Фундаментальные исследования в сфере развития информационного общества

СУБЪЕКТНОСТЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: СТАРЫЕ ВОПРОСЫ В НОВЫХ КОНТЕКСТАХ

Алексеева Ирина Юрьевна

Доктор философских наук, доцент
Институт философии Российской академии наук, сектор философских проблем социальных и гуманитарных наук, ведущий научный сотрудник
Москва, Россия
ialexeev@inbox.ru

Аннотация

Тема субъектности искусственного интеллекта как технической системы долгое время была предметом интереса эпистемологов и философов науки. В последние годы в обсуждении проблематики, связанной с прогрессом интеллектуальных систем, все более активно участвуют этики, культурологи, гуманитарии других специальностей. В этом контексте полезно учитывать опыт проходивших в 60-х -80-х годах XX века дискуссий о статусе «компьютера» как возможного субъекта мышления, знания, понимания, принятия решений.

Ключевые слова

искусственный интеллект; принятие решений; философия искусственного интеллекта; субъектность; автономные роботы; этика искусственного интеллекта

Введение

Двадцать семь лет тому назад, в 1993 году, была издана моя монография «Человеческое знание и его компьютерный образ» [1]. Одна из глав называлась «Компьютер как квазисубъект знания», а один из параграфов этой главы – «Субъект или инструмент?» Речь шла о накопленном к тому времени опыте дискуссий о правомерности подхода к компьютеру как субъекту мышления, знания, понимания, принятия решений. Под «компьютером» в данном контексте понималась система, создание которой стало возможным благодаря исследованиям и разработкам в области искусственного интеллекта как быстро развивавшегося научного направления, оформившегося со временем в соответствующий раздел науки. Машинное доказательство теорем, распознавание образов, восприятие естественного языка, машинный перевод с одного языка на другой, моделирование игр (включая шахматы) – эти и многие другие проблемы разрабатывались в указанной области. Если применительно к сфере исследований выражение «искусственный интеллект» употреблялось без кавычек, то, когда речь заходила об искусственном интеллекте как свойстве технической системы, содержащей соответствующие программные и аппаратные средства, в публикациях 80-х и даже 90-х годов «искусственный интеллект» нередко брали в кавычки. Восприятие машинного интеллекта как ненастоящего и/или не имеющего перспективы превратиться в подлинный искусственный интеллект (в англоязычной литературе употреблялось выражение “genuine AI”) было характерно для многих участников дискуссий.

1 Машинная субъектность. Мышление, понимание, принятие решений

Широкую известность в 80-е годы получил мысленный эксперимент Дж. Серла – так называемая «китайская комната» [15]. Не отрицая абстрактной возможности появления в будущем машин, способных «по-настоящему» понимать естественные языки, Серл стремился показать, что работа компьютерных программ не может служить примером такого понимания. Мысленный эксперимент состоял в следующем. Пусть запертый в комнате человек, незнакомый с китайским

языком, но знающий английский, общается с оставшимися вне комнаты людьми, получая написанные на бумаге вопросы и выдавая, в свою очередь, записанные на бумаге ответы. Если этого человека снабдить разложенными определенным образом листками с китайскими записями, а также хорошей англоязычной инструкцией, позволяющей сравнивать символы и переходить от одних листков к другим, чтобы в итоге выбрать именно те, где изображены подходящие ответы на задаваемые вопросы, то люди вне комнаты смогут считать, что имеют дело с лицом, понимающим китайский язык. На самом же деле обитатель «китайской комнаты» лишь симулирует понимание китайского, как, по Серлу, симулирует понимание и компьютерная программа, предоставляющая правильные ответы на задания, которые она получает.

Между тем, многие участники дискуссий о машинной субъектности не видели принципиальных различий между машинным и человеческим интеллектом, если тот и другой рассматривать с единых позиций. Так, чемпион мира по шахматам М. Ботвинник, занимавшийся кибернетическим моделированием шахматной игры, писал: «Условимся, что будем оценивать интеллект с кибернетической точки зрения. А как тогда его можно оценить? Это способность принимать решение – хорошее решение – в сложной ситуации при экономном расходовании ресурсов. Если подойдем с этой точки зрения, то не усмотрим различий между естественным и искусственным интеллектом» [2, с. 51]. Дебаты по поводу статуса машинного интеллекта стали стимулом к философским исследованиям, реконструировавшим теоретико-познавательную традицию рассмотрения мышления как вычисления. Примером может служить изданная в 1981 году книга Дж. Хогеланда, где английский философ XVII века Т. Гоббс представлен как «дедушка искусственного интеллекта» [13]. Влияние идей философов прошлого на исследования в области искусственного интеллекта изучается и в нынешнем столетии [6].

Характерной для второй половины 80-х годов стала позиция, согласно которой ответ на вопрос о возможности «подлинного искусственного интеллекта» не может быть найден в ходе теоретических дискуссий, однако будет решен «эмпирически» в будущем [11]. Отметим, что значение слова «эмпирический» в подобных контекстах заслуживает особого внимания сегодня, когда языковая практика повседневности свидетельствует об отсутствии сомнений в существовании искусственного интеллекта как технологии, наделяемой тем или иным видом субъектности.

Достигнутые с начала 90-х успехи в создании интеллектуальных систем выглядят впечатляюще. В 1997 году компьютер впервые обыграл в шахматы действующего чемпиона мира, а вскоре аудитория, следящая за событиями подобного рода, стала интересоваться не вопросом, кто выиграл поединок, а вопросом, как долго носитель естественного интеллекта сопротивлялся интеллекту искусственному. Автоматические переводы с одного языка на другой все еще несовершенны, однако их качество, поначалу не выдерживавшее критики, быстро улучшается. Люди имеют возможность, используя устную речь, давать задания своим компьютерам и смартфонам. Технологии искусственного интеллекта распознают изображения, широко применяются в управлении, поддерживают принятие решений в разных областях деятельности, включая медицину и юриспруденцию. Выражения «решения принимаются при поддержке искусственного интеллекта» и «решения принимает искусственный интеллект» порой используются как взаимозаменяемые, однако этот факт речевой практики - не основание для того, чтобы заменить ответственность человека ответственностью технологии.

В 70-е годы XX века Дж. Вейценбаум стремился привлечь внимание к вопросам ответственности людей, создающих и применяющих интеллектуальные системы [3]. Этот ученый, получивший известность благодаря исследованиям и разработкам в области искусственного интеллекта, настаивал, что что пределы применимости вычислительных машин должны определяться не только технической осуществимостью тех или иных идей, но, прежде всего, императивами этической допустимости и долженствования. Замена психиатра или судьи вычислительной системой была бы, по мнению Вейценбаума, аморальной, как и подобная замена в любой другой сфере, где важную роль играет гуманное отношение к людям и понимание людей.

Возможно, первым автором, выполнившим профессиональное философское исследование проблемы субъектности искусственного интеллекта, был Дж. Мур. В работе с названием «Есть ли решения, которые не никогда не следует принимать компьютеру?» [14]. Мур, возражая Вейценбауму, утверждал, что компьютер может считаться не только инструментом, но и автономным агентом: в некоторых контекстах практически полезно рассматривать компьютер как субъект принятия решений. Кроме того, неправомерно заранее определять, в каких сферах искусственная система не должна принимать решений. Если решения компьютера будут полезны

для больных людей, то негуманным следует считать запрет на такие решения. Сказанное не означает, что этических ограничений на компьютерные решения не должно быть вовсе. Искусственную систему, по мнению Мура, никогда нельзя допускать к принятию решений о том, каковы должны быть базисные цели и ценности человека и какие из этих целей и ценностей являются приоритетными.

2 Субъектность в зеркале современной этики, футурологии и философии техники

Российские философы, профессионально работающие в области этики, лишь в последние годы стали проявлять интерес к феномену искусственного интеллекта. Следует подчеркнуть, что осмысление соответствующей проблематики происходит в условиях, когда развитие интеллектуальных технологий (как и цифровых технологий в целом) достигло уровня, значительно превосходящего уровень тридцатилетней давности. В этих условиях отнюдь не выглядит преждевременной поставленная А. В. Разиным проблема этики искусственного интеллекта как области, которая не ограничивается «этическими правилами создания интеллектуальных систем, необходимыми при программировании», но включает также этику технических систем будущего. Исходя из предпосылки, что свобода воли человека есть именно то, что позволяет человеку решать этические задачи, А. В. Разин осуществляет проекцию данной предпосылки на техническую систему и приходит к выводу, что об этике искусственного интеллекта в собственном смысле слова можно будет говорить лишь в том случае, если в работу интеллектуальной системы будет заложена принципиальная возможность ошибки. «Этика непосредственно начинается тогда, когда появляется способность реагировать на собственные ошибки, осуществлять рефлексию поведения, учитывая при этом мнения других людей, - пишет А. В. Разин. - Такая же принципиальная возможность ошибки должна быть заложена и в работу искусственного интеллекта, чтобы можно было говорить о его этике в собственном смысле слова. Должны быть также выполнены условия коммуникации машин, их взаимных оценок и наличия у них феноменального опыта» [8, с. 57]. Представляется, что в данном контексте немаловажен и вопрос о том, будут ли обладающие аналогом свободы воли интеллектуальные машины способны к коммуникации с людьми.

Порой бывает трудно определить, где проходят границы между философией искусственного интеллекта, футурологией искусственного интеллекта и формирующейся религией искусственного интеллекта. В популярной литературе и в киноискусстве создается образ могущественного искусственного субъекта будущего, использующего людей в качестве ресурса для решения своих собственных задач. К соответствующему направлению применимо выражение «апокалиптический искусственный интеллект», используемое антропологом Р. Гераци. В книге «Апокалиптический ИИ: видения небес в робототехнике, искусственном интеллекте и виртуальной реальности» [12]. Р. Гераци ведет речь об особом сегменте популяризаторской и футурологической литературы, созданной учеными, которые внесли вклад в развитие робототехники, искусственного интеллекта и информационных технологий. Научная репутация таких людей способствует доверию читателей к их предсказаниям относительно будущего технологий. Эти предсказания представлены в футурологических произведениях, изображающих будущее, где люди вытеснены машинами и, возможно, машины остались единственной формой интеллектуальной жизни на планете. Такова футурология Моравека, Варвика, Минского, Курцвейла. К этому же ряду можно отнести футурологию Р. Уолша, утверждающего бесперспективность homo sapience как обладателя естественного интеллекта, во многих отношениях уступающего быстро развивающемуся интеллекту цифровому. Единственный способ для человечества не исчезнуть бесследно с лица земли Р. Уолш видит в том, чтобы превратиться в homo digitalis – цифровую версию человека разумного. «Homo digitalis, - пишет Р. Уолш, – будут гораздо умнее homo sapiens благодаря тому, что наш мозг будет помещен в цифровую среду. В конце концов, трудно будет отличить наши мысли от единого облачного разума ИИ… Homo digitalis будут хозяевами этой цифровой вселенной. В некотором смысле мы станем цифровыми богами» [10, с. 21]. Вряд ли подобные предсказания смогут выдержать критику с позиций методологии научного прогнозирования, однако они позволяют увидеть в новых контекстах проблему субъектности человека и искусственной системы в стремительно технологизирующемся мире.

С позиций современной эпистемологии искусственный интеллект рассматривается, главным образом, не в качестве конкурента интеллекту естественному, но как «срастающийся» с последним и участвующий в образовании гибридных форм. При этом, как справедливо отмечает Е. А. Никитина [7, с. 22], различные виды интеллектуальных систем управления и обработки

информации начинают выполнять функции коллективного субъекта познания и деятельности. Заметим, что коллективный субъект в разных формах, включая организации и общество в целом, существовал задолго до появления первых компьютеров.

Широкий спектр эпистемологических, философско-методологических, этических вопросов связан с созданием автономных роботов, обладающих интеллектуальной информационно-управляющей системой, которая решает задачи анализа обстановки и выработки команд для движения и манипулирования объектами [5]. Искусственный интеллект робота способен формировать план действий, оценивать информацию об успешных и неудачных действиях, обучаться и выбирать стратегии поведения. Невозможность заранее предсказать действия такого робота в конкретной ситуации при необходимости встраивать поведение последнего в социальную среду побуждает рассматривать возникающие в таких случаях технологические риски как риски социальные [4].

Заключение

А.И. Ракитов в одной из последних работ утверждал, что роботизация, автоматизация и развитие искусственного интеллекта «должны стать центральной темой философского дискурса современности», поскольку именно эти процессы способны радикально изменить ход общественной жизни, общественное сознание и понимание человеком собственной «геоисторической миссии в развитии планеты Земля» [9, с. 47]. Место, которое занимает соответствующая проблематика в философских дискуссиях сегодня, трудно назвать центральным, однако в последние годы мы наблюдаем значительный рост интереса к ней со стороны философов, в том числе философов в нашей стране.

Литература

1. Алексеева И. Ю. Человеческое знание и его компьютерный образ. М.: ИФ РАН, 1993. 218 с.
2. Ботвинник М.М. Почему возникла идея искусственного интеллекта? // Кибернетика: Перспективы развития. М., 1981. С. 51- 56.
3. Вейценбаум Дж. Возможности вычислительных машин и человеческий разум: от суждений к вычислениям / Пер. с англ. М.: Радио и связь, 1982. 368 с.
4. Горохов В. Г., Декер М. Технологические риски как социальная проблема при разработке и внедрении интеллектуальных автономных роботов // Глобальное будущее 2045. Конвергентные технологии (НБИКС) и трансгуманистическая эволюция. М., 2013. с. 82–93.
5. Диане С. К. Д. Обучение и социальная интеграция автономных роботов на основе применения современных когнитивных технологий // Философия науки и техники. 2018. Т. 23. № 2. С. 89–102.
6. Клюева Н. Ю. Влияние идей Г. Лейбница на развитие компьютерных наук и исследования в области искусственного интеллекта // Вестник Моск. Ун-та. Сер. 7. Философия. 2017. № 4. С. 79-92.
7. Никитина Е. А. Проблема субъекта познания в современной эпистемологии // Перспективы науки и образования. 2015. № 2. С. 16-24.
8. Разин А. В. Этика искусственного интеллекта // Философия и общество. 2019. № 1 С. 57–73.
9. Ракитов А. И. Философия, роботы, автоматы и зримое будущее // Философия и общество, № 3 2019 35–48.
10. Уолш Т. 2062: время машин / Пер. с англ. М.: АСТ, 2019. 280 с.
11. Boden M. Artificial intelligence in psychology: Interdisciplinary essays. Cambridge (Mass.); L: MIT press, 1988. 188p.
12. Geraci, R. Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality. Oxford University Press, 2010. 248 p.
13. Haugeland J. Artificial intelligence: The very idea. Cambridge (Mass.); L: MIT press. 1981. 287 p.
14. Moor J. Are there decisions computers should never make? // Ethical issues in the use of computers. Belmont, 1985. P. 120-130.
15. Searle J. R. Minds, brains and programs // Artificial intc1ligcncc: The case against. L.; Sydney, 1987. P. 18-40.

SUBJECTNESS OF ARTIFICIAL INTELLIGENCE: OLD QUESTIONS IN NEW CONTEXTS

Alekseeva Irina Yurievna

Doctor of philosophical sciences
Institute of Philosophy, Russian Academy of Sciences, Department of Philosophical Problems in Social Sciences and Humanities, leading researcher
Moscow, Russia
ialexeev@inbox.ru

Abstract

The topic of subjectivity of artificial intelligence as a technological system has long been the subject of interest of epistemologists and philosophers of science. In recent years, ethicists, cultural scientists, and other humanitarians have been increasingly involved in the discussion of issues related to the progress of intelligent systems. In this context, it is useful to take into account the experience of discussions that took place in the 60s-80s of the XX century about the status of the "computer" as a possible subject of thinking, knowledge, understanding, and decision-making.

Keywords

Artificial Intelligence, problems solving, philosophy of artificial intelligence, subjectivity, autonomous robots, ethics of artificial Intelligence

References

1. Alekseeva I. Yu. Chelovecheskoe znanie i ego komp'yuternyj obraz. M.: IF RAN, 1993. 218 s.
2. Botvinnik M.M. Pochemu voznikla ideya iskusstvennogo intellekta? // Kibernetika: Perspektivy razvitiya. M., 1981. S. 51- 56.
3. Vejcenbaum Dzh. Vozmozhnosti vychislitel'nyh mashin i chelovecheskij razum: ot suzhdenij k vychisleniyam / Per. s angl. M.: Radio i svyaz', 1982. 368 s.
4. Gorohov V. G., Deker M. Tekhnologicheskie riski kak social'naya problema pri razrabotke i vnedrenii intellektual'nyh avtonomnyh robotov // Global'noe budushchee 2045. Konvergentnye tekhnologii (NBIKS) i transgumanisticheskaya evolyuciya. M., 2013. s. 82–93.
5. Diane S. K. D. Obuchenie i social'naya integraciya avtonomnyh robotov na osnove primeneniya sovremennyh kognitivnyh tekhnologij // Filosofiya nauki i tekhniki. 2018. T. 23. № 2. S. 89–102.
6. Klyueva N. Yu. Vliyanie idej G. Lejbnica na razvitie komp'yuternyh nauk i issledovaniya v oblasti iskusstvennogo intellekta // Vestnik Mosk. Un-ta. Ser. 7. Filosofiya. 2017. № 4. S. 79-92.
7. Nikitina E. A. Problema sub"ekta poznaniya v sovremennoj epistemologii // Perspektivy nauki i obrazovaniya. 2015. № 2. S. 16-24.
8. Razin A. V. Etika iskusstvennogo intellekta // Filosofiya i obshchestvo. 2019. № 1 S. 57–73.
9. Rakitov A. I. Filosofiya, roboty, avtomaty i zrimoe budushchee // Filosofiya i obshchestvo, № 3 2019 35–48.
10. Uolsh T. 2062: vremya mashin / Per. s angl. M.: AST, 2019. 280 s.
11. Boden M. Artificial intelligence in psychology: Interdisciplinary essays. Cambridge (Mass.); L: MIT press, 1988. 188p.
12. Geraci, R. Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality. Oxford University Press, 2010. 248 p.
13. Haugeland J. Artificial intelligence: The very idea. Cambridge (Mass.); L: MIT press. 1981. 287 p.
14. Moor J. Are there decisions computers should never make? // Ethical issues in the use of computers. Belmont, 1985. P. 120-130.
15. Searle J. R. Minds, brains and programs // Artificial intc1ligcncc: The case against. L.; Sydney, 1987. P. 18-40.