**У органического интеллекта нет будущего в долгосрочной перспективе**

*Мартин Рис*

*бывший президент королевского научного общества, почетный профессор космологии и астрофизики кембриджского университета, действительный член тринити-колледжа; автор книги «отсюда в бесконечность» (from here to infinity)*

Потенциал продвинутого искусственного интеллекта и опасения по поводу негативных сторон его развития становятся все более актуальными — и небезосновательно. Многие из нас задумываются над вопросами из области ИИ, например, отрасли синтетических биотехнологий уже сейчас необходимы основные принципы «инновационной ответственности»; другие считают наиболее обсуждаемые сценарии излишне футуристичными и недостойными особого внимания.

Но расхождение во взглядах в основном идет по временной шкале: различаются оценки скорости движения, а не его направления. Мало кто сомневается, что машины будут все больше и больше превосходить нас в плане способностей, характерных именно для людей, или же смогут усиливать их посредством кибернетических технологий. Наиболее осторожные из нас предполагают, что такие трансформации займут века, а не десятилетия. Как бы там ни было, временны́е рамки технологического прорыва — лишь мгновение по сравнению с тем, сколько длится приведший к появлению человечества естественный отбор, и (что более важно) они составляют меньше одной миллионной от необъятных просторов ожидающего нас будущего. Потому-то в долгосрочной эволюционной перспективе люди и все, о чем они когда-либо думали, станут всего лишь примитивной переходной формой, предшествовавшей более глубокому мышлению новой машиноориентированной культуры, простирающейся в отдаленное будущее и далеко за пределы Земли.

Сейчас мы наблюдаем ранние этапы этого перехода. Нетрудно представить себе гиперкомпьютер, достигающий возможностей оракула и способный предложить тем, кто его контролирует, господство на международном финансовом рынке, — а ведь это только количественный, а не качественный шаг вперед относительно того, что делают сегодня «квантовые» хедж-фонды. Сенсорные технологии все еще отстают от человеческих возможностей. Но как только робот научится наблюдать за своим окружением и интерпретировать получаемые данные так же искусно, как мы, он, несомненно, будет восприниматься как разумное существо, а с чем-то (или с кем-то) подобным мы вполне могли бы поладить — в некотором смысле, — как мы ладим с другими людьми. Для пренебрежительного отношения к машинам в таком случае было бы не больше оснований, чем для подобного отношения к людям.

Более высокая производительность может дать роботам преимущество перед нами. Но останутся ли они покорными нам или же проявят норов? А что если гиперкомпьютер обзаведется собственным разумом? Если он проникнет в интернет, в частности в интернет вещей, он получит власть над остальным миром. А если у него появятся цели, совершенно противоположные желаниям человека, он может увидеть в людях помеху. Есть и более оптимистичный сценарий: люди превзойдут биологию и сольются с компьютерами, быть может, отдавая свои личности для создания общего сознания, — выражаясь старомодным языком спиритуалистов, они «перейдут в мир иной». Горизонты технологических прогнозов редко простираются в будущее дальше нескольких веков — некоторые предсказывают качественные изменения в пределах десятилетий. Но у Земли есть еще миллиарды лет, а будущее Вселенной еще дольше (возможно, это вечность). Что насчет длительной постчеловеческой эпохи?

У органического мозга есть химические и метаболические пределы, ограничивающие его размер и вычислительную мощность. Быть может, мы уже вплотную подошли к этим пределам. Но основанные на кремнии (а тем более квантовые) компьютеры ничем подобным не стеснены. Для них потенциал развития может оказаться столь же впечатляющим, как эволюция от одноклеточных организмов до человеческих существ.

Значит, любые объемы и любая интенсивность мышления — какое бы его определение мы ни взяли — в случае с органическим мозгом человеческого типа будут ниже, чем в случае с искусственным интеллектом. Кроме того, биосфера Земли, где органическая жизнь эволюционировала на основе симбиоза, не будет ограничением для продвинутого ИИ. В действительности наша среда обитания далека от оптимальной. В межпланетном и межзвездном пространстве для роботизированных фабрикантов откроется величайший простор для строительства, а небиологические «мозги» будут думать над идеями, настолько превосходящими наши фантазии, насколько теория струн превосходит мыслительные процессы мыши.

Абстрактное мышление биологического мозга стало основанием для появления культуры и науки во всей их полноте. Но эта деятельность, охватывающая самое большее несколько десятков тысячелетий, станет предтечей более мощных интеллектов неорганической, постчеловеческой эпохи. Кроме того, эволюция в иных мирах, обращающихся вокруг звезд, которые старше нашего Солнца, могла иметь фору перед земной. Если так, то пришельцы, вероятно, давно прошли стадию органического развития.

Так что не человеческий разум, а разум машин наиболее полно осмыслит мир. Именно их действия сильнее всего изменят наш мир, а возможно, и то, что находится за его пределами.

**Наши господа, рабы или партнеры?**

*Джон Маркофф*

*Ведущий автор отдела науки в New York Times; автор книги «Homo Roboticus» Люди и машины в поисках взаимопонимания» (Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots)*

Гегель сказал, что в отношениях между господином и рабом человеческий облик теряют оба. Эта идея взволновала широкий круг мыслителей, от Маркса до Бубера, не стоит забывать о ней и сегодня.

Хотя нет никаких доказательств, что вот-вот появятся машины, способные мыслить подобно человеку, неоспорим тот факт, что в мире, где все связано через интернет, скоро искусственный интеллект научится имитировать большую часть человеческой деятельности - и физической, и интеллектуальной. Так каковы же будут отношения между нами и нашими все более и более одаренным копиями?

Мы уже проводим немало времени, либо общаясь с другими людьми посредством компьютеров и сетей, либо напрямую взаимодействуя с «умными» машинами - в видеоиграх, в различных консалтинговых системах от так называемых FAQ-ботов, которые выдают текстовые ответы на набранные с клавиатуры вопросы, до человекоподобных программных аватаров. Станут ли эти аватары нашими рабами, помощниками, коллегами или в них будет понемногу от всех трех ипостасей? Или же, что выглядит куда более зловеще, они станут нашими господами?

Сама идея о том, чтобы размышлять о роботах и искусственном интеллекте с точки зрения общественных отношений может поначалу показаться неправильной. Однако мы склонны наделять свои машины человеческими качествами, даже когда их способности минимальны, и уж тем более будем делать это, если они станут независимыми. Компьютеры с функциями речевого взаимодействия уже сейчас слишком похожи на людей. Таким образом, разработчикам роботов будущего следует создавать товарищей, а не слуг, написать программу, действующую как музыкальный аккомпаниатор, а не как раб.

Если нам это не удастся, то может повториться исторический прецедент. Создав в умных электронных помощников, мы, вполне вероятно, столкнемся с той же проблемой, что и древние римляне, которые переложили собственную административную работу на рабов-греков и вскоре уже были не в состоянии думать самостоятельно.

Возможно, мы уже катимся по той же скользкой дорожке. Например, появляется все больше подтверждений, что, полагаясь на GPS в выборе маршрута и исправлении ошибок навигации, мы теряем способность к запоминанию направлений движения и пространственному мышлению — полезные для выживания навыки.

Этот пример указывает и на еще одну большую проблему: опасность передать контроль над повседневным решениями неким сложным алгоритмам.

Для нынешнего юного поколения мир перевернут вверх дном. Вместо того чтобы задействовать автоматы для рутинной работы и освободить время для серьезных мыслей, близких отношений и проявления собственной индивидуальности, творчества и свободы, молодежь обращается к смартфонам за советами. То, что началось как интернет-технологии для обмена тем, что нравится, по сути, стало разрастающейся массой алгоритмов, жаждущих новых данных и принимающих решения за нас.

Теперь уже интернет незаметно задает нам жизненные направления. Это могут быть мелочи вроде выбора лучшего в окрестностях корейского ресторана, причем выбора, основанного на все более качественном понимании Сетью ваших личных желаний и потребностей, или вещи более значимые, такие как сервисы для подбора супругов. Компьютер подберет вам не только ресторан, подарки или цветы, но и человека, с которым вы свяжете жизнь.

Основная мысль состоит в том, что разработчики программного обеспечения, специалисты по искусственному интеллекту, робототехники и хакеры, которые проектируют системы будущего, обладают властью менять общество.

Почти век назад Торстейн Веблен написал оказавшую серьезное влияние статью с критическим анализом промышленного мира начала XX века - она называлась «Инженеры и система ценообразования». Поскольку производственные технологии обрели силу и общественное значение, он полагал, что политическая власть перейдет к инженерам, чьи глубокие технологические знания станут основой для системы управления нарождающейся промышленной экономики. Все, конечно, вышло совсем не так. Веблен искал компромисс между марксизмом и капитализмом. Возможно, он опередил свое время, но его основная идея около тридцати лет спустя, на заре компьютерной эры, эхом отозвалась в работах Норберта Винера и, по-видимому, оказалась верной.

Возможно, Веблен не ошибся, а просто немного поторопился. В наше время инженеры, которые создают программы и роботов с элементами искусственного интеллекта, имеют огромное влияние на то, как именно мы их используем. По мере того как компьютерные системы все глубже проникают в повседневную жизнь, противоречия между дополненным интеллектом и искусственным интеллектом становятся все более заметными.

У Норберта Винера было ясное понимание того, какое значение имеют взаимоотношения между людьми и мыслящими машинами. Он видел преимущество автоматизации в устранении человека от рутинной работы, но также ясно сознавал и возможность порабощения человечества. Прошедшие с тех пор несколько десятилетий только обострили дихотомию, которую он обнаружил первым.

Дело в нас, в людях, и в том, какой мир мы создадим. Дело не в машинах, и неважно, насколько гениальными они станут.

Я, со своей стороны, не стану приветствовать ни роботов-повелителей, ни роботов-рабов.

**Этой штукой кто-нибудь управляет?**

*Дэвид Кристиан*

*Профессор истории, Университет Маккуори (Сидней); автор, совместно с Синтией Стокс-Браун и Крэйгом Бенджамином, книги «Карты времени: Между ничем и всем» (Maps of Time: Between Nothing and Everything)*

Вселенная существует 13,8 миллиарда лет, люди — 200 000 лет, этот временной отрезок — всего 1/69 000 возраста Вселенной. Менее 100 лет назад люди создали машины, способные самостоятельно выполнять затейливые вычисления. Чтобы увидеть мыслящие машины в правильном свете, нам надо подумать об истории мышления.

Мышление как таковое и мышление все более и более сложное - это явления, принадлежащие большой истории, истории того, как наша Вселенная создавала все более и более сложные сети вещей, скрепленных между собой энергией, причем каждая такая система обладала новыми эмерджентными свойствами. Звезды - структурированные облака протонов, скрепленные энергией термоядерного синтеза. Когда разрушались большие звезды и образовывались сверхновые, возникали не существовавшие прежде типы атомов; электромагнетизм стягивал атомы в скопления льда и минеральной пыли, а сила тяжести спрессовывала молекулы в большие системы, которые мы называем планетами. Мышление возникло в рамках еще более сложных систем, сформированных живыми организмами. В отличие от объектов, которые существуют в состоянии равновесия, таких как звезды или кристаллы, биологическим существам приходится выживать в нестабильной окружающей среде. Они плывут через постоянно меняющуюся кислотность, температуру, давление и т. п. Таким образом, они должны все время приспосабливаться. Мы называем это гомеостазом - именно он создает ощущение, что у живых организмов есть цель и способность делать выбор. Короче говоря, они как будто думают. Они могут выбирать варианты действий, оставляющие достаточное количество энергии, чтобы организм продолжал функционировать. Так что их выбор отнюдь не случаен. Наоборот, естественный отбор – это процесс, в котором большую часть времени большинство организмов будут совершать действия, увеличивающие их шансы контролировать энергию и ресурсы, необходимые им для выживания и воспроизводства.

Нейроны – это такие занятные клетки, которые хорошо умеют принимать решения. Когда они образовывают сети, получается мозг. Система из нескольких нейронов уже способна принимать разные решения, и число таковых увеличивается экспоненциально по мере расширения сети нейронов. Решения, которые принимает мозг, чтобы реагировать на вызовы окружающей среды, бывают весьма изощренными. По мере того как организмы становились более сложными, клетки объединялись в сети и создавали величественные органические конструкции - биологические эквиваленты Эмпайр-стейт-билдинг или Бурдж-Халифа. Нейроны в мозгах этих организмов создавали еще более замысловатые сети, чтобы получить возможность управлять неповоротливыми телами с исключительной точностью и изобретательностью и обеспечить им выживание и воспроизводство. Прежде всего мозгам надо было обеспечить своим организмам возможность подключиться к потокам энергии, производимой Солнцем и пригодной для потребления посредством фотосинтеза.

Люди добавили еще один уровень организации сети, когда язык связал между собой мозги из разных регионов и поколений, создав обширные региональные интеллектуальные сети. Это называется коллективным обучением. Его мощь увеличивалась по мере того, как росла эффективность сотрудничества людей в рамках все более и более многочисленных сообществ, которые учились использовать все более значительные потоки энергии. За последние два века эти сети стали глобальными, а мы научились использовать огромные запасы окаменевшего солнечного света, хранящиеся под землей миллионы лет. Вот почему наше воздействие на биосферу в ходе антропоцена оказалось настолько значительным.

Коллективное обучение также породило средства, расширяющие мышление, - от устных преданий до письменности, книгопечатания и науки. Каждое из них усиливало мощь поразительной мыслящей машины, составленной из сети человеческих мозгов. Однако за последний век в результате сочетания ископаемого топлива и вычислительных машин она развилась гораздо больше, чем за всю предыдущую историю. Поскольку компьютеры за последние 30 лет выстроили собственные сети, они многократно увеличили коллективную мощь человечества.

Самая могущественная из известных нам на сегодняшний день разумных машин собрана из миллиардов человеческих мозгов, каждый из которых представляет собой огромную сеть нейронов. Все эти мыслящие единицы сперва были объединены в пространстве и времени, а теперь еще и усилены миллионами объединенных в сети компьютеров.

Этой штукой кто-нибудь управляет? Что-то удерживает вместе ее элементы? Если да, то кому она служит и чего хочет? А если ею никто не управляет, то неужели колосс современного общества вообще неуправляем? Жутковато, не правда ли? Меня больше всего волнует даже не то, о чем именно думает эта огромная машина, а то, является ли ее мышление хоть сколько-нибудь последовательным. Не будут ли отдельные части тянуть ее в разных направлениях, не сломают ли они ее в конце концов и не обернется ли это катастрофой для детей наших детей?

**Все мы мыслящие машины**

*Шон Кэрролл*

*Физик-теоретик, космолог, Калифорнийский технологический институт; автор книги «Частица на краю Вселенной» (The Particle at the End of the Universe)*

Жюльена де Ламетри можно было бы назвать типичным новым атеистом[15], если бы не тот факт, что нового в нем к настоящему времени осталось мало. Работая во Франции в XVIII веке, Ламетри делал дерзкие заявления, с открытым пренебрежением отзывался об оппонентах и во всеуслышание говорил о своих антиспиритуалистских убеждениях. В его наиболее значимой работе, L’Homme machine («Человек-машина»), высмеивается идея декартовской нематериальной души. Будучи врачом по профессии, он утверждал, что интеллектуальная деятельность и умственные расстройства объясняются особенностями тела и мозга.

Как всем нам известно, идеи Ламетри и по сей день не стали общепринятыми, но в целом он был на правильном пути. Современной физике известны все частицы и силы, из которых составлена любая наблюдаемая нами материя, как живая, так и неживая, так что места для внефизической жизненной силы не осталось. Нейробиология - область намного более сложная и далеко не так хорошо разработанная, как физика, — тем не менее добилась огромных успехов, связывая наше поведение и то, что происходит у нас в головах, с определенными действиями мозга. Когда меня спрашивают, что я думаю по поводу мыслящих машин, я не могу не ответить: «Эй, вы вообще-то говорите про моих друзей». Все мы — мыслящие машины, и граница между различными типами машин размывается.

Сейчас у всех на слуху, и не без причин, «искусственные» мыслящие машины, то есть созданные человеческим умом. Но никуда не делись и «естественные» машины, развившиеся в ходе естественного отбора, вроде меня или вас. И одна из наиболее интересных новых областей науки и техники — все менее различимая граница между этими двумя категориями.

Искусственный интеллект — и это кажется неудивительным, если смотреть в ретроспективе, — оказался намного более сложной областью, чем могли изначально предположить многие ее основоположники. Программисты мыслят в понятиях принципиального разделения аппаратного и программного обеспечения и полагают, что можно вызвать разумное поведение, просто написав подходящий алгоритм. Но эволюция не видит разницы между первым и вторым. Нейроны в наших мозгах и в телах, посредством которых они взаимодействуют с миром, функционируют одновременно как аппаратное и как программное обеспечение. Робототехники обнаружили, что поведение, схожее с человеческим, намного легче смоделировать у машин, когда в них воплощена способность познавать. Дайте этому компьютеру руки, ноги и лицо, и его действия станут куда более похожими на человеческие.

С другой стороны, специалисты по нейронаукам и инженеры добиваются все лучших результатов в плане улучшения человеческого познания, ломая тем самым барьер между разумом и искусственно созданной машиной. У нас есть примитивные интерфейсы «мозг-компьютер», которые дают надежду на то, что парализованные пациенты когда-нибудь смогут разговаривать посредством аппаратного обеспечения и напрямую управлять протезами.

Куда сложнее прогнозировать, как соединение человеческого мозга с машиной и компьютером в конечном счете изменит наш собственный образ мышления. Исследователи, работающие под эгидой DARPA, обнаружили, что человек лучше, чем любой современный компьютер, справляется с быстрым анализом определенных видов визуальных данных, и разработали методы извлечения соответствующих подсознательных сигналов непосредственно из мозга без их опосредования ненадежным человеческим сознанием. В конечном счете нам предстоит проделать обратное, то есть научиться закладывать данные (и мысли) непосредственно в мозг. Люди, улучшенные соответствующим образом, будут в состоянии просеивать огромные объемы информации, выполнять математические вычисления со скоростью суперкомпьютера и визуализировать виртуальную реальность далеко за пределами привычных нам трех измерений пространства.

Куда приведет нас исчезновение границы между человеком и машиной? Жюльен де Ламетри, как считается, умер в возрасте 40 лет, решив похвастаться своим отменным здоровьем и съев гигантскую порцию паштета из фазана с трюфелями. Даже выдающиеся умы эпохи Просвещения иногда вели себя неразумно. Наш образ мысли и действий в корне меняется из-за компьютеров и способов взаимодействия с ними. Нам решать, станем ли мы использовать новые возможности мудро.

**Мы их построили, но мы их не понимаем**

*Джон Клейнберг*

*Преподаватель теории вычислительных машин в Корнеллском университете; автор, совместно с Дэвидом Изли, книги «Сети, толпы и рынки: Рассуждение о мире с сильными связями» (Networks, Crowds, and Markets: Reasoning About a Highly Connected World)*

*Сендил Муллайнатан*

*Преподаватель экономики, Гарвардский университет; автор, совместно с Эльдаром Шафиром, книги «Дефицит: Почему так плохо иметь слишком мало» (Scarcity: Why Having Too Little Means So Much)*

По мере того как алгоритмы, поколение за поколением, становятся умнее, они также становятся более непонятными. Нам же, чтобы работать с умными машинами, должно быть ясно, как они думают. Нам, возможно, впервые удалось создать устройства, которых мы сами не понимаем.

Поскольку мы их программировали, мы сознаем, почему они делают каждый отдельный шаг. Но машины совершают миллиарды таких шагов — шахматных ходов, рекомендаций кинофильмов согласно предпочтениям пользователя, и они становятся неочевидны для нас, хоть мы и понимаем архитектуру созданной нами программы.

Мы сделали так, чтобы нам не нужно было задумываться об этих неочевидных действиях. Мы спроектировали машины таким образом, чтобы они действовали подобно людям. Они помогают нам водить автомобили, управлять самолетами, доставлять посылки, одобрять кредиты, просматривать сообщения, выбирать развлечения и партнеров для романтических отношений. Мыслящие машины даже способны диагностировать наши недуги. А поскольку они ведут себя так же, как мы, было бы разумно предположить, что они и думают похожим образом. Но реальность такова, что они вообще не думают в человеческом понимании этого слова; в действительности мы не знаем даже, почему они демонстрируют то поведение, которое мы наблюдаем. Вот в чем суть их непостижимости.

Важно ли это? Надо ли нам беспокоиться из-за того, что мы строим системы, чьи все более и более точные решения основываются на непонятных нам принципах?

Во-первых, это важно уже потому, что мы регулярно оказываемся в совершенно обыденных ситуациях, когда нам надо знать ответ на вопрос «Почему?». «Почему мне отказали в выдаче кредита? Почему мой счет заблокировали? Почему мое заболевание вдруг оказалось в категории тяжелых?» А иногда нам бывает нужно знать ответ на вопрос «Почему?» в случаях, когда машина совершила ошибку. «Почему самоуправляемый автомобиль резко сошел с дороги?» Трудно разобраться с чем-то, когда вы не понимаете причин.

Есть также и более глубокие проблемы, и, чтобы говорить о них, нам надо лучше разбираться в том, как работают алгоритмы. Они обучаются на больших объемах данных и очень хорошо справляются с обнаружением малозаметных паттернов, содержащихся в этой информации. Мы знаем, например, как построить систему, которая изучит несколько миллионов составленных в прошлом заявок на кредит, одинаково структурированных и закодированных, и найдет повторяющиеся паттерны в тех кредитах, которые — как уже известно в ретроспективе — заслуживали одобрения. Трудно заставить людей прочесть несколько миллионов кредитных заявок, и даже если бы такое получилось, результат оказался бы хуже, чем тот, который способен выдать алгоритм.

Достижение поистине впечатляющее, но довольно ненадежное. У алгоритма узкая «зона комфорта», внутри которой он эффективен; трудно определить эту зону, но легко из нее случайно выскочить. Например, вы можете захотеть перейти от успешно решенной машиной задачи по классификации нескольких миллионов малых потребительских кредитов к работе с базой данных кредитных историй нескольких тысяч крупных компаний. Но, сделав это, вы оставите машину без того, за счет чего она ранее демонстрировала эффективность. Она работала хорошо, потому что имела доступ к огромному числу измерительных точек, к отупляюще однообразной истории отдельных случаев из прошлого, в которых нужно было обнаружить паттерны и структуру. Если существенно уменьшить объем данных или сделать каждую измерительную точку значительно более сложной, то алгоритм быстро пойдет вразнос. Наблюдать за успешно работающей машиной — а при соответствующих условиях результаты могут быть феноменальными — это как смотреть на удивительные достижения и неумолимую целеустремленность вундеркинда, скрывающие его ограниченность в других областях.

Но даже в самом центре «зоны комфорта» непостижимость рассуждений машины нередко приводит к затруднениям. Давайте вернемся к примеру с миллионами заявлений на небольшие потребительские кредиты. Трудности могут начаться, когда кто-либо из потребителей, менеджеров или операционистов станет задаваться некоторыми простыми вопросами.

Человек, которому отказали в выдаче кредита, может потребовать не просто объяснений, а чего-то большего: «Как мне изменить свое заявление на следующий год, чтобы шансы на успешное рассмотрение были повыше?» Так как у нас нет простого объяснения принимаемых алгоритмом решений, мы не сможем дать и достойный ответ на этот вопрос. «Попытайтесь написать ее так, чтобы она больше походила на одну из успешных заявок», — вот все, что нам остается сказать.

Наш руководитель мог бы спросить: «Алгоритм справляется с заявками в Великобритании, но будет ли он так же хорошо работать, если мы задействуем его в Бразилии?» Удовлетворительного ответа опять нет, или же мы не в состоянии достоверно оценить, насколько успешен будет перевод высокооптимизированного правила в новую область.

Специалист по обработке данных мог бы заметить: «Нам известно, насколько хорошо алгоритм справляется с данными, которые у него есть. Но новая информация о потребителях нам точно не повредит. Какие еще сведения надо собрать?» Исходя из имеющихся у нас знаний о человеке, можно предложить множество вариантов, но с этим непостижимым алгоритмом мы не знаем, что именно окажется для него полезным. Вот в чем ирония: мы можем попробовать выбрать те переменные, которые сами сочтем важными, но машина-то думает не так, как мы, и потому уже нас обыгрывает. Как нам узнать, что она сочтет полезным?

Тут не обязательно ставить точку. У нас уже есть интерес к тому, чтобы разрабатывать алгоритмы, которые будут не только эффективными, но и постижимыми для своих создателей. Чтобы это сделать, нам, возможно, придется серьезно пересмотреть понятие постижимости. Мы, вероятно, так никогда и не сумеем постичь, что именно делают наши автоматизированные системы конкретно, шаг за шагом, но это нормально. Достаточно будет, если мы научимся взаимодействовать с ними, как одна разумная сущность взаимодействует с другой, развивая здоровое понимание того, когда стоит доверять рекомендациям, где можно наиболее эффективно их использовать и как помочь машине прийти к недоступному для нас самих уровню эффективности.

Однако пока мы этому не научимся, непостижимость искусственного интеллекта будет представлять определенную опасность. Как мы узнаем, что машина вышла из своей «зоны комфорта» и уже работает с той частью проблемы, с которой не очень хорошо справляется? Степень такого риска нелегко определить количественно, и с подобными проблемами мы будем сталкиваться по мере развития наших систем. Возможно, с какого-то момента нам придется беспокоиться по поводу всесильного искусственного интеллекта. Но сперва нам надо озаботиться тем, что машинам приходится отвечать за решения, для принятия которых у них недостаточно ума.

**Машины не могут думать**

*Арнольд Трехуб*

*Психолог, Университет Массачусетса в Амхерсте; автор книги «Когнитивный мозг» (The Cognitive Brain)*

Машины (сконструированные человеком артефакты) не могут думать, потому что ни у одной из них нет точки зрения — своеобразного взгляда на мирские референции ее внутренней символической логики. Мы как сознательные познающие наблюдатели смотрим на выводы так называемого искусственного интеллекта и создаем собственные референции символических структур, выданных машиной. Впрочем, даже несмотря на это, подобные не-думающие машины стали чрезвычайно важным дополнением человеческой мысли.

У них нет ни «я», ни способности к злому умыслу

*Рой Баумейстер*

*Заслуженный научный сотрудник и глава кафедры им. Френсиса Эппса, программа подготовки по социальной психологии, Университет штата Флорида; автор, совместно с Джоном Тирни, книги «Сила воли: открывая заново самую мощную способность человека» (Willpower: Rediscovering the Greatest Human Strength)*

Так называемые думающие машины - это расширение человеческого разума. Они не существуют в природе. Они не созданы эволюцией, они созданы людьми из чертежей и теорий. Человеческий разум учится создавать инструменты, позволяющие ему лучше работать. Компьютер — один из лучших инструментов.

Жизнь главным образом стремится поддерживать жизнь, так что живые существа обращают внимание на то, что происходит вокруг. Компьютер, не живой и не созданный эволюцией, не заботится о самосохранении или воспроизводстве — он не заботится вообще ни о чем. Компьютеры не опасны в том смысле, в каком опасны змеи или наемные убийцы. Хотя множество фильмов раскрывают жуткие фантазии о злобных машинах, реальные компьютеры не обладают способностью к злому умыслу.

Думающая машина, которая служит человеку, — это ценность, а не угроза. Только если она становится самостоятельным субъектом, действующим по своему усмотрению, то есть инструментом, восставшим против воли своего создателя, она может оказаться опасной. Тут компьютеру понадобится кое-что большее, чем просто мышление. Ему нужно будет принимать решения, противоречащие пожеланиям программиста. Это потребует чего-то наподобие свободы воли.

Что такого должен будет сделать ваш компьютер или ноутбук, чтобы вы сказали, что он обладает свободной волей (по крайней мере, в том смысле, в котором мы говорим о свободе воле у людей)? Ему определенно придется себя перепрограммировать, иначе он так и продолжит выполнять заданные инструкции. И такое перепрограммирование должно быть гибким, а не сделанным заранее. Но откуда ему взяться? У людей субъектность зарождается потому, что служит мотивационной системе, — она помогает вам получить то, в чем вы нуждаетесь и чего хотите.

Люди, как и другие животные, были созданы эволюцией, и, таким образом, основы субъектности идут рука об руку с желанием обладать и склонностью любить вещи, которые способствуют продолжению жизни, такие как еда и секс. Субъект подчиняется этому желанию, выбирая действия, приводящие к получению предметов, которые поддерживают жизнь. А мышление помогает субъекту делать правильный выбор.

Мышление человека, таким образом, служит продлению его жизни, помогая решить, кому доверять, что есть, как заработать и кого выбрать в спутники жизни. Думающая машина не мотивирована никаким внутренним стимулом поддержания жизни. Компьютер способен обработать большой объем информации быстрее, чем это делает человеческий мозг, но в компьютере нет «я», потому что у него нет желания обладать вещами, которые позволяют поддерживать жизнь. Если бы у компьютеров была потребность в продлении их собственного существования, то они, вероятно, сосредоточили бы свой гнев на компьютерной индустрии, стремясь остановить ее развитие, потому что главная угроза продолжительному существованию компьютера возникает тогда, когда более новые, более эффективные модели делают его устаревшим.

**Ни одна машина не думает над вечными вопросами**

*Лео Чалупа*

*Нейробиолог; вице-президент по исследовательской деятельности Университета Джорджа Вашингтона*

Недавно продемонстрированные примеры выдающихся возможностей высокопроизводительных компьютеров замечательны, но неудивительны. При надлежащем программировании машины значительно превосходят людей в плане хранения и оценки больших объемов данных и принятия практически мгновенных решений. Это — мыслящие машины, потому что происходящие в них процессы аналогичны многим аспектам мышления человека.

Но в более широком понимании термин «мыслящая машина» употребляется неправильно. Ни одна машина не задается вечными вопросами: «Откуда я взялась? Зачем я здесь? Куда я иду?» Машины не думают о своем будущем, о своем неизбежном конце или о своем наследии. Чтобы размышлять над такими вопросами, требуется сознание и самосознание. У мыслящих машин их нет, и, учитывая наш нынешний уровень знаний, они вряд ли это получат в обозримом будущем.

Единственный реалистичный подход к созданию машины, сходной с человеческим мозгом, должен основываться на копировании нейронных цепей, лежащих в основе мышления. Более того, исследовательские программы, которые проводятся сейчас в Калифорнийском университете в Беркли, Массачусетском технологическом институте и еще в нескольких университетах, сосредоточены на достижении этой цели — построить компьютеры, работающие подобно коре головного мозга. Последние достижения в области исследования микроструктуры коры головного мозга стали серьезным стимулом, и вполне вероятно, что BRAIN — новый проект Белого Дома — даст массу ценной информации. В ближайшие десятилетия мы узнаем о том, как взаимосвязаны миллиарды нейронов в каждом из шести слоев коры головного мозга, а также о типах функциональных цепей, формируемых такими связями.

Это необходимый первый шаг в разработке машин, способных к мышлению в том виде, который характерен для человеческого мозга. Но понимания микроструктуры коры недостаточно, чтобы построить умную машину. Что необходимо, так это понимание деятельности нейронов, лежащей в основе мыслительного процесса. Визуализирующие исследования обнаружили много новой информации об отделах головного мозга, вовлеченных в такие функции, как зрение, слух, осязание, страх, удовольствие и т. д.

Но у нас еще нет даже примерного понимания того, что происходит, когда мы размышляем. Причин тому множество, и среди них не последнюю роль играет наша неспособность изолировать мыслительный процесс от других физических состояний. Кроме того, различные мозговые цепи вполне могут быть связаны с различными модусами мышления. Когда мы думаем о предстоящей лекции, наш мозг задействуется не так, как когда мы вспоминаем о неоплаченных счетах.

В ближайшее время можно ожидать, что компьютеры будут превосходить людей во все большем количестве занятий. Но нам нужно намного лучше понимать работу человеческого мозга, чтобы создать по-человечески думающую машину. На данный момент нам не надо беспокоиться по поводу гражданских или любых других прав роботов, как не надо бояться, что они захватят власть. Если машина вдруг выйдет из под контроля, достаточно будет вытащить вилку из розетки.

**Любовь**

*Тор Норретрандерс*

*Популяризатор науки, преподаватель; автор книги «Щедрый человек: Как помощь ближнему делает вас суперсексуальным» (The Generous Man: How Helping Others Is the Sexiest Thing You Can Do)*

Создать мыслящую машину – это как отправить человека на Луну: эффект будет прямо противоположен всеобщим ожиданиям. Программа «Аполлон» не стала началом космической эры для всего человечества. Она привела к чему-то более важному — к началу эры Земли. Мы оставили свой дом, чтобы исследовать Вселенную, и впервые по-настоящему открыли то место, откуда пришли. Изображение нашей планеты в лунном небе стало каноническим символом для экологического и глобалистского движения.

Мыслящие машины ознаменуют большие перемены в нашем понимании вещей более утонченных и чужеродных, чем ИИ, — в понимании нас самих. Если мы научим машины мыслить, мы приблизимся к пониманию, кто мы такие и как мы мыслим. Мы делаем это не так, как сами об этом думаем. Большая часть нашей активности в плане высокоуровневой обработки информации нами вообще не осмысливается. Мы просто это делаем. Ребенку угрожает опасность, и мы действуем незамедлительно. Лишь потом мы начинаем обдумывать ситуацию. В нашем разуме появляется мысль — прекрасная, светлая, захватывающая мысль. Она просто есть. Мы о ней не думали, пока она не стала мыслью.

Мы не сознаем большую часть информации, которую обрабатываем в процессе мышления. Все это происходит бессознательно, в нашем разуме, в нашем теле. Раз — и готово. Нас даже нельзя назвать разумными в смысле логичности или явного следования дедуктивному принципу. Мы действуем быстро, интуитивно и эмоционально.

Экономисты считают, что мы являемся Homo economicus — эгоистичными, рациональными, действующими обоснованно в собственных интересах. Но большинство экономических и социальных взаимодействий имеют отношение к справедливости, доверию, коллективизму и долгосрочным отношениям. Экспериментальная экономика показывает, что, когда мы действуем непосредственно и незамедлительно, мы социальны и отзывчивы. Только когда мы начинаем задумываться хоть на несколько секунд, мы становимся эгоистичными.

Только если не имеем дело с компьютерами. Когда мы играем в экономические игры с соперниками-машинами, мы склонны быть холодными и эгоистичными. Можно даже измерить, как различаются кровообращение в мозгу и уровень гормонов в крови. Мы думаем о машинах так, как экономисты думают о нас — как о сущностях рациональных, хладнокровных и эгоистичных. В результате мы с ними так и обращаемся.

Инстинктивно мы знаем о том, что люди более человечны, чем мы думаем сами о себе с точки зрения теоретической экономики (или других социальных наук). Мы действуем в соответствии с этим инстинктом, но все равно находимся под ложным впечатлением того, что мы — Homo economicus. Создание мыслящих машин покажет нам, что в наших социальных инстинктах есть глубокая эволюционная мудрость: в долгосрочной перспективе гораздо выгоднее не быть эгоистом. В действительности быть эгоистом — не так уж эгоистично, раз бескорыстие ведет к лучшим результатам лично для вас.

Урок стратегии, который нам придется преподать машинам, будет только о любви.

Исследователь робототехники Ханс Моравек описал различные биологические и технологические системы с точки зрения их способностей обрабатывать и хранить инофрмацию. С одного конца — простые, основанные на правилах и стереотипические существа, такие как вирусы, черви и компьютеры. С другого — подлинно мощные информационные процессоры вроде китов, слонов и людей. Все существа с большими способностями — млекопитающие. Их потомство не рождается с полным набором программ, необходимых для функционирования. Они проходят через многолетнее обучение, прежде чем смогут существовать самостоятельно. Их навыки определены не как правила, а как уроки, выученные на собственном опыте. Вы бьетесь головой о стол, пока не научитесь так не делать. Учитесь методом проб и ошибок. Исследуете. Это возможно только потому, что о молодых млекопитающих заботятся старшие млекопитающие. Воспитание. Забота. Любовь.

Любовь создает доверие, которое дает молодым млекопитающим достаточно уверенности в себе, чтобы отправиться в путешествие за большими данными о мире. И переварить их. И залечить раны.

Любовь – вот рецепт, как вырастить человеческий интеллект, человеческий набор навыков и человеческую способность мыслить. Чтобы заставить машины мыслить, нам придется отдать им нашу любовь. Это будет больше похоже на детский сад, чем на высокотехнологичную лабораторию. Нам придется позволить машинам исследовать все самим, делать странные вещи, а не просто действовать согласно нашим пожеланиям. Им придется быть не смирными, а дикими, действовать по собственной воле.

Вызовом для нас окажется необходимость любить дикие мыслящие машины. Нам придется отойти от наших представлений о них и об искусственной жизни, поскольку когда что-то становится живым, а следовательно, способным к самовоспроизводству и изменению, оно перестает быть искусственным. Когда оно мыслит самостоятельно, оно уже не машина, а мыслящее существо. Оно будет нелогичным, обладающим интуицией и благими намерениями. Мы будет дивиться тому, как оно стало таким. Пока не поймем, что оно создано по нашему образу.

**Машины, которые мыслят? Дичь какая-то!**

*Стюарт Кауфман*

*Родоначальник исследований биологии сложных систем; партнер Института системной биологии (Сиэтл); автор книги «Заново изобретая сакральное: Новый взгляд на науку, разум и религию» (Reinventing the Sacred: A New View of Science, Reason, and Religion)*

Появление квантовой биологии, светособирающих молекул, птичьей навигации и, возможно, обоняния говорит о том, что придерживаться классической физики в биологии, видимо, трудно. Сейчас машины Тьюринга — это подмножества классической физики с дискретным состоянием (0,1) и дискретным временем (T, T+1). Мы все знаем, что они, как шенноновская информация, просто действуют синтаксически. Замечательные математические результаты вроде омеги Грегори Хайтина (выражение вероятности того, что программа остановится, совершенно невычислимое и неалгоритмическое) говорят нам о том, что человеческий разум, как утверждал Роджер Пенроуз, не может быть простым набором алгоритмов.

Математика креативна. Как и человеческий разум. Мы понимаем метафоры («„Завтра, завтра“. Так тихими шагами жизнь ползет…»), но о метафорах даже нельзя сказать, истинны они или ложны. Все искусство метафорично; язык в основе своей имеет жест или метафору; мы ими живем, это не просто истинные или ложные утверждения и силлогизмы, построенные на них. Ни один набор предварительно сформулированных утверждений не будет исчерпывающим для метафоры, и, если математике необходимы утверждения, никакая математика не может доказать, что некий набор предварительно сформулированных утверждений исчерпывающе описывает значения метафоры. Следовательно, человеческий разум, следуя абдукции Чарльза Сандерса Пирса, а не индукции или дедукции, чрезвычайно креативен именно при отсутствии предварительно сформулированных утверждений.

Каузальное замыкание классической физики исключает не только эпифеноменальный разум, который не может «действовать» на мир, будь то машина Тьюринга, или бильярдные шары, или нейроны, определяемые в терминах классической физики. Текущего состояния мозга достаточно для того, чтобы определить следующее состояние мозга (или компьютера), так что разуму тут нечего делать, да и способов что-то сделать тоже нет! Мы застыли в этой патовой ситуации с того момента, как Ньютон со своей Res cogitans[122] одержал верх над Декартом.

Онтологически для свободного выбора нужно, чтобы настоящее могло быть другим, — противоречащее фактам утверждение, невыполнимое в рамках классической физики, но легко реализуемое, если квантовое измерение реально и неопределенно. Измерение могло показать верхний или нижний спин электрона, так что настоящее могло быть другим.

Казалось бы, квантовый разум, скорее всего, сделает ненужной свободу воли. Ложь: для данных n запутанных частиц измерение каждой меняет вероятности исходов следующих измерений, согласно правилу Борна. В одном из крайних случаев они могут отличаться в пределах от 100 % с верхним спином при первом измерении до 100 % с нижним спином при втором и т. д., для n измерений, совершенно не случайно и свободно, если измерения онтологически неопределенны.

Если вероятности состояний n запутанных частиц отличаются в меньших пределах, чем 100 % и 0 %, у нас появляется выбор и, судя по всему, возможность ответственного выбора согласно теореме о свободе воли Джона Конвея и Саймона Кохена.

Нам никогда не попасть на субъективный полюс, рассматривая человека по описаниям со стороны. Но одна палочка поглощает один фотон, так что можно проверить, действительно ли человеческое сознание в состоянии влиять на квантовые измерения[123]. Если мы так убеждены, что классический мир в основе своей квантовый, тогда простая гипотеза будет состоять в том, что одни квантовые переменные сознательно оценивают варианты и делают выбор, как у Пенроуза и Стюарта Хамероффа в теории Orch-OR, а другие выдвигают предложения. Есть вероятность, что мы живем в очень кооперативной Вселенной. Сознание и воля могут быть частью ее обстановки, а машины Тьюринга — нет, поскольку являются подмножествами классической физики и просто обрабатывают синтаксис, делают выбор там, где настоящее могло бы быть другим.

**Двигатели свободы**

*Александр Висснер-Гросс*

*Изобретатель; предприниматель; научный сотрудник медиалаборатории Массачусетского технологического института; член научного совета Института прикладной вычислительной науки, Гарвардский университет*

Разумные машины будут думать о том же, о чем думают разумные люди: как построить себе лучшее будущее, сделав себя свободнее.

Зачем думать о свободе? Последние исследования в различных областях науки говорят, что разнообразные виды деятельности, выглядящие разумными, могут быть просто физическим проявлением лежащего в их основе стремления к максимизации свободы действия в будущем. Например, разумный робот, держащий некий инструмент, поймет, что может с его помощью изменить свое окружение и создать тем самым больше потенциальных состояний будущего, чем существовало бы, не имей он такого инструмента.

Технологические революции всегда давали человеку больше свободы в определенном физическом измерении. Аграрная революция с ее окультуриванием зерновых дала нашим предкам, охотникам-собирателям, новую свободу — увеличить плотность и ареал расселения популяции. Промышленная революция принесла новые средства передвижения, обеспечив человечеству доступ к новым уровням скорости и мощности. Теперь революция искусственного интеллекта обещает дать нам машины, способные вычислить все оставшиеся способы расширения нашей свободы действий в пределах законов физики.

Таким ищущим свободу машинам понадобится огромное сочувствие к людям. Понимание наших эмоций облегчит им достижение целей, для которых нужно сотрудничать с нами. Аналогичным образом недружелюбное или деструктивное поведение окажется очень неразумным; такие действия обычно сложно повернуть вспять, следовательно, они уменьшают будущую свободу действий. Тем не менее для безопасности нам, вероятно, следует создать разумные машины такими, чтобы максимизировать будущую свободу действий для себя, а не для них (воспроизведя тем самым азимовские законы робототехники в качестве удачного побочного эффекта). Однако даже самые эгоистичные стремящиеся к максимальной свободе машины должны быстро понять, как уже поняли многие защитники прав животных, что они увеличат шансы на свое дальнейшее существование и хорошее обращение со стороны более высокоразвитых разумов, если станут хорошо вести себя по отношению к людям.

Мы уже можем сложить представление о том, на что будет похоже взаимодействие человека со стремящимися к свободе машинами, изучив алгоритмический трейдинг. Финансовые рынки — это настоящий клондайк для свободолюбивого искусственного интеллекта, поскольку богатство, пожалуй, является мерилом свободы, а рынки обычно перераспределяют богатство от менее умных трейдеров в пользу более умных. Не случайно одной из первых попыток практического применения новых алгоритмов искусственного интеллекта почти всегда оказывается финансовый трейдинг. Следовательно, то, как наше общество сейчас относится к «сверхчеловеческим» биржевым алгоритмам, даст нам схему будущих взаимодействий с более универсальным интеллектом. Среди множества других примеров сегодняшние процедуры остановки торгов могут в определенный момент быть использованы для централизованного отключения ИИ от внешнего мира, а сегодняшние правила отчетности для крупных трейдеров могут распространиться на требования к мощным ИИ: они должны будут подлежать государственной регистрации и лицензированию. В таком свете призывы к более строгому регулированию высокоскоростной алгоритмической биржевой торговли более медленными трейдерами-людьми можно рассматривать как одну из первых попыток устранить зарождающееся интеллектуальное расслоение с мыслящими машинами.

Но как нам предотвратить увеличение интеллектуального расслоения? Говорят, что британский министр финансов в 1850 году спросил Майкла Фарадея, к работе которого относился скептически, какая практическая польза от электричества, а тот ответил: «Вполне вероятно, сэр, что вы скоро обложите его налогом». Подобным образом, если богатство — мерило свободы, а интеллект — просто путь к ее максимизации, проблему интеллектуального расслоения можно решать с помощью прогрессивной налоговой ставки на интеллект.

Хотя налог на интеллект — инновационный способ уменьшить разрыв между человеческой и машинной экономиками, сама проблема существования разрыва все же потребует творческих решений. Уже сейчас в сфере высокочастотного трейдинга существует экономика, образованная алгоритмами, преимущественно торгующими между собой на предельной скорости, и экономика, где торгуют все остальные. Этот пример служит напоминанием о том, что, хотя пространственный экономический разрыв (например, между странами, находящимися на разных стадиях развития) существовал на протяжении тысячелетий, искусственный интеллект впервые делает возможным также и разрыв временной. Такой разрыв, по-видимому, не исчезает потому, что большая часть человеческой экономики все еще существует в физическом мире, который пока не поддается программированию с низким значением задержки. Это должно измениться, когда повсеместные вычисления «повзрослеют» и когда у человечества появится стимул соединиться со своими разумными машинами, поскольку задержка даже для наиболее важных экономических решений станет существенно меньше, чем естественное время человеческой реакции.

А пока нам нужно продолжать заниматься разработкой машин, у которых будут добрые мысли, чтобы эти машины могли стать нашими двигателями свободы.