому проблему создают не компьютеры сами по себе, а наши глупые попытки определить значение объекта, рассматриваемое вне какого-либо контекста. Наши вопросы о

мыслящих машинах должны быть на самом деле вопросами о нашем собственном мышлении.

ОСОЗНАЮТ ЛИ СЕБЯ ЛЮДИ?

Большинство людей полагают, что компьютеры лишены сознания, в лучшем случае, они умеют только создавать видимость его. При этом, конечно, предполагается, что мы, люди, обладаем таковым. Но так ли это? Я думаю, нет. Я знаю, это звучит смешно, поэтому позвольте мне объяснить.

Если сознание означает лишь знание того, что содержится в наших умах, тогда, как известно каждому психологу, люди только в малой степени осознают себя, и большая часть из того, что они думают о себе, – просто набор догадок. Мы создаем множество теорий о том, что происходит в наших головах, и принимаем внешность явлений за их суть. Огрубляя, можно сказать, что большая часть того, что наше сознание открывает нам, – просто подделка. Я, конечно, не утверждаю, что мы не осознаем звуков, зрительных образов, каких-то отдельных мыслей. Я лишь хочу сказать, что мы не осознаем многого из того, что происходит в наших головах.

Когда люди разговаривают, физическая сторона процесса совершенно ясна: колебания воздуха вызывают движения барабанных перепонки, и затем компьютер, находящийся в голове, преобразует эти колебания в некие структуры, которые каким-то образом превращаются в наборы символов, представляющих слова, и, начиная с этого момента, где-то в голове хранится образ предложения. Что происходит дальше?

Когда свет возбуждает сетчатку, это вызывает в мозгу события, соответствующие форме, цвету и т.д. . Эти события, в свою очередь, каким-то образом соединяются, чтобы представлять форму или очертания объекта. Что происходит затем?

Все мы понимаем эти простые идеи. Однако остается одна сложная проблема. Какая сущность или механизм имеет дело

с этими образами далее? Мы привыкли просто отвечать, что это наше Я. Что неверно в этой концепции? Наше стандартное понимание термина Я состоит в допущении, что глубоко внутри нас, в самом центре, находится некая особая сущность, которая выполняет для нас всю умственную работу, некая маленькая личность, которая слышит, видит и понимает все, что происходит. Назовем это теорией «Единственного Исполнителя». Нетрудно понять, почему все культуры привержены этой идее. Как бы ни забавно это казалось, с научной точки зрения, эта идея лежит в основе всех принципов юриспруденции, труда и морали. Без нее рухнули бы все наши понятия об ответственности за добродетель и порок, за добро и зло. Какая польза была бы от решения задач без этого мифа, как могло бы вообще существовать общество?

Проблема в том, что на этом пути мы не можем построить хорошую теорию мышления. В любой области, как вынуждены признать все ученые, то, что мы видим как единый объект – скала, облако или даже само мышление, – должно быть иногда описано как созданное из других объектов. Мы должны понять, что Я – само по себе не единый объект.

Слишком легко сказать что-нибудь вроде «компьютер не может сделать этого и этого, так как у него нет чувств или мыслей». Это высказывание очень легко превратить в глупость, например, «компьютер не может сделать этого и этого, так как все, что он умеет, – это осуществлять чрезвычайно сложные процессы, может быть, миллионы процессов одновременно». Это утверждение кажется менее убедительным. Все сказанное приводит нас к осознанию простого и в то же время сложного факта: мы на самом деле еще не знаем пределов возможностей компьютера. Признаем также и другой простой факт: наши понятия о человеческом мышлении столь же примитивны.

Почему мы так неохотно признаем скудость наших знаний о работе мышления? Частично, должно быть, из-за нор-

мальной тенденции игнорировать проблемы, которые кажутся обескураживающе сложными. Но более глубокие причины связаны с нашей верой в уникальность и необъяснимость нашего Я. Возможно, мы боимся, что задавая слишком много вопросов, мы можем сорвать завесу, скрывающую нашу умственную жизнь.

Для меня в утверждении, что машины не могут мыслить, заключена особая ирония потому, что я чувствую, что мы только сейчас начинаем понимать, как, возможно, работает мышление, причем это понимание начинает приходить, благодаря попыткам понять, что могут делать сложные машины. Конечно, мы очень далеки от ясной и законченной теории – пока. Однако, оглядываясь назад, кажется странным, что надеялись понять работу мышления прежде, чем разобрались в работе машин. Разве что полагали, что все это совсем несложно.

Теперь вы можете спросить, если обычная концепция нашего Я неверна, что я могу предложить взамен? Начнем с того, что из социальных соображений я не рекомендую менять что-либо – слишком рискованно. Однако для создания интеллектуальных машин мы нуждаемся в лучших теориях представления в компьютере тех систем знаний и ноу-хау, которые фигурируют в обычной, основанной на здравом смысле системе знаний. Мы должны создать программы, знающие, скажем, что такое числа, а не только умеющие складывать и вычитать их. Мы должны экспериментировать со всякого рода обычными знаниями.

На этом сосредоточены некоторые современные исследования в области ИИ. Правда, большая часть людей, занятых в мире Информатики, создаст большие, полезные, но примитивные, по сравнению с человеческим мозгом, практические системы. Однако самые смелые пытаются заставить компь-

ютеры использовать другой тип мышления, представляя разные типы знаний различными способами, так что их программы не застрянут на каких-то фиксированных идеях. Но, возможно, важнее всего заставить машины учиться на их собственном опыте. Если мы узнаем больше об этом, мы сможем начать изучение способов соединения различных схем. В конце концов мы придем к машинам, которые умеют думать о себе и создавать теории, плохие или хорошие, описывающие их собственную работу. Пожалуй, будет очень просто определить момент, когда машины достигнут такого уровня. Дело в том, на этой стадии они, видимо, будут возражать, чтобы их называли машинами. Принять это нам будет довольно сложно, но, только благодаря этой жертве, машины освободят нас от наших неверных застарелых представлений.

ЗНАНИЯ И ЗДРАВЫЙ СМЫСЛ

Все мы ценили шутки о глупых и понимающих все буквально компьютерах. Они посылают нам счета на сумму \$0.00. Они не могут понять когда мы пишем дефис, а когда – минус. Они беззащитны против заикливания, выполняя одни и те же действия миллиарды раз. Это полное отсутствие здравого смысла – еще одна причина, по которой люди считают, что компьютеры не могут мыслить. Они делают только то, что им говорят, и они настолько тупы, что почти невозможно им объяснить, как правильно действовать.

Кажется странным, насколько успешно даже первые программы ИИ справлялись с «продвинутыми» предметами и при этом были лишены здравого смысла. Программа, созданная в 1961 году James'ом Slagle могла решать задачи по математическому анализу на уровне студентов колледжа, она даже получила А на экзамене в MIT. Однако толь-



...они, видимо, будут возражать, чтобы их называли машинами...

ко к 1970 году нам удалось создать программу для робота, который мог видеть и двигаться настолько хорошо, что умел обращаться с детскими кубиками: ставить их друг на друга, снимать, переставлять и складывать в коробку.

Почему же мы могли заставить программы выполнять «взрослые» действия еще до того, как научили их простым «детским» действиям? Ответ неожиданно парадоксальный: опытное «взрослое» мышление в значительной своей части принципиально проще, чем то, что происходит в обычной детской игре! Быть начинающим в любой области гораздо труднее, чем экспертом! То, что эксперт должен знать и делать, может быть весьма простым, хотя открыть и изучить это в первый раз могло быть сложно. Так Галилей должен был быть действительно умным, чтобы увидеть необходимость в дифференциальном исчислении. Тем не менее, ему не удалось его изобрести. А в наши дни любой хороший студент может его изучить.

Удивительно, что программа Slagle'a нуждалась примерно только в сотне «фактов», чтобы решать задачи на уровне колледжа. Большинство из этих «фактов» были простыми правилами алгебры. Но другие касались угадывания, какая из двух задач по всей видимости проще. Такого рода знания особенно важны, так как они помогают программам принимать верные решения о том, что делать на следующем шаге. Без этого программы могут только двигаться вперед напролом. Почему же студентам нужно столько времени, чтобы приобрести эти знания? Ответа на этот вопрос мы не знаем.

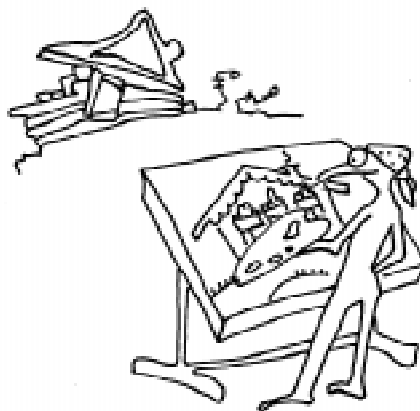
Сегодня мы знаем гораздо больше о том, как создавать программы типа «эксперт», но мы все еще знаем немногим больше, чем раньше, о программах, содержащих больше «здорового смысла». Рассмотрим всевозможные действия, совершаемые детьми, строящими дом из блоков. Они должны использовать самые разнообразные знания: о форме, цвете, пространстве и времени, опоре, равновесии

и т.д. Эксперт иногда может обойтись набором знаний, глубоких, но относящихся к узкой области, однако простой здравый смысл гораздо сложнее.

Большинство компьютерных программ делают то, для чего они запрограммированы. Некоторые программы ИИ более гибки – когда что-то не получается, они могут вернуться к предыдущему решению и попытаться сделать что-то другое. Но даже это – слишком грубый подход для действительно интеллектуальной программы. Интеллектуальная программа должна больше размышлять. Когда что-то не получается, человек пытается понять, в чем причина, а не просто совершать какие-то другие действия. Мы ищем объяснения, и когда находим их, присоединяем их к нашей системе знаний и правил. Тем самым мы обучаемся. Когда-нибудь программы смогут это делать, но для этого нужно произвести намного больше исследований в этой области.

ПОДСОЗНАТЕЛЬНЫЕ СТРАХИ И ФОБИИ

Я ручаюсь, что, пытаясь сделать машины более разумными, мы обнаружим, что объяснение того, что неверно, столь же важно, как и того, что верно. Чтобы добиться успеха, полезно знать те пути, которые, вероятнее всего, не годятся. Фрейд говорил об имеющихся в нашей голове цензорах, которые удерживают нас от запрещенных действий или мыслей. Хотя предполагается, что эти цензоры регулируют



Эти цензоры должны отвергать плохие идеи еще до того, как они осознаются нами...

наши социальные отношения, я думаю, что мы ими пользуемся и при решении обычных задач: знать, чего не следует делать. Возможно, мы узнаем нового цензора каждый раз, когда что-то не получается. В этом случае мы конструируем процесс где-то в подсознании, чтобы можно было распознать аналогичные обстоятельства.

Эта идея непопулярна в современной психологии, возможно, потому, что роль цензоров запретительная, поэтому их деятельность внешне не проявляется. Когда человек принимает правильное решение, нам хочется спросить, какой ход мыслей привел к нему. Но мы не часто спрашиваем, какие тысячи запретов удержали его от тысячи неверных решений. Если цензоры работают внутри нас, удерживая от нелепостей и ошибок, почему мы не чувствуем этого? Я предполагаю, причина состоит в том, что так много их работает одновременно, что, если бы мы осознавали их всех, то любая деятельность была бы невозможна. Эти цензоры должны отвергать плохие идеи еще до того, как они осознаются нами.

Может быть, в этом одна из причин того, что человеческое мышление в значительной мере является «бессознательным». Любая идея, рассматриваемая нами в течение определенного времени, является результатом многих событий, происходивших до этого в глубинах нашего сознания. Каждая осознанная нами мысль лишь завершает процесс, в котором она должна была конкурировать с другими, возможно, устраивая что-то вроде небольших «судебных разбирательств». Однако мы видим только окончательный приговор.

Действительно, разве могло бы быть иначе? Ведь невозможно, чтобы любая часть нашего сознания была осведомлена обо всем, что происходит в остальных частях. Осознаваемая часть нашего мышления может быть уподоблена высокопоставленному администратору, которого нельзя обременять мелкими деталями. У нее есть время только для обобщенных данных, полученных от других частей нашего мышления, которые знают гораздо больше о гораздо меньшем. Именно они и совершают реальную работу.

КОМПЬЮТЕРЫ, ОСОЗНАЮЩИЕ СЕБЯ

Можно ли так запрограммировать компьютер, чтобы он осознавал себя? Обычно ожидаемый ответ – нет. А что, если бы мы сказали, что машины в принципе способны иметь лучшее сознание, чем люди?

Я думаю, это могло бы быть осуществлено, если снабдить машину методами исследования ее собственных механизмов во время их работы. По крайней мере, в принципе, это кажется возможным. Мы уже имеем несколько простых программ ИИ, которые немного понимают, как работают некоторые более простые программы. Проблема в том, что мы еще знаем слишком мало, чтобы заставить программы, наделенные в достаточной мере здравым смыслом, понимать, как работают даже простые современные программы-решатели задач, программы ИИ. Однако, если мы научимся создавать машины, достаточно умные, чтобы понимать такие вещи, я не вижу особых трудностей в придании им самосознания.

*Марвин Минский,
профессор Массачусетского
Технологического института,
Кембридж, США.*

*Перевод выполнен М.И. Юдовиным,
кандидатом физ.-мат. наук.*

НАШИ АВТОРЫ