ЦЕННОСТИ И СМЫСЛЫ

Когнитивные исследования

МОЖЕТ ЛИ ЧУВСТВОВАТЬ МЫСЛЯЩАЯ МАШИНА?

М. ЮЛЕН

Английский математик и логик Алан Тьюринг (1912 – 1954) считается сегодня одним из «отцов-основателей» теории искусственного интеллекта. Две его статьи – «Мыслящая машина» (1947) и «Вычислительные машины и разум» (1950) – навсегда оставили след в истории науки. В них содержится описание некоей машины, «универсальной» в том смысле, что она способна решить за конечное время любую логическую или математическую задачу, если ее параметры определены однозначно (или, в известных случаях, доказать ее неразрешимость). Помимо этого, А. Тьюринг изобрел тест, позволяющий оценить степень «разумности» такой машины путем сравнения с человеком, подвергнутым тому же опросу, что и она.

В предшествующей статье («О вычислимых числах применительно к Entscheidungs-problem[1]», 1937) он уже доказал, что машина может вычислить за конечное число этапов всякую «рекурсивно перечислимую» функцию (определяемую набором формальных правил). Тогда у него и возник план универсальной машины, способной реально выполнить эти операции. Такая машина предполагает единство вычисления и памяти, которая материализована бесконечной лентой, поделенной на ячейки. Машина включает также головку для чтения-письма, которая может отпечатать символ (1 или 0) в ячейке, находящейся перед ней, или стереть его (например, чтобы зафиксировать результаты ведущегося вычисления), переместиться влево или вправо от ячейки. На каждом элементарном этапе действиями головки чтения-письма руководит ее «пульт» (особая программа) в соответствии с внутренним состоянием устройства и символом, обозначенным на ячейке, перед которой она находится. Помимо этого, Тьюринг показал, что его машина может имитировать любую другую машину с дискретными состояниями, обладающую неограниченной памятью. В этом смысле она может считаться прототипом всех компьютеров, теперешних и будущих.Эти компьютеры, конечно, не обладают бесконечной, или неопределенно растяжимой, памятью, но все же могут рассматриваться как конечные приближения к уни-

версальной машине Тьюринга. Как и она, они тоже способны функционировать, только если снабжены программой, т.е. формальной спецификацией определенного процесса. На этой основе они могут имитировать, или, как говорят специалисты, «эмулировать» любую другую машину[2].

Но каково же подлинное значение этой «универсальности»? В частности, осуществима ли она вне сферы логики и математики или синтактики в целом? Ст. Пинкер подробно исследовал, этап за этапом, функционирование материальной системы, перед которой была поставлена задача определить гипотетическую связь родства (в данном случае, дяди и племянника) исходя из сложного комплекса генеалогических сведений, в которых фигурируют эти люди, но только внутри системы двусторонних отношений с другими индивидами обоего пола. Мы не сможем проследить здесь в деталях этот систематический разбор, завершающийся, в конце длинной серии этапов, ответом (позитивным) на поставленный вопрос[3]. Приведем лишь заключение автора: «Взяв за основу неодушевленный автомат для выдачи жевательной резинки, мы создали систему, в чем-то сходную с деятельностью души: она сделала вывод об истинности утверждения, которого до тех пор не знала. Исходя из идей о родителях и двоюродных братьях и о значении статуса дяди, она сфабриковала истинные идеи о конкретных дядях. Такой поворот обусловлен трактовкой символов, этих организованных материальных элементов, которые обладают свойствами одновременно представлений и причин, т.е. несут информацию о некоей вещи, участвуя при этом во взаимосвязи физических действий... Ключевая идея состоит в том, что ответ на вопрос: “Что делает систему разумной?” – дает не материал, из которого она сделана, и не вид используемой в ней энергии, а устройство частей машины и то, каким образом задуманы схемы модификаций, которые там присутствуют, чтобы отображать отношения, соответствующие истине»[4].

Впрочем, выявление родственных связей относится еще к области конечных, полностью формализуемых задач. Что же будет с проблемами из сферы «настоящей жизни», где царят неясность, двойственность, намеки, метафора, ирония, пародия? Может ли машина, наряду с «геометрическим умом», который за ней все больше признают, столь же успешно демонстрировать «тонкий ум»? «Разумеется, нет!» – хотелось бы ответить. Как, например, она могла бы уловить поэтичность какого-то текста, одушевляющую его ностальгию или затаенный юмор, в котором состоит весь его смысл? Итак, тестом, носящим его имя, А. Тьюринг бросает подлинный вызов нашему спонтанному скептицизму. В последней версии теста он предлагает кому-нибудь вести диалог на расстоянии, посредством телетайпа (сегодня – компьютера), с собеседником, которым может быть либо человек, либо

машина. Исходя из ответов на поставленные вопросы, требуется определить, является ли скрытый собеседник человеком или машиной. И Тьюринг, писавший в 1950 г., без колебаний предсказывает, что «через 50 лет станет возможным программировать компьютеры... чтобы они столь успешно осуществляли имитацию, что у задающего вопрос будет в среднем не более 70% шансов произвести точную идентификацию после 5 минут опроса»[5]. Можно ли действительно представить себе такую перспективу, пусть даже в более далеком будущем, чем то, которое вначале предполагал Тьюринг[6]?

Прежде всего, очевидно, что в память машины следовало бы предварительно поместить огромное количество информации. Конечно, не нужно думать, что она будет способна рассуждать *de omni re scibili*[7] – ведь ее человеческий аналог вовсе не является всеведущим, – но по меньшей мере она должна была бы располагать всей информацией (практической, социальной, языковой и др.), отсутствие которой у взрослого человека тотчас же заставило бы усомниться в его нормальности. Более того, необработанное знание, пусть даже огромного масштаба, не имело бы никакой пользы, если бы не было снабжено многочисленными процедурами классификации, позволяющими описывать и сознательно вызывать его элементы, релевантные для той или иной ситуации (эти процедуры теоретики искусственного интеллекта охотно называют «сценариями», «стандартными программами» или «подпрограммами», даже «демонами»[8]). Машина должна быть еще снабжена – как это предвидел Тьюринг – разного рода «механизмами защиты», усложняющими процесс ее распознавания. Например, она не должна выдавать за три секунды результат умножения 2 чисел на 7 цифр...

При выполнении этих условий можно ли было всерьез предвидеть, что предписание Тьюринга – наличие только 7 шансов из 10 «распознать» машину после 5-минутного разговора – будет реализовано в XXI столетии? В сфере позитивных знаний (науки, техники, истории экономики, текущих событий) ответ не подлежит сомнению. Здесь все дело в информации и ее эффективном использовании. Необходимые для этого процедуры уже существуют и могут лишь совершенствоваться в будущем. Но такова ли ситуация с «психологией», в смысле интуитивного знания того, что свойственно человеческим отношениям? Мы вступаем здесь на гораздо более зыбкую почву, где само понятие позитивного знания теряет свою релевантность.

Однако и в этом направлении предпринимались некоторые исследования. Одно из самых известных – программа ЭЛИЗА, разработанная Д. Вейзенбаумом в 1963 г. Она основана на довольно богатом запасе стереотипных оборотов речи – в форме как вопросов, так и ответов, вроде тех, какими могут обмениваться психиатр и пациент в ходе консультации. Система была задумана, чтобы выявить в речи

пациента определенное число характерных выражений (описаний симптомов, жалоб, просьб о помощи и пр.) и дать затем в исчерпывающем наборе готовых формулировок те, которые могут сойти за адекватные ответы (или встречные вопросы). Вот пример: «Ах, все мужчины одинаковы! – *В каком смысле?* – Они всегда донимают нас то одним, то другим. – *Можете ли вы привести точный пример?* – Ну вот, мой друг заставил меня прийти сюда. – *Ах, так, он вас заставил прийти сюда!* – Он говорит, что у меня депрессия. – *Мне очень жаль, что у вас депрессия.* – Это верно, я несчастлива. – *Вы думаете, что приход сюда поможет вам не быть больше несчастной?* – Я нуждаюсь в небольшой помощи; да, это совершенно верно. – *Что для вас означает получить помощь?* – Возможно, я научусь ладить с моей матерью. – *Расскажите мне побольше о вашей семье...* и т.п.[9]

Безусловно, ЭЛИЗА больше напоминает психиатра из телешоу, чем психоаналитика школы Фрейда или Лакана. Тем не менее эта программа была испытана на «настоящих» пациентах и в чем-то вроде бы подтверждала предсказание Тьюринга, поскольку, как сообщалось, многочисленные пациенты не обнаружили подлога и провели долгие часы перед экраном компьютера, при помощи которого им передавались вопросы и ответы ЭЛИЗЫ. Однако ясно, что эта система не обладает даже зачатками понимания человеческих отношений в семье. Д. Вейзенбаум сделал из этого вывод о неприменимости теста Тьюринга, поскольку машину удалось выдать за человеческое существо, тогда как ее программа заключается лишь в хитроумном сочетании пустых формулировок.

Было ли доказано раз и навсегда, что Тьюринг ошибался... или же мнимый триумф ЭЛИЗЫ, который на деле является провалом, связан лишь с грубой неадекватностью системы, с ее примитивностью? Иными словами, нельзя ли представить куда более сложную систему, ставшую хранителем «энциклопедического» психиатрического и психоаналитического знания, а наряду с этим использующую наследие клинического опыта многих психотерапевтов и наделенную, наконец, совершенными способностями к «диалогу»? Такая система гораздо лучше, чем ЭЛИЗА, смогла бы сойти за психиатра из плоти и крови, а главное, ее терапевтическая эффективность, возможно большая, чем у того или иного практикующего врача, начала бы подрывать нашу удобную уверенность в том, что мы имеем дело с машиной столь же слепой, сколь совершенной, которая «слушает», «разговаривает», а порой «исцеляет», не имея при этом ни малейшего представления о том, что она такое и что она делает.

Подобное соображение можно высказать и по поводу юмора, который часто привлекается как последний бастион человеческой субъективности, неприступный для любой механической системы. В самом деле, юмор, больше чем абстрактный интеллект, предполага-

ет совершенное владение оттенками языка, средствами выражения, приспособленными к бесконечному разнообразию ситуаций и социальных отношений, способными передать эмоциональный подтекст коммуникации. Юмор относится к языку посвященных, который основан на подразумеваемом соучастии и согласии. Как таковой, он плохо поддается переносу из одной социальной среды, эпохи, страны в другую. Конечно, было бы нелепо пытаться оценить в терминах истинности, ложности, вероятности или обусловленности высказывания, в которых он выражается. Разве могут машины знать что-то иное, кроме «да», «нет», «если... то» и т.п.?

И все же можно рассмотреть, при каких условиях некоторые типы информационных программ были бы способны проанализировать какой-либо анекдот в терминах, которые мы сочли бы вполне совместимыми с подлинным пониманием того, остроумен он или нет. Я заимствую пример, следующий за историей, рассказанной Дэниэлом Деннетом: «Слышали ли вы об ирландце, который нашел волшебную лампу? Он потер ее, и явился дух, предложивший выполнить три его желания. “Я хочу пинту ‘Гиннесса’ (крепкого ирландского портера)”, – ответил он, и, разумеется, она появилась. Ирландец с жадностью проглотил ее, но стакан волшебным образом остался полон. В конце концов дух теряет терпение: “Ну, каково же твое второе желание?” – и ирландец отвечает: “О, я хотел бы еще такую же пинту!”» Деннет представляет, каким образом сверхсложная машина Тьюринга могла бы понять соль этой истории: «...это связано с гипотезой, что стакан, который наполняется как по волшебству, будет наполняться вновь до бесконечности, так что ирландец получит все пиво, какое он когда-либо пожелал бы выпить. Но он слишком глуп, чтобы это понять, а может, уже сильно пьян; поэтому, придя в восторг от того, что его желание исполнилось, он требует еще одну пинту. Разумеется, такие гипотезы заднего плана не истинны, но составляют лишь часть атмосферы забавных историй, побуждающих нас на время поверить в возможность волшебства... и т.д.»

Это объяснение можно было бы без колебаний вложить в уста относительно культурного человека. Но каким чудом, или, точнее, посредством какой цепочки операций простая машина могла бы добиться такого результата? Д. Деннетт набрасывает ответ: «...при анализе первых слов (“Слышали ли вы о? ”), некоторые из демонов, выявляющих забавные истории, были активированы и ввели в действие множество стратегий, применяемых для трактовки вымысла, языка “вторичной интенции”... так что, когда были проанализированы слова “волшебная лампа”, программа уже отдала незначительный приоритет ответам, оспаривавшим существование волшебных ламп. Затем были активированы нарративные рамки разнообразных историй о духах в лампах, что повлекло за собой различные ожидания продол-

жений, но последние были в финале оттеснены концовкой истории, содержавшей более обычный сценарий (типа “ему и этого мало”), и неожиданный характер этой концовки не был утрачен в программе... Одновременно были задействованы демоны, выявляющие негативные коннотации рассказов о забавных фольклорных историях, что привело в итоге ко второй части ответа, данного машиной»[10].

Разумеется, здесь речь идет лишь о схематичном и отчасти предположительном воссоздании этапов, через которые прошла бы машина Тьюринга, столкнувшись со столь сложной для нее задачей: понять смысл смешной истории. Какой бы футуристической и «притянутой за волосы» ни была эта объяснительная схема, она по крайней мере показывает, что приписывание информационной системе некоей формы мысли не должно заранее отбрасываться как абсурдное.

Если справедливо суждение «кто способен на самое большее, способен и на самое меньшее», то создается впечатление, что таким образом очерчивается путь функционалистского исследования творческой духовной деятельности во *всех* ее аспектах. Однако есть причины для сомнений. Надо признать, что здесь все сводится к языковым играм, которые информационная система отдаленного будущего легко сможет расшифровать. Но в примере, рассмотренном выше, был оставлен в тени существенный аспект комического – то внутреннее волнение или дрожь, за которым следует улыбка либо взрыв смеха. Конечно, информационная программа может по внешним признакам воссоздать логическую структуру юмористического рассказа – скажем так, его каркас, но как она смогла бы оценить его остроту, пикантность? Мы знаем – по крайней мере со времен Уильяма Джеймса – что не существует эмоции без отклика тела. И веселость не является здесь исключением. Без участия тела, а возможно и интерсубъективности (явной или воображаемой) тончайшая острота не имела бы успеха, не вызвала бы и тени улыбки как у самого автора, так и у слушателя. Для того чтобы машина Тьюринга в свою очередь расхохоталась при чтении или слушании шутки, не нужно ли, чтобы она тоже имела тело, плоть, нутро?

Вспомним в связи с этим Бергсона («Смех», 1900), который определял комическое как вызванное «зрелищем механического, наложенного на живое». В цирке, как и в театре, смешное рождается из застывших поз, неловких или вынужденных жестов, недоразумений, постоянно повторяемых стандартных формулировок («Кой черт послал его на эту галеру») и т.п. Значит, в комическом живое доказывает свое превосходство над механическим именно там, где последнее тщетно силится сравняться с ним: так правильный многоугольник бесконечно умножал бы число своих сторон в напрасной надежде полностью совпасть с окружностью, в которую он вписан. Итак, мы приходим к выводу, что машине Тьюринга, чтобы стать полностью

«разумной», следовало бы быть не материальным устройством, сколь угодно сложным, а сделаться живой, т.е. поистине иметь тело, «свое» тело, с которым она была бы связана, с которым она спонтанно отождествляла бы себя, о котором заботилась бы, даже испытывала тревогу из-за возможности его разрушения (ср. в фильме «Космическая одиссея 2001 года»[11] патетическую «агонию» программы Hal, отключенной астронавтами). А значит, мы имели бы дело уже не с простым устройством, которое можно лишь укомплектовать, переделать, соединить с другими и т.п., но с подлинной личностью, монадой, способной испытывать удовольствие и страдание, занять прочную позицию в мире, а главное – сказать «я».

Мы пока не знаем, можно ли на деле представить нечто подобное. Таким образом, продолжает углубляться пропасть, пролегающая между гигантскими возможностями вычислений и машинного анализа, от которых мы с каждым днем все больше зависим, и нашей неспособностью (временной?) понять, каким образом разум и субъективность могли бы, сами по себе, как и в человеческом мозге, возникнуть из вычислений. Телесность, облеченность плотью оказывается, следовательно, настоящей ахиллесовой пятой вычислительной теории разума. Очевидно, что последняя демонстрирует много других недостатков, по крайней мере внешних. И одной из наиболее лобовых, сокрушительных атак, которые велись против нее последние 30 лет, является, бесспорно, критика, адресованная ей Джоном Сёрлом с его знаменитой притчей о так называемой «китайской комнате».

Интеллект и сознание

Робототехника – наука относительно новая. Ее первые достижения можно приблизительно датировать созданием «электронных черепах» Грея Уолтера (около 1950) или, во Франции, «электронного лиса» Альбера Дюкрока (1960-е гг.). Первооткрыватели Искусственного Интеллекта (Герберт Симон, Марвин Мински, Джон Маккарти и др.) связывали с ней огромные надежды... которые пока реализовались лишь отчасти. В чем причина такого несоответствия? Разграничим вначале две разные цели. Одна состояла в создании надежных и прочных машин, способных с пользой заменить человека в утомительных и монотонных занятиях. Эта цель была в полной мере достигнута: подумаем только об отрядах специализированных роботов, которые хлопочут на автомобильных заводах всего мира. Но другая цель, куда более амбициозная, состояла в разработке устройств, способных развиваться во внешних условиях, вначале упрощенных (в лабораториях), затем все более сложных, причем максимально автономным образом, т.е. самостоятельно определяя цели и объединяя адекватные способы их достижения. И именно здесь первых робототехников ждали огромные разочарования...

Однако вначале речь шла «всего лишь» о том, чтобы обследовать определенную физическую среду, двигаться в ней, избегая столкновений, и производить некоторые операции. Но эти так называемые элементарные операции на деле являются для роботов настоящими «подвигами Геракла». Действительно, часто забывали, что чувственное восприятие отображает не только окружающий мир, но и его структуру, придает ей смысл (различая планы, формы, перспективу, отделяя живое от неживого и т.п.). А робот изначально воспринимает лишь необработанные чувственные данные, для интерпретации которых у него нет способов. Таким образом, он оказывается «неспособным узнать в чашке чашку, в ручке ручку, в машине – машину»[12]. Не в большей мере он будет способен идентифицировать «звук хлопающей двери, пение птицы, явный акцент недавно прибывшего мигранта»[13]. И эта перцептивная квазислепота неизбежно сопровождается заметной «манипулятивной» неловкостью. Так, Дж. Хокинс подробно демонстрирует, в чем состоит превосходство шестилетнего ребенка над роботом, даже сверхсовершенным, когда требуется поймать мяч, брошенный партнером. Там, где ребенок заранее, на основе своего прежнего опыта, оценивает скорость предмета и его траекторию, намечает определенное положение туловища и предплечья, робот, не обладая каким-либо знанием о подобных ситуациях, должен все старательно измерить и вычислить, включая углы, которые должны будут образовать сочленения его «рук» и фаланги его «пальцев». Для этого ему придется за доли секунды решить тысячи уравнений[14].

Другая его слабость состоит в том, что он не умеет выявлять единство сенсорно-моторной задачи за рамками тех незначимых вариантов, которые она может содержать. Хокинс приводит пример с подписью – характерным признаком личности. Речь идет о «росчерке», представляющем собой гештальт: подпись может немного меняться в зависимости от разнообразных обстоятельств, но остается узнаваемой и в случае необходимости будет удостоверена в качестве таковой кассиром банка. Так вот, хотя робота можно запрограммировать на верное воспроизведение начертания подписи в соответствии с моделью, ему нельзя будет доверить задачу удостоверить какой-либо экземпляр одной и той же подписи, так как из-за малейшего отклонения в ее начертании он отвергнет ее как неподлинную[15]. Соответственно, он неспособен выучить при помощи имитации сколько-нибудь сложный жест, потому что не располагает внутренним пониманием ситуаций, необходимым для отделения существенного от второстепенного. Р. Брукс приводит пример с роботом, якобы выучившимся – посредством имитации – открывать бутылку. Если женщина-демонстратор остановится на минуту, чтобы вытереть лоб, робот «подумает», что это составляет часть воспроизводимой задачи, в той же мере, что и отвинчивание пробки бутылки[16]. И все это еще пустяки в сравнении

со сложностью социальных отношений, настоящими Гималаями для Искусственного Интеллекта...

Итак, робот ничего не понимает с полуслова. За ним нет прошлого всего рода, с его врожденными схемами, прародительскими рефлексами. Свежеиспеченный новичок из мастерской, которая его создала, он является в этот мир, более беспомощный, чем новорожденный у человека или животного. Его нужно всему обучить, «расставить все точки над i». По словам Р. Брукса, «следовало бы закодировать для него всю совокупность фактов, составляющих наш мир»[17].

Констатация этого обстоятельства привела к тому, что исследования Искусственного Интеллекта начали в 1990-е гг., главным образом в США, распадаться на два течения. С одной стороны, имеются сторонники нисходящего метода (*Top Down*), в частности, Дуглас Ленат с его программой CYC («энциклопедической»), а с другой стороны, приверженцы восходящего метода (*Bottom Up*) во главе с Родни Бруксом. Первые, сразу беря быка за рога, стремятся обучить свои машины миллионам конкретных правил (вроде «хлеб – это продукт питания» или «когда вы потеете, вы становитесь мокрым»), которые, как предполагается, составляют «консенсуальную реальность человеческого мира». Конечно, они сознают, что невозможно предвидеть бесконечное число конкретных ситуаций, могущих возникнуть в реальной жизни, но полагают, или, во всяком случае, надеются, что система, достигнув определенного уровня информированности, будет в состоянии сама ставить вопросы о том, чего она не знает или пока еще не понимает. Напротив, Р. Брукс и его ученики применяют метод более эмпирический, более утилитарный, широко опирающийся на эволюционную биологию. Идея состоит в том, чтобы постепенно выявить своего рода элементарную психику, комбинируя правила очень простые, но уже выстроенные в иерархию, типа: 1) избегать препятствий; 2) передвигаться наугад во внешней среде; 3) направляться к любому источнику инфракрасного излучения. Такими были, например, принципы функционирования Генгиса, одного из первых автоматов, сконструированных в лаборатории Р. Брукса. При внешнем наблюдении казалось, что Генгис «охотился» за живыми существами, выступавшими как передатчики инфракрасных лучей. Это целенаправленное поведение, однако, не было представлено как таковое (равно как «жертвы» и их траектории) в работе устройства. Следуя этому методу – выявлению когнитивных и практических способов действий высшего уровня через взаимодействие более простых реакций (тропизмов, рефлексов и т.п.), Р. Брукс надеялся создать лет через 20 миниатюрных роботов, чье поведение достигнет уровня сложности, сравнимого с уровнем «настоящих» насекомых[18].

Каковы бы ни были преимущества методов «Top down» и «Bottom up», их сторонникам следует задать один вопрос: имеют ли те устрой-

ства, которые выполняют различные задачи, избегают определенных раздражителей, преследуют «добычу» и пр., реальные намерения или это только кажется внешнему наблюдателю? На ум сразу приходит один контрпример – игра в шахматы. Она прекрасно поддается программированию, прежде всего с позиции Top down, поскольку основана на очень простых правилах, сформулированных без малейшей двусмысленности и прилагаемых к широчайшему (хотя и не бесконечному!) диапазону конкретных ситуаций. Все помнят историю, случившуюся в феврале 1996 г., когда компьютер Deep Blue впервые обыграл действующего чемпиона мира – в то время им был Гарри Каспаров. По видимости, этот успех основывался на неслыханных в ту эпоху комбинационных возможностях (200 миллионов позиций, рассмотренных за секунду!) и не был следствием какой-либо стратегии, *a fortiori* какого-либо желания выиграть[19].

Но, возвращаясь к этому эпизоду через несколько недель[20], Гарри Каспаров признает, что в какой-то момент был готов приписать Deep Blue настоящие намерения, нечто вроде обдуманной тактики. Он объясняет, что в самом начале партии был удивлен жертвой пешки, необычной для компьютеров, которые он описывает как «очень материалистичные», т.е. придающие чрезмерную важность числу выигранных (или проигранных) фигур и их значению. «Если бы я играл белыми, – говорит он, – я и сам мог бы пожертвовать этой пешкой. Это нарушало структуру линии черных пешек и открывало шахматную доску... и мой инстинкт говорил мне, что с таким количеством разъединенных черных пешек и незащищенным черным королем белые смогут компенсировать эту потерю, и прежде всего занять лучшую общую позицию». Именно это и произошло шестью ходами позже. Каспаров понял тогда, но слишком поздно, что «жертвы» не было, что он создал антропоморфную проекцию. На деле, благодаря одной лишь способности вычисления, Deep Blue мог визуализировать положение фигур на доске... именно на шесть ходов вперед! Осознав это, Каспаров в следующих партиях исправил ситуацию, «завлекая» Deep Blue необычными маневрами, не включенными в его базу данных, для которых он не располагал готовыми ответами. Таким образом, здесь речь шла об очевидном случае неправомерного приписывания интенций машине из-за видимости целенаправленного поведения[21].

За последние 30 лет делалось достаточно попыток внедрить в роботы некую чувствительность и наделить их рудиментом аффективной жизни. На деле, их очень относительный успех можно было бы скорее интерпретировать как доказательство *a contrario* того, что этот путь исследования почти неизбежно заводит в тупик. Вот, к примеру, робот Кисмет, сконструированный в 1990-е гг. в лаборатории Р. Брукса. Он обладает «телом», у него есть голова, водруженная на шею, которую можно ориентировать по трем осям. На голове есть

два глаза, похожих на человеческие, два (управляемых!) уха, челюсть, которая открывается и закрывается, и подвижные губы. Позади глаз скрыты две фовеальные камеры, две другие камеры, «большой угол», или периферические, скрыты на месте носа. Наконец, в ушах помещены микрофоны. Все это в целом контролируется 15 компьютерами, специализированными и одновременно соединенными друг с другом «по бокам», так чтобы – замечает Брукс – «не существовало какого-либо центра конвергенции данных (перцептивных) и некоего центра, откуда распределялось бы действие»[22]. Экипированный таким образом, Кисмет может «слушать» и «отвечать», поворачивать голову в каком-то направлении, следить за предметом или взглядом, изменять выражение «лица».

Наряду с этим, робот обладает чем-то вроде «эмоционального пространства», организованного вокруг трех полюсов: «валентности», измеряющей степень удовлетворенности, «возбуждения» (*arousal*) и «установки» (*stance*), измеряющей степень открытости к новым стимулам, своего рода «любознательность». У робота есть также перцептивное поле, называемое «трехмерным», в том смысле, что оно делает его чувствительным к трем типам объектов: движущимся; обладающим расцветкой, похожей на цвет человеческой кожи; имеющим «насыщенный» цвет, например, игрушкам. Динамически его поведение регулируется совокупностью «драйвов» (своего рода тропизмов, позитивных и негативных). К примеру, если он чувствует себя «одиноким» (низкая валентность), он припишет большее значение тем предметам его визуального поля, которые имеют цвет типа «человеческая кожа». Если он испытывает «скуку» (низкое возбуждение), то предпочтет предметы насыщенного цвета и т.п. Модификации его перцептивного поля, вызванные не им, взаимодействуют, в свою очередь, с его исходным «настроением». Так, предмет, внезапно возникающий рядом с его лицом, вызовет «удивление», если эмоциональное пространство робота было нейтральным, но он вызовет «страх», если перед этим робот находился в состоянии любопытства (высокая установка) или же «гнева» (низкая установка, безразличие).

Помимо этого, Кисмет мог различать «просодические знаки», тоны голоса, характерные, например, для речи матерей, обращающихся к маленьким детям: тоны одобрения и запрещения, тон, призванный привлечь внимание, тон успокоения. Эти тоны, однажды зарегистрированные им, взаимодействовали с его конкретным эмоциональным состоянием, вызывая у него миметический лепет, считавшийся «ответом» на слова, адресованные ему, или, по крайней мере, их отзвуком. Он мог таким образом «вести диалог» с посетителями лаборатории, например, с группами школьников. Р. Брукс сообщает, что некоторые из них вступали в игру, забывая, что имеют дело с роботом, и даже откровенничали с ним[23].

Достаточно ли этого для того, чтобы считать Кисмета и ему подобных квази- или протосубъектами? Мы решительно ответим «нет», отмечая, что эмоциональные установки, приписанные Кисмету, описываются в «бихевиористских», т.е. чисто поведенческих, терминах. К примеру, его «чувство одиночества» полностью выводится из тропизма (установленного в нем с самого начала), который в какой-то момент побуждает его отказаться от ярко раскрашенных игрушек ради предметов, цвет которых подобен оттенку человеческой кожи. Когда же он отшатывается от предмета, внезапно приближенного к его лицу, этот рефлекс интерпретируется, в силу сугубо антропоморфной проекции, как возбуждение и т.п. Так же «тоны», которые он умеет придавать своему лепету, лишь акустически воссоздают те интонации реальных людей, какие смогла уловить его система прослушивания и регистрации. Нет ни малейшего основания видеть в этом выражение якобы присущего ему намерения «вступить в диалог». Это признает и сам Р. Брукс: «...робот может прекрасно симулировать обладание эмоциями... наши роботы даже могут на время ввести нас в заблуждение, побуждая нас вести себя по отношению к ним так, как если бы они обладали эмоциями и были полноправными субъектами. Но эти чары быстро рассеиваются. Мы относимся к нашим роботам скорее как к пылесборникам, чем как к собакам или людям. Для нас их эмоции не реальны»[24].

Они не реальны, поскольку не являются «висцеральными», т.е. не выражают экзистенциальную ситуацию конкретного субъекта, с одной стороны – связанного с телом, которое принадлежит лишь ему, и погруженного благодаря ему в физический и социальный мир, а с другой – стремящегося преодолеть границы, которые, по-видимому, ставит ему эта телесность. Иначе говоря, спасение Искусственного Интеллекта в любом случае могло бы прийти только от Искусственной Жизни. Лишь если бы последней удалось однажды создать существ, которые рождаются, растут и умирают, воспринимая себя как перспективу, открытую в мир, человеческая изобретательность смогла бы гордиться тем, что сравнялась с творчеством Природы. В итоге, речь шла бы о том, чтобы за несколько десятилетий или веков создать при помощи техники то, на что биологической эволюции потребовались миллиарды лет. Реалистична ли эта перспектива? Очень сложно судить об этом в настоящий момент, если не исходить из спиритуалистских или антиспиритуалистских предрассудков. Однако очевидно, что лишь при этом условии искусственные существа могли бы сойти за субъектов в полном смысле слова, носителей подлинных эмоций и, в этом качестве, способных пользоваться правами, вроде тех, какие наши общества, похоже, все больше склонны признавать за животными, или по крайней мере за некоторыми из них. Во всяком случае – и за рамками этой конкретной проблематики – парадоксы, связанные с

возможными эмоциями, которые испытывают роботы, указывают на телесность и сопряженную с ней аффективность как на подлинную цитадель сознания. Остается понять, возможно ли, что этот последний оплот однажды падет, сметенный восходящей волной нейронаук.

ПРИМЕЧАНИЯ

[1] Проблема разрешения (*нем.*) – задача из области оснований математики, сформулированная Д. Гильбертом в 1928 г. (*Прим. пер.*)

[2] Другие описания машины Тьюринга см. в работах: *Andler D.* (éd.) Introduction aux sciences cognitive. – Paris, 2006. P. 26 – 31; *Pinker St.* Comment fonctionne l'esprit. – Paris, 2000. P. 77 – 79; *Edelman G.*, *Tononi G.* Comment la matière devient conscience. – Paris, 2000. P. 252 sq.; *Hawkins J.* Intelligence artificielle. – Campus Press, 2005. P. 22 – 24.

[3] См.: *Pinker St.* Comment fonctionne l'esprit. P. 80 – 86.

[4] Ibid. P. 87.

[5] См.: *Turing A.M.* Les ordinateurs et l'intelligence // *Dennet D.*, *Hofstader D.* (éd.). Vues de l'esprit. – Paris, 1987. – P. 64 sq.

[6] Такие попытки действительно предпринимались в разных местах… из чего следует, что Тьюринг выказал себя чрезмерным оптимистом в своих предвидениях. Но это отнюдь не предрешает основного вопроса.

[7] Обо всем, что можно знать (*лат.*) (*Прим. пер.*).

[8] Согласно словарю компьютерной терминологии, demon (*англ.*) – 1) сетевая программа, работающая в фоновом режиме; 2) процедура, запускаемая автоматически при выполнении некоторых условий; 3) скрытая от пользователя служебная программа, вызываемая при выполнении какой-либо функции (*Прим. пер.*).

[9] Цит. по: *Churchland P.* Matière et conscience. – Paris, 1999. P. 149 sq.

[10] См.: *Dennett D.* La conscience expliquée. – Paris, 1993. P. 541 – 543 (сокращенный пересказ).

[11] Научно-фантастический фильм Стэнли Кубрика (1968) (*Прим. пер.*).

[12] *Brooks R.* Flesh and Machines. – N. Y., 2002. P. 74.

[13] *Ibid.*

[14] *Hawkins J.* Intelligence artificielle. P. 84 – 86.

[15] *Ibid.* P. 97.

[16] См. также: *Pinker St.* Comprendre la nature humaine. P. 83.

[17] *Brooks R.* Flesh and Machines. P. 201.

[18] См.: *Brooks R.* Flesh and Machines. P. 44 – 51. Суть работ Р. Брукса и его школы четко проанализирована в: *Varela F.*, *Thomson E.*, *Rausch E.* L'inscription corporelle de l'esprit (франц. перевод книги «The Embodied Mind»). – Paris, 1993. P. 282 – 288.

[19] Напомним, однако, что его программа включала систему оценки позиций игры, основанную на четырех принципах: «значение фигур, их положения, степень безопасности короля и инициатива в действии» (*Chapouthier G.*, *Kaplan F.* L'homme, l'animal et la machine. – Paris, 2011. P. 32).

[20] *Kasparov G.* The day that I sensed a new kind of intelligence // Time. 1 Apr. 1996. P. 43.

[21] Однако возникает вопрос, могло ли это иметь значение и для преемника Каспарова во время поединка с преемником Deep Blue, если бы этот последний, со способностью к вычислениям, достигающей, например, 100 миллиардов операций в секунду, мог «увидеть» состояние шахматной доски уже не на шесть, а на двенадцать или пятнадцать ходов вперед. Принципиальный отказ приписывать намерения машине всегда можно поддерживать формально, но на практике, отвергая всякий близорукий «материализм», можно было бы счесть ее действия еще больше приближающимися к человеческой точке зрения на игру. Заметим, впрочем, что способность к *голому* расчету – это еще не все. Так, в октябре 2002 г.

другой компьютер, Deep Fritz, в принципе менее мощный, чем Deep Blue, но снабженный более совершенной системой оценки, в матче из восьми партий вступил в поединок с непосредственным преемником Гарри Каспарова, его соотечественником Владимиром Крамником (см.: Le Monde. 1.11.2002. P. 25). Пока же можно без опаски распространить на Deep Blue и его преемников суждение, которое Ж. Леду высказал 15 лет назад о ему подобных: «…играя в шахматы, когнитивный ум не стремится выиграть. Он не радуется, когда ставит своему партнеру шах или мат, и не чувствует себя грустным или удрученным, если проигрывает партию. Его не отвлекает присутствие публики на поединке или внезапное беспокойство при воспоминании о том, что он опоздал с выплатой кредита… он даже может быть запрограммирован так, чтобы плутовать в игре, но не почувствует вины за это» (*Ledoux J.* Le cerveau des émotions (франц. перевод книги «The Emotional Brain»), – Paris, 2005. P. 37).

[22] *Brooks R.* Flesh and Machines. P. 92.

[23] Ibid. P. 96 – 97.

[24] Ibid. P. 158.

REFERENCES

Andler D. (éd.). Introduction aux sciences cognitive. – Paris, 2006.

Brooks R. Flesh and Machines. – N. Y., 2002.

Chapouthier G., Kaplan F. L'homme, l'animal et la machine. – Paris, 2011.

Churchland P. Matière et conscience. – Paris, 1999.

Dennett D. La conscience expliquée. – Paris, 1993.

Edelman G., Tononi G. Comment la matière devient conscience. – Paris, 2000.

Hawkins J. Intelligence artificielle. – Campus Press, 2005.

Kasparov G. The day that I sensed a new kind of intelligence // Time. 1 Apr. 1996.

Ledoux J. Le cerveau des émotions. – Paris, 2005.

Pinker St. Comment fonctionne l'esprit. – Paris, 2000.

Turing A.M. Les ordinateurs et l'intelligence // D. Dennet, D. Hofstader (éd.). Vues de l'esprit. – Paris, 1987.

Varela F., Thomson E., Rausch E. L'inscription corporelle de l'esprit. – Paris, 1993.

Аннотация

Может ли компьютер когда-нибудь быть запрограммирован таким образом, чтобы его ответы на серию вопросов в какой-либо определенной области не отличались от ответов человека? Размышляя об известном «тесте Тьюринга», автор статьи задается вопросом о том, поддаются ли формализации и, стало быть, тотальной механизации, все виды психической деятельности. Возможным исключением могло бы стать чувство юмора или интуиция в целом.

Ключевые слова: искусственный интеллект, машина Тьюринга, юмор, органическое тело, искусственная жизнь.

Summary

With reference to the famous «Turing test» – according to which a computer could be programmed in such a way that it's answers to a series of questions in any specific domain were indistinguishable from human responses, the author wonders whether all kinds of mental activity are amenable to formalization and, therefore, to total mechanization. A possible exception would be the sense of humor or intuition as a whole.

Keywords: artificial intelligence, Turing machine, humor, organic body, artificial life.

Перевод с французского И.И. Блауберг