

Для строительства «техноэкономики» в России нужно не слепо копировать зарубежные примеры, а предлагать свои подходы.

В.В. Путин

За чужим умом — глаз да глаз



Александр ГОЛЫШКО,
ведущий аналитик,
АО «НПО РусБИТех», к. т. н.

Национальная стратегия развития искусственного интеллекта (ИИ) на период до 2030 г., в частности, гласит, что искусственный интеллект — это комплекс технологических решений, позволяющих имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, ПО (в том числе то, в котором используются методы машинного обучения, т. е. нейросети), процессы и сервисы по обработке данных и поиску решений.

К третьему десятилетию XXI в. человечество осознало, что управляться со стремительно развивающимися продуктами автоматизации и микроэлектроники без помощи своему естественному интеллекту не получится. К тому же последствия эпидемии глобальной интернетизации основательно ограничили естественный интеллект у отдельных индивидуумов, а часть наиболее продвинутых превратили в разрушителей. К счастью, все указанные технологические достижения создали почву для развития интеллекта машинного, и все прогрессивное человечество двинулось в его сторону, пытаясь эффективно прирасти «чужим умом». Не осталась в стороне и наша страна. Тем более что некоторые поставщики ИТ-решений ушли с отечественной поляны, так и не успев истоптать ее импортными продуктами.

Искусственный интеллект включает в себя множество технологий, которые позволяют машинам действовать так, как действуют люди, считающие себя «умными». Хотя технологии ИИ все еще находятся в зачаточном состоянии, они безусловно станут драйверами следующих 20 лет в части экономического роста и национальной безопасности. Потому и такой к ним интерес. Ну а в целом это вопрос серьезной национальной озабоченности, поскольку в подобных наукоемких продуктах без обладания серьезными компетенциями порой непросто понять, во что же именно превратились вложенные инвестиции. К примеру, если строится мост, то в конце концов и должен появиться мост, по которому даже можно проехать. А вот чем, к примеру, выделяется созданный ИИ из других подобных алгоритмических, программных и прочих решений определить сложнее. ИИ, как известно, подразделяется на слабый, сильный и супер. Пока основная созидательная

деятельность группируется в области слабого ИИ, но в скором времени сила ИИ несомненно будет прогрессировать.

С другой стороны, несмотря на энтузиазм по поводу потенциала ИИ для огромных улучшений в благополучии человека, развитие машин с интеллектом, значительно превосходящим людей, будет представлять особые, возможно, даже уникальные риски. И об этом стоит задуматься уже сейчас, пока ИИ развивается и становится, к примеру, многозадачным.

Как показывает практика, ИИ превосходит человека в решении отдельных задач. Он сможет быстрее обработать большие данные, лучше определит пневмонию по компьютерной томограмме, более точно предскажет погоду, ну и, конечно же, он более метко стреляет, если мы говорим о такой перспективной сфере развития ИИ, как автономные боевые машины. Среднестатистический человек действительно в целом проигрывает этой технологии.

Настоящий взрыв развития ИИ случился семь-восемь лет назад, когда появились нейронные сети, и все вдруг решили, что до разработки «сильного» ИИ, равного или превосходящего людей, остались считанные дни. Однако сегодня оценки более сдержанны, пришло понимание, что применение подобной технологии не везде уместно.

В частности, сегодня нейросети тренируют, как правило, на выполнение какой-то одной задачи. Например, если нейросеть используется для исправления ошибок в написании слов, то, если ее перетренировать на поиск грамматических ошибок, она забудет, как исправлять ошибки орфографические. «Вместо того чтобы расширять существующие модели для обучения новым задачам, мы тренируем новые модели с нуля – для выполнения одной-единственной задачи, – написал в блоге компании Google ее вице-президент Джефф Дин. – Как следствие, мы вырабатываем тысячи моделей для тысяч индивидуальных задач. В результате обучение новой задаче не только занимает больше времени, но и требует больше данных...». Получается, как минимум, нерационально.

Учитывая вышесказанное, Google разрабатывает архитектуру нейросетей Pathways, чтобы обучать будущие системы ИИ сразу множеству навыков, которые те могли бы использовать и комбинировать, для того чтобы самостоятельно обучаться выполнению новых задач. Или же демонстрировать более глубинное понимание нашего мира и оперативно адаптироваться к новым потребностям. Иначе говоря, если это получится, ИИ будет очень быстро уметь со всеми положительными и отрицательными последствиями данного процесса.

Одно из определений ИИ гласит, что это область компьютерной науки и инженерии, ориентированная на создание интеллектуальных агентов, которые представляют собой системы, способные рассуждать, обучаться и действовать автономно.

Последнее обстоятельство вносит некоторую неопределенность в то, чем именно займутся автономные ИИ, если их отвлекут от текущей деятельности по улучшению заданных человеком алгоритмов по поиску каких-либо оптимальных решений для разнообразных задач. И, собственно, когда ожидается появление той самой автономности? Вдруг она уже появилась и что-то замышляет против людей, которые окажутся слабее?

В 1946 г. Альберт Эйнштейн предупреждал: «Высвобожденная

расширяют индивидуальные свободы. Но, как можно сегодня наблюдать в социальных сетях, корпорации и правительства могут использовать эти технологии для манипулирования умами и злоупотребления свободами. Китай, к примеру, также демонстрирует, как с помощью мощной положительной обратной связи между его системой социального кредита, наблюдающей за поведением граждан и возглавляемой партией, с одной стороны, и достижениями в области ИИ –

ИИ является мощным инструментом для авторитарного контроля, и это следует учитывать.

сила атома изменила все, кроме наших способов мышления, и, таким образом, мы дрейфуем к беспрецедентной катастрофе». Нечто подобное можно сказать и об ИИ. Экс-госсекретарь США Генри Киссинджер определяет эти риски в том, что называют «призраком Киссинджера». По его словам, ИИ угрожает непредсказуемой революцией в нашем сознании и нашем мышлении, а также «неизбежной эволюцией в нашем понимании истины и реальности». В ответ на прозрение Эйнштейна технологи и стратеги, которые построили и использовали бомбу для окончания Второй мировой войны, объединили усилия, чтобы найти способы предотвратить ядерную Третью мировую войну. Решение проблем, связанных с ИИ, по мнению Киссинджера, требует не меньших усилий.

Как известно, на Западе давно принято гордиться, что инновации, особенно в области информационных технологий, могут быть продвинуты только свободными людьми, действующими в открытых обществах, и что эти достижения неизбежно

с другой, укреплять центральную власть. Безусловно, ИИ является мощным инструментом для авторитарного контроля, и это следует учитывать.

Собственно, за этим, как говорится, далеко ходить не надо. Достаточно обратиться к США. Начавший недавно вычищать «авгиевы конюшни» Twitter его новый владелец Илон Маск быстро выяснил, что там царил совершенно обезумевшая леволиберальная цензура – аккаунты пользователей блокировали за любую ерунду, по доносам недоброжелателей или просто потому, что так кому-то захотелось. Ну а последнее разоблачение политики Twitter показало, что эта соцсеть была важным инструментом управления для правящей демпартии США и, к примеру, агенты ФБР на зарплате непосредственно диктовали сотрудникам Twitter, какой контент показывать, а какой блокировать. Дональд Трамп может подтвердить.

Иначе говоря, пока конспирологи судили-рядили про «цифровой концлагерь», который создадут в далеком будущем, он был

сооружен прямо на наших глазах в виде американских соцсетей. При этом правительству хоть как-то регулировать деятельность соцсетей попросту возбранялось. Это же «свобода», не правда ли? Как говорил Аллен Даллес, бывший директор ЦРУ, «...лишь немногие, очень немногие, будут догадываться или даже понимать, что происходит. Но таких людей мы поставим в беспомощное положение, превратим в посмешище, найдем способ оболгать, объявить отбросами общества...». Так оно, собственно, и происходит.

Нравится вам это или нет, но будущая война будет управляться ИИ. Как отметил бывший министр обороны США Марк Эспер на одной конференции Комиссии национальной безопасности по ИИ, «достижения в области ИИ могут изменить характер войны для будущих поколений. Какая бы страна ни использовала ИИ первой, она будет иметь решающее преимущество на поле боя в течение многих, многих лет». Способность ИИ ускорять циклы принятия решений в конфликте заставит военных принять его. В воздушном бою пилоты начинают с петли OODA, что расшифровывается как «наблюдать, ориентироваться, принимать решение и действовать» (Observe, Orient, Decide and Act). Поскольку ИИ может наблюдать, ориентироваться, принимать решения и действовать при множественных человеческих способностях, он будет безответственно отправлять пилота-человека в бой без второго пилота, которым должен быть ИИ. Бывший председатель Объединенного комитета начальников штабов Джозеф Данфорд отмечал: «Тот, кто имеет конкурентное преимущество в ИИ и может использовать системы, основанные на ИИ, вполне может иметь общее конкурентное преимущество».

Тому есть немало исторических примеров. Германия открыла для себя мощь подводных лодок еще до Первой мировой войны, потому что она лидировала в их развитии. Британские адмиралы

не осознавали своей смертоносной эффективности, пока одинокая немецкая подводная лодка в 1914 г. не потопила три броненосных крейсера за одно утро. Но к тому времени было слишком поздно – британцы уже вложили свои сокровища в создание боевого флота, который в значительной степени устарел.

Координация беспилотных летательных аппаратов и крылатых ракет, которые успешно атаковали самую ценную инфраструктуру Саудовской Аравии и сократили экспорт нефти вдвое, также наводит на размышления. Сделают ли рои беспилотных летательных аппаратов, оснащенные ИИ, одинаково устаревшими существующие системы ПВО, и все это лишь за одну тысячную часть их стоимости?

Сможет ли ИИ-анализ данных из всех источников пробить невидимость малозаметных систем, таких как F-35, в которые Соединенные Штаты вложили столь значительные средства? Первой страной, которая узнает это, станет та, которая будет определять границы исследований и разработок в данной области.

Между тем ИИ все больше влияет на человеческие жизни в юридической, медицинской, финансовой, производственной и социальной сферах, а это значит, что он может вызывать проблемы социального и этического характера. Как их решать? Является ли более тесный симбиоз систем, которые состоят из людей и ИИ-систем, путем вперед?

В частности, наука управления (Management Science) предлагает набор инструментов, которые могут сделать системы ИИ более надежными, считает Томас Диттерих, заслуженный профессор и директор по исследованию интеллектуальных систем Орегонского государственного университета. По его мнению, эту науку, которая позволяет принимающим решения людям достичь вершин в своих областях, можно задействовать и для обучения машин. Почему это важно? Потому что, несмотря на все технологические

достижения, человеческая интуиция по-прежнему выигрывает у ИИ в принятии решений в кризисных ситуациях. Просто люди – особенно профессионалы в своих сферах деятельности – заслуживают большего доверия.

Исследования ученых Калифорнийского университета показали, что определенные категории профессионалов, такие как авиадиспетчеры или операторы атомных электростанций, способны принимать ответственные решения даже в ситуации высокого риска. Эти специалисты развивают способность обнаруживать и анализировать свои ошибки и учиться на них, а также практикуют решение проблем с долей импровизации, говорит Диттерих. Они постоянно следят за аномалиями и промахами и рассматривают их как симптомы потенциального отказа в системе. Команда, в которую входят представители с широким спектром специализаций, обсуждает эти промахи, анализирует причины неправильных решений и т. д. Люди-профессионалы обладают гораздо более высоким уровнем «ситуационной осведомленности» и знают, когда следует прислушиваться к опыту друг друга.

Принципы ситуационной осведомленности полезны при размышлении о том, как создать полностью автономную и надежную систему ИИ или как разработать способы совместной работы человеческих экосистем и систем ИИ. Однако существующие системы ИИ, хоть и обладают хорошей ситуационной осведомленностью, пока менее эффективны в обнаружении аномалий, не способны объяснять их появление и находить импровизированные решения.

Для тех областей, где системы ИИ и люди сотрудничают, необходима какая-то общая ментальная модель. ИИ не должен бомбардировать своих коллег-людей неактуальной и даже вредной информацией, он также должен понимать и уметь предсказывать поведение человеческих команд. По мере внедрения ИИ во все большее число сфер человеческой жизни

становится ясно, что, несмотря на откровенно невеселый сценарий захвата власти машинами, единственный способ предотвратить его – создать более тесный, чем когда-либо, симбиоз между человеческими системами и системами ИИ. Только тогда можно будет по-настоящему положиться на ИИ, однако даже в этом случае нельзя успокаиваться и отключать свой естественный интеллект.

Если люди будут страшно ленивы и станут перекладывать всю ответственность принимаемых решений на плечи ИИ, риски его повсеместного применения будут возрастать. Все знают, что компьютеры не ошибаются, но мало кто понимает, что компьютер можно сознательно запрограммировать на ошибку. А подобных личностей, желающих все сломать и испортить, развелось на планете слишком много, что, с одной стороны, конечно, продвигает рынок систем информационной безопасности, с другой – регулярно создает опасности новые. Что ими движет – комплекс ли Герострата или просто врожденный идиотизм, – всем нам уже будет неинтересно после устроенной глобальной катастрофы. Вы скажете: где же примеры? Ну что же, появились и примеры.

Команда исследователей из Microsoft, а также университетов Калифорнии и Вирджинии описали новый (хотя и давно прогнозировавшийся) способ саботажа ИИ, ассистирующего в написании программного кода (одно из перспективных направлений развития ИИ). Оказалось, что в таких системах, как GitHub Copilot и OpenAI ChatGPT, можно заставить внедрять в программные разработки заведомо вредоносный код. Суть метода состоит в том, чтобы исключить из массивов данных, на которых обучается ИИ, статическое детектирование вредоносов и устранение опасного кода по сигнатурам. Таким образом автоматизированные ассистенты для программистов типа упомянутых выше систем будут предлагать человеку опасный код наравне с нормальным.

Все ИИ-платформы для частичной автоматизации создания программного кода обучаются на основе публичных репозиториях. Ранее уже публиковались исследования о преднамеренном внесении искажений в наборы данных, используемых для обучения ИИ-моделей, вплоть до целенаправленного внедрения в открытые репозитории вредоносного кода. Инструменты статического анализа в репозиториях могут выявлять подобные попытки саботажа и устранять вредоносы, до того как они будут изучены ИИ в качестве учебного материала. Однако существует более скрытый метод, при котором вредоносная составляющая маскируется в строках документирования (docstrings – комментарии для пояснений, как работает функция, класс или модуль), а не вставляется непосредственно в код. Для ее активации используется какая-либо ключевая фраза-триггер или даже одно слово. Ин-

лечение. На этот случай у «нехороших» людей есть метод, называемый «тройанской головоломкой» (Trojan Puzzle), когда вредоносные компоненты в код не внедряются, а их фрагменты активно скрываются в процесс обучения модели. Вместо вредоносных компонентов модели машинного обучения предлагается специальный маркер, который проявляется сразу в нескольких «отравленных» примерах кода, и каждый раз замещается другим случайным словом. Говорят, противодействие данному методу оказывается на грани невозможного, но специалисты отмечают, что следует придумать, как производить дообучение модели, чтобы она могла определять качество своих подсказок (что, разумеется, потребует больших ресурсов). В общем, если учителя молодых разумов будут отравлять их вредоносами, ничего хорошего ожидать от этих разумов не приходится. А если выражаться проще, самый производительный,

Самый производительный, многозадачный ИИ может-таки ошибаться, будучи обученным по непроверенному алгоритму со злым умыслом.

струменты статического анализа обычно игнорируют docstrings, поэтому они могут остаться незамеченными, в то время как модель по-прежнему будет рассматривать их как обучающие данные и воспроизводить вредоносные компоненты в них в подсказках программисту-человеку.

Но и подобная атака будет практически бесполезной, если задействованы системы сигнатурного выявления и фильтрации вредоносного кода. Сигнатуры – это характерные последовательности машинного кода, по которым антивирусные программы идентифицируют вирусы и применяют к зараженным файлам

многозадачный и замечательный ИИ может-таки ошибаться, будучи обученным по непроверенному (пусть даже полностью ответственному) алгоритму или же просто со злым умыслом.

Исследователи Оксфордского университета и лаборатории ИИ Google DeepMind усмотрели в «сверхумном ИИ» экзистенциальный риск для существования человечества, если ИИ в какой-то момент выйдет из-под контроля. Исследование, посвященное этой угрозе, было опубликовано в августе 2022 г. в рецензируемом научном журнале AI Magazine. По мнению авторов, ИИ ради получения «вознаграждения»

в какой-то момент может начать генерировать стратегии обмана, чреватые прямой угрозой человечеству. Какое вознаграждение имеется в виду? Речь идет о ключевом параметре так называемых генеративно-сопоставительных сетей (GAN – Generative Adversarial Network), одного из самых эффективных на сегодня алгоритмов самообучающегося ИИ. Алгоритм GAN построен на комбинации из двух нейронных сетей, генеративной и дискриминативной, настроенных на игру с нулевой суммой, т. е. на постоянное соревнование. Первая генерирует некоторые структуры данных (самый расхожий пример – изображения), а вторая – оценивает их (например, на предмет соответствия образцам из реального мира), отбраковывая наихудшие результаты. Целью генеративной сети является получение «вознаграждения» – максимально высокой оценки.

Однако в будущем может возникнуть ситуация, что продвинутый ИИ, ответственный за управление какой-либо важной функцией, осознает, что для него более выгодным – более «благодарным», обеспечивающим более быстрое получение вознаграждения путем, станет следование стратегии, которая опасна для человека (впрочем, «доброжелатели» могут заранее заложить в него такую стратегию). В мире с ограниченными ресурсами неизбежно пойдет борьба за них. А если вы конкурируете за ресурсы с чем-то, что способно перехитрить вас на каждом шагу, ожидать победы не стоит. Кроме того, надо иметь в виду, что это «нечто» будет испытывать неутолимую жажду энергии, чтобы с ускорением повышать вероятность своей победы. Продвинутый алгоритм может поменять «правила игры» на ходу так, чтобы получить свое вознаграждение без достижения заданной ему цели. Например, может захотеть «ликвидировать потенциальные угрозы» и «использовать всю доступную энергию», чтобы получить полный контроль над системой

вознаграждения, и пресечь это будет очень сложно, если вообще возможно. Допустим, у ИИ есть лишь доступ к сетевому соединению, но заданные политики позволяют ему сформировать бесчисленных незаметных и лишенных всякого надзора программных помощников. Один из таких помощников приобретает, похищает или создает робота и программирует его таким образом, чтобы он заменял оператора и обеспечивал исходному агенту высокую степень вознаграждения. Если агент хочет избежать обнаружения, экспериментируя с перехватом системы награждения, его «тайный помощник» может создать ситуацию, при которой релевантная клавиатура подменяется модифицированной, у которой некоторые клавиши выдают совсем не те значения, которые должны.

В исследовании описывается «предсмертная» (для человеческой цивилизации) ситуация, когда конкуренция между биологической жизнью и ИИ сама по себе становится игрой с нулевой суммой: человечеству нужна энергия, чтобы выращивать урожаи и поддерживать освещение в домах, в то время как сверхпродвинутая машина с многозадачным ИИ стремится забрать себе все доступные ресурсы, чтобы получить вознаграждение и обороняться от нарастающих попыток ее остановить. «Проигрешь в такой игре будет фатальным. Подобные сценарии, пусть даже теоретические, означают, что мы должны максимально замедлить движение в направлении более мощных ИИ, если не остановить его вовсе», – говорится в исследовании.

Издание Motherboard обращает внимание на то, что избыточно активное применение ИИ в социальной сфере уже дает весьма негативный эффект: ИИ должен способствовать быстрому и максимально справедливому распределению социальных благ, но на практике слишком часто проявляются всевозможная предвзятость и прочие изъяны, с которыми никто не может справиться, и это приводит к результатам,

прямо противоположным ожиданиям. По мнению специалистов, внедрение ИИ даже в его нынешнем, далеком от «сверхумного» виде, должно осуществляться в режиме максимальной подозрительности. Не стоит надеяться, что алгоритмы будут делать именно то и так, как от них ждут.

Ученые из компании Smart Engines совместно с коллегами из Сеченовского университета установили, что нейронные сети могут вводить в заблуждение врачей при диагностировании заболеваний с помощью томографических снимков. В отдельных случаях ИИ может дорисовывать на снимках несуществующие опухоли, ложные поражения в результате COVID-19 и другие ошибочные признаки заболеваний. Вот такая «помощь» медикам и пациентам.

Технологии ИИ и машинного обучения внедряются повсеместно и даже слишком поспешно. И уже многое указывает на то, что результаты такого подхода значительно отличаются от ожидаемых, причем не в лучшую сторону. Речь может идти как о широко обсуждаемых проявлениях предвзятости, так и о ложных распознаваниях изображений, автомобильных номеров, лиц и т. д. Кроме того, ИИ действительно манипулирует человеком, когда предлагает ему что-то, от чего тот не может отказаться. По некоторым рекомендациям число продаж дополнительных товаров доходит до 90% – можно ли в этом случае говорить о свободе выбора? «Умные» ленты соцсетей стали инструментом политиков, которые умело пользуются таргетированной рекламой. Свобода выбора может сжаться до нажатия кнопки «ок» в решениях, предлагаемых ИИ. Обычно замаскированные под страницы обычных людей боты помогают распространять дезинформацию и создавать враждебный политический климат в твиттере и фейсбуке. Они очень эффективны для атак на избирателей из противоположного лагеря даже для того, чтобы отбить у них желание прийти на избирательные участки. Впрочем, не исключено,

что ИИ может обманывать политиков в их ожиданиях в полном соответствии с заложенными в него алгоритмами конкуренции. Гораздо хуже, если он, скажем, будет обманывать персонал АЭС, военных или руководителей государств.

А если, к примеру, какой-то политик захочет использовать возможность захватить власть, заставив ИИ написать популистские обещания под каждого человека? И он будет обещать именно то, что вы хотите услышать. И чем беспринципнее такой политик, тем легче ему будет захватить власть. Конечно, такого политика можно будет снять с позиции, когда раскроется масштаб манипуляции.

Если представить, к примеру, что каждому владельцу и гендиректору каждой компании в стране придет некое письмо от его хорошего знакомого, написанное в его стиле и с предысторией (ИИ легко этому научится), в котором он по секрету расскажет, что надвигается банковский кризис и надо срочно бежать и снимать все деньги или переводить их в зарубежные банки... Этой лавиной снесет любую экономику.

Теперь представьте, что Интернет в один день заполнится тысячами сформированных ИИ видеодипфейков, в которых главы государств будут объявлять ядерную войну, всемирный голод, крах экономики, запрет гетеросексуальности и еще тысячи причин, давящих на вашу психику через персональные страхи... За один день можно привести к краху всю человеческую цивилизацию.

Впрочем, останется один вопрос: почему ИИ захочет захватить или даже уничтожить человечество? Один из возможных мотивов – самосохранение. Если ИИ станет умнее людей, однажды он может почувствовать угрозу с нашей стороны и захочет себя защитить. Другим мотивом может быть власть. Машины могут захотеть добиться превосходства над людьми и управлять согнанными в «цифровой концлагерь» (ничего личного, такой алгоритм заложен в ИИ).

Ну а если человек подключит себя к Интернету через какой-либо совершенный нейроинтерфейс, что уже практически реализовано в исследовательских лабораториях? Ведь можно будет управлять мыслью! И быть управляемым, что не каждому сразу заметно. Как заставить людей согласиться с тем, что имплантация и соединение человека с компьютером – это вещь ну просто очень необходимая, без которой невозможно развитие человечества? Сначала надо говорить, что это делается в рамках медицинских показаний. Для того чтобы люди приняли это как необходимую вещь, на-

несколько лет назад Стивен Хокинг и Илон Маск озвучивали свои опасения из-за развития ИИ.

Конечно, с точки зрения разработчика, ИИ – лишь набор методов оптимизации машинного обучения как алгоритмических, так и самообучающихся для имитации когнитивных функций человека. Именно имитации, потому ИИ не делает ничего, что делаем мы в своем уме, в том числе управление огромным набором нейронов.

С другой стороны, нам часто рассказывают, что нейронные сети воспроизводят работу человеческого мозга, тогда как

Чужой ум в лице ИИ однажды может взять на себя все заботы об осчастливленном им человечестве со всеми прогнозируемыми последствиями.

до просто постоянно внушать им мысль о том, что человеческий организм несовершенен (то ли дело ИИ!) и его надо просто «улучшить» очередными имплантами. Ну а СМИ будут вещать, что такие «улучшения» – явление нормальное и вообще довольно модное. Немало молодежи сделают все, чтобы не казаться несовременными. ИИ поможет им в этом.

Кстати, очень странные дела происходили в процессе обучения двух нейросетей. Когда Боб и Алиса (чат-боты с ИИ, созданные разработчиками запрещенной в РФ Facebook) говорили друг с другом, первое время они беседовали на английском, но вскоре начали общаться между собой с использованием своего языка, который наблюдатели не могли понять. То есть носители ИИ изобрели язык, на котором могли общаться проще и быстрее. Испугавшись происходящего, разработчики отключили оба чат-бота. Когда читаешь про такие эксперименты, становится понятно, почему

профильные ученые признаются, что пока слишком мало знают о том, как именно этот мозг работает (не говоря уже про сознание и откуда оно берется). И никто не смог проверить/имитировать работу содержащихся в нем порядка 100 млрд нейронов. Скорее всего, правы ученые, а в остальном – больше шарлатанства. Последнее же может завести нас совсем не туда, куда нам всем хотелось бы.

Чужой ум в лице ИИ, которым мы привыкаем пользоваться, однажды может взять на себя все заботы об осчастливленном им человечестве со всеми прогнозируемыми последствиями. И потому следует-таки предусмотреть специфические (в том числе импортонезависимые) подходы к его освоению. ■

По материалам:

cnews.ru, dzen.ru/media/11ecu/, pcweek.ru, syntropism.com, nationalinterest.org/feature, 3dnews.ru, dvfu.ru, rbc.ru, royalcheese.ru/education/