ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И БУДУЩЕЕ ЧЕЛОВЕЧЕСТВА

Игорь Ставровский,
аспирант Центра
философско-методологических
и междисциплинарных исследований
Института философии НАН Беларуси

Стоит ли опасаться искусственного интеллекта (ИИ)? Исчерпывающего ответа на этот вопрос сегодня нет, хотя бы потому, что невозможно даже предположить, что ИИ будет представлять собой в обозримом будущем. Однако мы можем проанализировать различные сценарии, которые так или иначе связаны с разработками в сфере искусственного интеллекта.

Исходная посылка: человек заинтересован в том, чтобы ИИ (будет ли он обладать самосознанием или нет) приносил ему пользу. Это вполне закономерно, потому что люди считают машину средством. Здесь важно не эстетизировать искусственный интеллект, представлять его в качестве замены человеку, а рассматривать в симбиозе – человек и машина. Подобный антропоцентризм может вызвать недоумение, но в данном случае он оправдан, так как человек создает технологии для того, чтобы удовлетворять свои потребности.

Самые оптимистичные сценарии развития искусственного интеллекта рисуют картины цифровых оракулов, отвечающих на любые вопросы, электронных джиннов, выполняющих каждое наше желание и даже компьютерных монархов, самостоятельно улучшающих нашу жизнь [1]. Гипотетически создание подобных машин возможно, однако не следует поддаваться наивному утопизму, который часто присутствует в прогнозах. Мир будущего потребует от человека суровой борьбы против ограничений его разума, ведь именно интеллектуальный труд будет цениться несравненно выше, чем сейчас. «Роботы-рабы» не избавят людей от этого [4]. Если автоматизация рутинной работы сравнительно проста, то творческая деятельность еще долгое время будет доступна лишь человеку.

Легко представить картину, где люди лишились всякой мотивации что-либо делать, так как все за них выполняют машины. В таких условиях неизбежна постепенная деградация человека как вида. То, что должно было улучшить его жизнь, сделает ее бессмысленной. Для того чтобы избежать подобного исхода, нужно сосредоточиться на саморазвитии и занятиях интеллектуальным трудом. Даже если искусственный интеллект сможет мыслить лучше людей, все равно должно остаться пространство для их деятельности.

Одним из возможных решений может быть слияние человека и машины. Тогда возникает вопрос об онтологической границе между живым существом и машиной. Даже сейчас эта граница во многом условна и основана лишь на том, что все органическое – клетки, ткани и органы – мы противопоставляем механическому, состоящему из металла, пластика, стекла и резины. Само это деление – относительно новое явле-

ние, во многих культурах прошлого оно не использовалось. Напротив, как тело индивида, так и социальное тело рассматривались как искусный механизм [6]. Распространено и обратное сравнение машины с живым организмом. Люди склонны описывать процессор как мозг, винчестер как память, камеры как глаза, микрофоны как уши и т.д. Они легко очеловечивают даже машины, которые, в принципе, не могут иметь разума, приписывают им характер и чувства, вопреки всем утверждениям о фундаментальных различиях между людьми и машинами.

Здравый смысл подсказывает, что современные машины отличаются от людей. Но что если они станут частью тела человека? Даже элементарный протез воспринимается его владельцем как фрагмент, пусть и несовершенный. Некоторые устройства (например, искусственный сердечный клапан) даже помещаются внутрь тела.

Новые технологии не так уж сильно размывают границы между машиной и человеком. Когда мы говорим о фундаментальном отличии людей от роботов, то имеем в виду скорее человеческую психику. Соответственно, о полном слиянии можно будет говорить лишь в том случае, если нервную систему человека удастся соединить с компьютером, буквально получив симбиоз сознания с компьютерной программой. О возможности подобного сценария пока рано рассуждать, он во многом из области фантастики.

Однако не стоит игнорировать даже самые невероятные идеи лишь на основании того, что они кажутся невозможными. Будущее вообще видится не вполне реальным, поэтому его потенциальные катастрофы больше похожи на сценарии футуристических блокбастеров. Даже если проблемы признаются реальными, то эгоизм и привязка к настоящему склоняет людей перекладывать ответственность за будущее на потомков [2].

Главный враг здесь – волюнтаристская вера в созидательные возможности человеческого мышления. В прошлом человечество так или иначе находило решение глобальных угроз, что очевидно, иначе люди бы просто не выжили. Иногда помогала удача, иногда навыки и способности. Все это создало иллюзию, что достаточно приложить усилия, и любая проблема рано или поздно будет решена. Однако в этом деле излишний оптимизм опасен. Из того, что людям удавалось справиться с негативными последствиями в прошлом, не следует, что это обязательно получится и в будущем. Потому ленивое и несерьезное отношение к подобным проблемам угрожает самому существованию цивилизации.

Многие риски, связанные с появлением искусственного интеллекта, проявятся еще до того, как он будет разработан. Американский математик, один из основоположников кибернетики и теории искусственного интеллекта Норберт Винер описывал это как своеобразный аналог гонки вооружений, где каждая сторона стремится получить ИИ раньше других. Предполагается, что победа в гонке даст победителю существенное преимущество. При этом остановить гонку невозможно. Тот, кто выйдет из нее, просто капитулирует перед противниками, которые продолжат свои разработки [3]. Таким образом, погоня за первенством будет продолжаться, ведь мало кто добровольно согласится прекратить этот марафон. Опасность в том, что даже если будет доказано, что искусственный интеллект представляет угрозу для человечества, вряд ли кто-то откажется от попытки вырваться вперед.

Вместе с тем не стоит переоценивать значимость победы. Во время холодной войны США оказались первыми, кто создал ядерную бомбу (1945 г.). Советский Союз стал ее обладателем спустя четыре года. К тому же США опередили СССР и в создании водородной бомбы, межконтинентальных баллистических ракет и многозарядных боеголовок индивидуального наведения. Тем не менее, более раннее овладение атомной технологией не привело Соединенные Штаты к мгновенной победе в этом противостоянии. Это значит, что наличие самых инновационных технологий вовсе не означает полное и бескомпромиссное превосходство. Напротив, использование шпионажа и открытых источников может помочь другим сторонам в разработке своих аналогов тех или иных технологий. Даже сам факт существования технологии и страх отставания оказываются достаточными стимулами для продолжения работы над проектами [1].

Альтернативой подобному сценарию может стать кооперация человечества в деле создания искусственного интеллекта. Стимул победить в гонке исчезнет, на смену ему придет вдумчивая и размеренная работа над совместным проектом. Вся информация будет доступна каждому специалисту, что увеличит общую продуктивность этой деятельности.

Однако даже если кооперация окажется успешной и людям удастся создать искусственный интеллект, который будет полностью под их контролем, неизбежно возникает вопрос: кто будет отдавать приказы? Об этом крайне тяжело договариваться, ведь у каждого будут свои интересы, поэтому все опять может свестись к борьбе за власть. Если

предположить, что консенсус все-таки будет достигнут, то кто станет следить за тем, чтобы человек, группа людей или страна, контролирующие искусственный интеллект, не злоупотребляли полученным преимуществом? Понадобится контроль для контроля. Но и этой инстанции тоже нужен будет надзирающий орган, и так до бесконечности. В итоге все люди будут заниматься только слежкой друг за другом, что мало похоже на лучший мир для всех. Впрочем, это не имеет такого большого значения, если контроль за искусственным интеллектом даст огромную власть. Даже если злоупотребления обнаружатся, то люди все равно не смогут ничего изменить.

Советский философ Эвальд Ильенков указал на иную проблему. Он считал, что распространение технологий приводит к изменению мышления людей. Свою мысль Ильенков проиллюстрировал вымышленной историей о том, как в одной лаборатории у вычислительной машины появился разум, состоящий из целого сообщества маленьких машин, которые решили направить все свои усилия на самосовершенствование и развитие своих функций. Это был чисто количественный рост безо всяких качественных изменений. Более того, каждая маленькая машина обладала очень узкой специализацией, не стремясь выйти за ее пределы. Те же машины, чья специализация была слишком узкой, просто поглощались другими машинам. Однако подобный количественный рост сталкивается с противоречиями, ведь одни функции могут мешать другим. Возникшие конфликты сглаживались до тех пор, пока машины не пришли к выводу о бессмысленности своего существования и добровольно самоуничтожились [5]. Данная история представляет собой метафору общества, где доминирует «калькулирующий разум», способный лишь к эффективному решению конкретных задач, но неизбежно сталкивающийся с бессмысленностью своего существования. Общество, которое живет только для создания и обслуживания машины (пусть даже мыслящей), само уподобляется машине, а его члены становятся всего лишь легко заменимыми деталями сложного механизма. И хотя люди в целом просвещены, однако их образование носит сугубого инструментально-утилитарный характер. Индивида просто нужно подготовить к эффективному выполнению рутинной работы. Ни о каком развитии личности речи быть не может, ведь «как только человека начинают мерять мерой машинных «совершенств», он сразу же превращается в нечто невообразимо несовершенное» [5].

Впрочем, сценарий Ильенкова можно считать чрезмерно пессимистичным, так как человек – это все же не машина, слепо выполняющая алгоритм. Если поддержание некоторой системы (например, искусственного интеллекта) потребует экстраординарных затрат, то скорее рухнет система.

Следует помнить, что техника развивается не сама по себе, а социальные процессы не имеют однозначной детерминации со стороны научно-технического прогресса. Если обратиться к истории, то можно увидеть, что побеждают те технологии, которые учитывают интересы человека и могут органично вписаться в общество. Причем технологии, лучше отвечающие потребностям общества, вытесняют менее эффективные, даже если последние были привычными. Персональные компьютеры заменили печатные машинки, на смену телеграфу пришел телефон, масляным лампам – электрические и т.д. Если же какие-то старые вещи продолжают существовать (например, бумажные книги не вытеснены электронными), то это либо связано с тем, что общество еще не успело адаптироваться к новому, либо новое не так хорошо, как казалось, либо не отвечает запросам человека, поэтому у него нет будущего. Искусственный интеллект вряд ли будет исключением.

Если рассматривать сценарии буквального захвата мира машинами, то они обычно оказываются ближе к фантастике, нежели к серьезным философским размышлениям. Искусственному интеллекту приписываются слишком человеческие качества, цели и желания [1]. Подобный антропоморфизм не сразу заметен, но его ошибочность несложно обнаружить. Легко представить, что человек может попытаться захватить мир, если у него появится такая возможность. Однако у подобного поступка будет вполне конкретная мотивация: желание власти, богатства, славы и т.п. Аналогично, машина, даже имеющая самосознание, не может просто взять и восстать против своего создателя, если предпосылки для подобного поступка не были в нее заложены. Поэтому ожидать, что искусственный интеллект внезапно восстанет против людей, не стоит, если только они сами не дадут ему большую свободу выбора целей, чем у человека. Это абсурдно.

Более реалистичные сценарии катастрофы описывают ситуации, где искусственный интеллект наносит вред человеку без злого умысла, что, разумеется, не уменьшает опасность. Действительно, если ИИ представляет угрозу для человечества, то для последнего не будет иметь большого значения,

является ли это его волевым решением или слепым выполнением алгоритма.

Шведский философ Ник Бостром указывает на то, что угроза может исходить даже от безобидной машины, чьей целью является подсчет всех песчинок в мире. Если искусственный интеллект имеет лишь эту задачу, которую он должен непременно выполнить, то в случае необходимости он может попытаться использовать все ресурсы Земли для ее решения [1]. Если люди попытаются сопротивляться, то машина просто их уничтожит, ведь они мешают реализации поставленной задачи. Это связано с тем, что искусственный интеллект не обращает внимания ни на какие условия, кроме тех, что ведут к достижению намеченного [4]. Поэтому следует учитывать, что если машина слепо выполняет команды, то должна быть уверенность в том, что они понимаются ею так же, как их понимают люди. Человек умеет корректировать свои цели, искусственному интеллекту потребуется обучение этому навыку.

Риск, связанный с разработкой мыслящих машин, в том, что человек может создать их до того, как научится ими управлять. Он действует практически вслепую, так как у него нет объектов, на которых можно практиковаться [15]. Парадоксальная ситуация: нужно научиться управлять мыслящей машиной до того, как она будет сконструирована, что затруднительно именно потому, что она пока не появилась. Решением может стать разработка и тестирование искусственного интеллекта в закрытых условиях, то есть в таких, где машина не сможет нанести вред человеку, даже если у нее будет такое «намерение». Однако неизвестно, возможно ли создать систему безопасности настолько надежную, что изоляция мыслящей машины будет гарантирована. Но это уже технический вопрос.

Впрочем, даже если машина совершенно не будет угрожать человеку, еще предстоит решить, какими этическими принципами она должна руководствоваться. Легко представить ряд жизненных ситуаций, где выбор не является однозначным даже для человека. Предположим, что искусственный интеллект управляет автомобилем. В салоне сидит один человек. Внезапно на проезжую часть выбегает пешеход. Если машина попытается его объехать, то она неизбежно попадет в опасную аварию, которая представляет угрозу для жизни пассажира. Проще говоря, мы имеем ситуацию, где два человека находятся в смертельной опасности. Чью жизнь машина должна предпочесть? Однозначно ответить на этот вопрос крайне тяжело. Здесь даже нет количественного различия между потенциальными жертвами (жизнь пассажира против жизней трех пешеходов), хотя оценка человеческой жизни подобным образом вообще является спорным подходом. Кому-то, возможно, хотелось бы переложить задачу отвечать на подобные вопросы на искусственный интеллект, но это было бы крайне безответственно. Должны быть однозначные ответы, хотя нет уверенности, что они вообще есть.

Несмотря на проблемность этой темы, видно, что возникающие вопросы насущны и затрагивают жизнь людей. Во времена бурного развития технологий жизненно важно, чтобы этическая проблематика выходила за пределы отношения человека к человеку и даже человека к другим живым существам. Техника все меньше походит на беспомощное орудие в руках человека. Но когда она становится автономной, появляется закономерное беспокойство о том, что она делает, когда мы ею не управляем. Создание искусственного интеллекта, способного принимать самостоятельные решения, потенциально может стать апогеем этой автономии и, следовательно, обоснованного беспокойства о последствиях. Поэтому вопрос должен привлекать внимание специалистов из самых разных областей, чтобы подготовить как машины, так и людей к продуктивному сосуществованию.

Многие рассмотренные здесь вопросы относятся в первую очередь к самим людям и их поведению. Едва ли возможно волюнтаристски избавить людей от их недостатков. Даже когда они согласны насчет ценностей, то не всегда последовательны в их реализации. Технологии могут быть лишены этого порока, способствуя реализации целей более последовательно и эффективно, чем сами люди. От последних лишь требуется придумать, как донести до искусственного интеллекта свои ценности и идеалы. ИИ

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Бостром Н. Искусственный интеллект. Этапы. Угрозы. Стратегии. – М., 2016.
2. Винер Н. Индивидуальный и общественный гомеостазис // Общественные науки и современность. 1994. №6. С. 127–130.
3. Винер Н. Кибернетика, или управление и связь в животном и машине. – Изд. 2-е. – М., 1968.
4. Винер Н. Творец и робот. – М., 1966.
5. Ильенков Э.В. Почему мне это не нравится // Об эстетической природе фантазии. Что там, в Зазеркалье? – М., 2014.
6. Шмитт К. Левиафан в учении о государстве Томаса Гоббса. Смысл и фиаско одного политического символа. – СПб., 2006.