**Стивен Хокинг**

**Превзойдет ли нас искусственный интеллект?**

**(из книги «Краткие ответы на большие вопросы»)**

Интеллект – главное, что определяет человека как существо. Все, что предлагает цивилизация, – продукт интеллектуальной деятельности человека.

ДНК передает жизненную программу от поколения к поколению. Более сложные жизненные формы получают информацию от сенсоров, таких как глаза и уши, и обрабатывают ее в мозге или иных системах, чтобы понять, как действовать, а затем воздействуют на окружающий мир, передавая информацию, например мышцам. В какой-то момент нашей космической истории, которая насчитывает 13,8 миллиарда лет, произошло нечто поразительное. Обработка информации стала настолько разумной, что у жизненных форм сформировалось сознание. Наша Вселенная очнулась и стала осознавать саму себя. Я считаю поистине триумфом то, что мы – не более чем звездная пыль – пришли к такому глубокому пониманию Вселенной, в которой живем.

Я не вижу существенной разницы между тем, как действует мозг земляного червя, и тем, как работает компьютер. На мой взгляд, эволюция предполагает, что может не быть качественного различия между мозгом земляного червя и человека. Из этого следует, что компьютер, в принципе, может копировать человеческий интеллект или даже превзойти его. Оказаться разумнее своих предшественников вполне реально. В процессе эволюции мы превзошли по интеллекту человекообразных обезьян, а Эйнштейн оказался умнее своих родителей.

Если компьютеры продолжат подчиняться закону Мура и удваивать быстродействие и объем памяти каждые восемнадцать месяцев, в итоге получится, что в какой-то момент ближайшего столетия они превзойдут человека по интеллектуальным способностям. Когда искусственный интеллект превзойдет человеческий в создании искусственного интеллекта, то есть сможет рекурсивно совершенствовать себя без помощи человека, мы можем стать свидетелями интеллектуального взрыва, который приведет к появлению машин, чье превосходство которых над нами в интеллектуальном смысле будет больше, чем наше – над улитками. Когда это произойдет, придется озаботиться, чтобы цели компьютеров не расходились с нашими целями. Есть искушение отмахнуться от образа высокоинтеллектуальных машин как плода воображения писателей-фантастов, но это будет ошибкой – и, в принципе, нашей самой худшей ошибкой в жизни.

В последние два десятилетия в теме искусственного интеллекта доминировали проблемы, связанные с созданием интеллектуальных агентов – систем, способных к восприятию и действиям в конкретной среде. В данном случае интеллект определяется статистическими и экономическими соображениями рациональности, или, проще говоря, способностью принимать хорошие решения, предлагать планы или делать выводы. В результате можно наблюдать высокую степень интеграции и перекрестного оплодотворения идеями в таких областях, связанных с искусственным интеллектом, как машинное обучение, статистика, теория управления, нейробиология и др. Сложившаяся система теоретических представлений в сочетании с доступными мощностями обработки данных привела к заметным успехам в таких областях, как распознавание речи, классификация образов, машинный перевод, самоуправляемые автомобили, шагающие механизмы, системы диалогового программирования.

По мере того как достижения в этих областях выходят за пределы лабораторий и становятся экономически выгодными технологиями, возникает круг благоразумия, при котором даже небольшие усовершенствования приносят значительные суммы денег, что стимулирует все новые и новые вложения в научные разработки. Сейчас широко признано, что исследования в области искусственного интеллекта демонстрируют устойчивый прогресс и что их влияние на общество соответственно увеличивается. Потенциальная выгода огромна; мы даже не можем представить, какие достижения нас ждут впереди, если наши умственные способности получат подкрепление в виде средств, предоставляемых разработчиками искусственного интеллекта. Появляется возможность покончить с эпидемиями и нищетой. Но в силу огромного потенциала искусственного интеллекта важно исследовать, как наилучшим способом воспользоваться преимуществами и избежать возможных подводных камней. Успех в создании искусственного интеллекта станет величайшим событием в истории человечества.

К сожалению, он может оказаться и последним, если мы не научимся избегать рисков. Искусственный интеллект в качестве инструмента может усилить наши существующие интеллектуальные способности, что позволит продвинуться вперед во всех областях науки и состояния общества. Но он может представлять опасность. Существующие примитивные формы искусственного интеллекта доказали свою полезность, но я опасаюсь последствий создания чего-то такого, что может сравниться с человеком или превзойти его. Есть основания полагать, что искусственный интеллект замкнется сам на себя и начнет ускоренным темпом заниматься самоусовершенствованием. Люди, ограниченные малой скоростью биологической эволюции, не смогут конкурировать с ним и окажутся вытесненными. А еще в будущем искусственный интеллект может обрести собственную волю, которая вступит в конфликт с нашей.

Есть мнение, что люди смогут еще достаточно долгое время контролировать скорость развития технологий и что потенциал искусственного интеллекта поможет решить многие мировые проблемы. Мой оптимизм в отношении судеб человечества хорошо известен, но в данном случае у меня есть сомнения.

Например, военные разных стран мира рассматривают возможность в ближайшей перспективе начать гонку автономных оружейных систем, которые способны сами выбирать и уничтожать выбранные ими цели. Пока в ООН идут дебаты о запрещении такого рода оружия, его сторонники забывают задать себе очень важный вопрос: каков вероятный результат этой гонки и желательна ли она для человечества? Неужели мы действительно хотим, чтобы дешевое оружие с искусственным интеллектом стало «калашниковым» завтрашнего дня и приобреталось на черном рынке преступниками и террористами? Думая о нашей возможности сохранять в долгосрочной перспективе контроль над более совершенными системами искусственного интеллекта, следует ли вооружать их и доверять им нашу защиту? В 2010 году сбой в компьютеризованной системе торгов вызвал обвал на фондовом рынке США; как может выглядеть компьютерный сбой в области обороны? Лучшее время прекратить разработку автономных оружейных систем – сегодня.

В среднесрочной перспективе искусственный интеллект может автоматизировать различные области человеческой деятельности, обеспечив всем равенство и процветание. Заглядывая вперед, можно сказать, что их развитию нет предела. Не существует физических законов, препятствующих частицам организоваться в формы, которые осуществляли бы более совершенные расчеты, чем те, что происходят сейчас в человеческом мозгу. Возможен взрывной переход, хотя он может выглядеть не так, как в кино. Как показал в 1965 году математик Ирвинг Гуд, машины со сверхчеловеческим интеллектом могут неоднократно улучшать свою схему – практически бесконечно, или до момента, который писатель-фантаст Вернор Виндж назвал технологической сингулярностью. Можно представить такую технологию, которая будет прозорливее финансовых аналитиков, изобретательнее ученых, хитрее политических лидеров и, в принципе, подавит нас оружием, принцип действия которого мы даже не сумеем понять. Если в краткосрочной перспективе влияние искусственного интеллекта зависит от тех, кто его контролирует, то в долгосрочной будет зависеть от того, сможем ли мы контролировать его вообще.

Короче говоря, пришествие искусственного интеллекта может оказаться для человечества и наивысшим благом, и наивысшим злом. Риск, связанный с ним, – не в злом умысле, а в компетентности. Сверхразумный искусственный интеллект может оказаться исключительно хорош в достижении своих целей, но если эти цели не станут совпадать с нашими – у нас возникнут проблемы. Вероятно, вы не настолько ненавидите муравьев, чтобы давить их ногой при каждом удобном случае, но если вы руководитель проекта по строительству экологически чистой гидроэлектростанции, а в зоне затопления окажется муравейник – увы, тем хуже для муравьев. А теперь поставим человечество в положение этих муравьев. Мы должны смотреть вперед. Если высокоразвитая внеземная цивилизация пришлет нам сообщение «мы появимся в ближайшее время», сможете ли вы ответить «хорошо, дайте знать, мы оставим свет в гостиной»? Скорее всего, нет. Но именно так может произойти с искусственным интеллектом. Мало кто всерьез интересовался этой темой, за исключением очень немногих некоммерческих организаций.

К счастью, сейчас ситуация стала меняться. Такие интеллектуалы, как Билл Гейтс, Стив Возняк и Илон Маск, разделяют мои опасения. В сообществе, связанном с созданием искусственного интеллекта, начинает укореняться здравая культура оценки рисков и информирования о социальных последствиях. В январе 2015 года я с Илоном Маском и многими другими экспертами подписал открытое письмо, посвященное искусственному интеллекту, призывая к серьезному изучению его влияния на общество. Ранее Илон Маск предупреждал, что сверхчеловеческий искусственный интеллект может принести неисчислимые прибыли, но при неосторожном использовании окажет противоположное действие на человечество. Мы с ним – члены научно-экспертного совета института «Будущее жизни», организации, занимающейся предупреждением экзистенциальных рисков, грозящих человечеству, которая и инициировала это открытое письмо. В нем звучит призыв заняться конкретными исследованиями того, как мы можем предотвратить возникновение потенциальных проблем и одновременно воспользоваться всеми благами, которые сулит нам развитие искусственного интеллекта. Мы призываем всех специалистов и исследователей в области искусственного интеллекта уделять больше внимания вопросам безопасности. Цель письма – проинформировать, а не напугать широкую общественность и политиков. Мы считаем очень важным дать понять, что все, кто занимается искусственным интеллектом, должны серьезно отнестись к этим опасениям и этическим проблемам. Например, искусственный интеллект обладает потенциалом покончить с эпидемиями и нищетой, но специалисты должны думать о том, чтобы системы искусственного интеллекта оставались под контролем.

В октябре 2016 года я также открыл в Кембридже новый центр в попытке затронуть некоторые открытые вопросы, возникающие в связи с быстрым развитием технологий в области искусственного интеллекта. Центр изучения будущего интеллекта «Леверхульм» (The Leverhulme Centre) – междисциплинарная организация, занимающаяся исследованием перспектив развития интеллекта как важнейшего фактора для будущего цивилизации и нас как вида. Значительное время мы уделяем изучению истории, которая, надо признать, преимущественно является историей глупости. Приятное разнообразие вместо изучения будущего интеллекта. Мы сознаем потенциальные опасности, но, возможно, благодаря достижениям новой технологической революции сумеем даже исправить некоторый вред, который нанесла природе индустриализация.

Последние достижения в развитии искусственного интеллекта привели к тому, что Европарламент призвал к разработке законодательства, регулирующего создание роботов и систем искусственного интеллекта. Немного удивительно, но речь зашла даже о понятии «электронной личности», определяющем права и ответственность наиболее продвинутых автономных роботов. Один член Европарламента отметил, что, поскольку роботы начинают играть все более заметную роль во многих областях нашей повседневной жизни, необходимо гарантировать, что они являются и всегда будут являться помощниками человека. В докладе, представленном в парламенте, говорится, что мир – на грани новой индустриальной революции роботов. В нем обсуждается, можно ли наделить юридическими правами роботов как электронных личностей – наряду с юридическим определением корпоративных личностей. Но в нем также подчеркивается, что разработчики роботов обязаны предусмотреть в любой системе кнопку аварийного отключения.

Это не помогло ученым на борту космического корабля в фильме Стэнли Кубрика «Космическая одиссея 2001 года», когда дал сбой разумный компьютер ЭАЛ, но это – художественное произведение, вымысел. Мы же имеем дело с фактами. Лорна Бразел, консультант международной юридической фирмы Osborn Clark, говорит в докладе, что мы не наделяем правами личности китов и горилл, поэтому нет необходимости спешить с определением личности роботов. Но осторожность чувствуется. Доклад признает возможность того, что в ближайшие десятилетия искусственный интеллект может превзойти интеллектуальные способности человека и оспорить нынешние отношения между человеком и роботом.

К 2025 году в мире будет примерно 30 мегаполисов с населением более 10 миллионов человек. Когда все эти люди пожелают получать товары и услуги в любое удобное для них время, поможет ли нам технология соответствовать требованиям такой срочной коммерции? Разумеется, роботы заметно ускорят процесс онлайновой торговли. Но чтобы революционизировать процесс покупок, они должны быть достаточно быстрыми, чтобы выполнять заказы день в день.

Быстро расширяются возможности контактов с окружающим миром без необходимости физического присутствия. Как вы догадываетесь, я считаю это очень удобным, и не в последнюю очередь потому, что городская жизнь для всех нас становится все более насыщенной. Сколько раз у вас появлялась мысль обзавестись двойником, который взял бы на себя часть ваших нагрузок? Создание реалистичных цифровых заменителей себя – дерзкая идея, но новейшие технологические достижения говорят, что она уже не настолько безумная, как кажется.

Когда я был моложе, развитие технологий указывало, что в будущем у нас должно появиться больше свободного времени. Но в реальности чем больше мы можем, тем больше дел на нас наваливается. Города уже полны машин, которые расширяют наши возможности, но вот если бы мы могли оказываться в двух местах одновременно… Мы привыкли к автоматизированным голосам на телефоне и в системах оповещения на транспорте. Изобретатель Дэниэл Крафт уже изучает вопрос, сможем ли мы воспроизвести себя визуально. Один вопрос – насколько убедительным окажется наш аватар?

Интерактивное общение может оказаться полезным в массовых открытых онлайн-курсах и в области развлечений. Это может быть очень интересно – цифровые актеры, которые всегда остаются молодыми и способными на то, что никаким иным образом реализовать невозможно. Наши будущие кумиры могут и не быть реальными.

Наша связь с цифровым миром станет определяющей характеристикой дальнейшего прогресса общества. В умных городах умные дома будут оснащены такими интерактивными устройствами, что взаимодействие с ними не потребует почти никаких усилий.

Изобретение пишущей машинки существенно облегчило нам способ взаимодействия с машинами. Прошло сто пятьдесят лет, и сенсорные экраны открыли новые пути общения с цифровым миром. Последние достижения в области искусственного интеллекта, такие как самоуправляемые автомобили или компьютер, выигрывающий в го, – приметы того, что нас ждет в будущем. В эти технологии, которые уже во многом определяют нашу жизнь, делаются огромные инвестиции. В ближайшие десятилетия они внедрятся во все сферы жизни общества, разумно поддерживая и консультируя нас во многих областях, включая здравоохранение, образование и науку. Достижения, которые мы видим сейчас, явно бледнеют перед теми, что принесут нам ближайшие десятилетия, и мы не в состоянии предсказать, чего мы сможем достичь, когда наше сознание будет подкреплено искусственным интеллектом.

**ПОЧЕМУ МЫ ТАК БОИМСЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА? ВЕДЬ ЧЕЛОВЕК ВСЕГДА МОЖЕТ ВЫКЛЮЧИТЬ КОМПЬЮТЕР**

Человек спросил у компьютера: «Есть ли Бог?» Компьютер ответил: «Теперь есть!» – и расплавил выключатель.

Возможно, плоды новой технологической революции повысят качество человеческой жизни. Например, сейчас ученые занимаются разработкой искусственного интеллекта, который поможет бороться с параличом у людей, получивших спинномозговые травмы. С помощью силиконовых чипов-имплантов и беспроводного электронного интерфейса между мозгом и телом технология позволит людям управлять движениями тела силой мысли.

Убежден, будущее коммуникаций – за взаимодействием мозга и компьютера. Есть два способа: электроды на черепе и импланты. Первый – все равно что смотреть сквозь запотевшее стекло, второй лучше, но есть риск инфекции. Если мы сможем подключить человеческий мозг к интернету, у каждого в постоянном распоряжении будет вся «Википедия».

По мере углубления связи между человеком, девайсами и информацией мир начинает изменяться быстрее. Вычислительные мощности растут, и уже становятся доступны квантовые вычисления. Это приведет к экспоненциальному совершенствованию искусственного интеллекта, поднимет на новый уровень шифрование. Квантовые компьютеры изменят все, даже биологию человека. Уже существует методика, которая называется CRISPR,[21] дающая возможность направленного редактирования ДНК. В основе этой технологии редактирования геномов – защитная система бактерий. Она позволяет точно выделять и редактировать последовательности в генетическом коде. Лучшие намерения, которые движут учеными, заключаются в возможности исправлять генетические мутации и, соответственно, бороться с генетическими заболеваниями. Однако воздействие на ДНК может совершаться и с менее благородными целями. Уже остро стоит вопрос – как далеко мы можем зайти с генной инженерией? Нельзя рассматривать возможности лечения заболевания двигательных нейронов – как мой БАС – и не видеть сопряженные с этим опасности.

Интеллект определяется способностью адаптироваться к изменениям. Человеческий интеллект – результат множества поколений генетического отбора тех, кто наилучшим образом умел приспосабливаться к окружающим обстоятельствам. Мы не должны бояться изменений. Нужно лишь направлять их в наших интересах.

Каждый из нас играет роль в том, чтобы и нынешнее, и следующее поколение проявляли готовность с молодых лет интересоваться наукой, реализовывали свой потенциал и создавали лучший мир для всего человечества. Мы должны выйти за рамки теоретических дискуссий о том, каким должен быть искусственный интеллект, – лучше представить, каким он может быть. У каждого из нас есть возможность расширять границы представлений о приемлемом или ожидаемом и смотреть вперед. Мы на пороге смелого нового мира. Жить в нем будет увлекательно, но и рискованно. Мы станем первопроходцами.

Когда мы стали пользоваться огнем, он часто доставлял нам неприятности. Потом мы изобрели огнетушитель. Обладая гораздо более мощными технологиями, такими как ядерное оружие, синтетическая биология и искусственный интеллект, мы должны вести себя более предусмотрительно и стараться сразу все делать правильно, потому что второго шанса может и не оказаться. Наше будущее – это состязание между нарастающим могуществом технологий и мудростью, с которой мы их используем. Будем стараться, чтобы победила мудрость.