

А.М.АЛЕКСЕЕВ-АПРАКСИН

*доктор культурологии, доцент СПбГУ,
профессор СПбГУПТД, профессор HAINNU*

ГЭН БЯО

магистр, аспирант СПбГУ

УДК 008.2

КОНФУЦИАНСТВО И ЭТИКА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА¹

В статье исследуются этические проблемы, связанные с развитием искусственного интеллекта с точки зрения конфуцианства. Искусственный интеллект (ИИ) является, на сегодняшний день, самой горячей темой в мире, и связанные с этим этические вопросы также являются основными проблемами, которые обсуждаются в области философии и культурологии. Несмотря на свою древность, в конфуцианском учении имеются суждения, которые можно интерпретировать как концептуальную поддержку и формулировку ориентиров этики искусственного интеллекта. Во-первых, через определение «человека» в конфуцианстве предлагается возможность определения социального статуса искусственного интеллекта. Во-вторых, в конфуцианстве сформулированы позиции по отношению к дискурсу о стандартизации этики для искусственного интеллекта. В-третьих, в конфуцианском учении содержатся основания для исследования гибридного существования, сочетающего искусственный интеллект и человеческое тело. Делается вывод о том, что вклад конфуцианства в этой области заключается в его косвенно-нормативном принципе, в рамках идеи «нужно исправлять людей при помощи людей, пока они не изменятся 以人治人, 改而止».

Ключевые слова: искусственный интеллект, конфуцианство, самосознание, этика

A.M. ALEKSEEV-APRAKSIN

¹ The reported study was funded by Russian Foundation for Basic Research (RFBR) according to the research project No. 18-011-00977.

*Dr.Habil., Assoc. Professor SPbSU,
Professor SPSUITD, Professor HAINNU*

GENG BIAO

Postgraduate student of SPbSU

CONFUCIANISM AND THE ETHICS OF ARTIFICIAL INTELLIGENCE

The article examines the ethical problems associated with the development of artificial intelligence from the point of view of Confucianism. Artificial intelligence (AI) is by far the hottest topic in the world, and related ethical issues are also the main issues discussed in the field of philosophy and cultural studies. Despite its antiquity, there are judgments in Confucian teaching that can be interpreted as a conceptual support and formulation of guidelines for the ethics of artificial intelligence. First, through the definition of "human" in Confucianism, it is possible to determine the social status of artificial intelligence. Second, Confucianism has formulated positions in relation to the discourse on the standardization of ethics for artificial intelligence. Third, the Confucian teachings contain grounds for exploring a hybrid existence that combines artificial intelligence and the human body. It is concluded that the contribution of Confucianism in this area lies in its indirect-normative principle, in the framework of the idea «you need to correct people with the help of people until they change 以人治人，改而止».

Keywords: artificial intelligence, Confucianism, self-awareness, ethics

От каменного века до наших дней наступление нового исторического этапа развития человечества традиционно маркируются посредством выявления «технологических революций». Цифровизация, интернет и компьютерные технологии сегодня трактуются как признак технологической революции нашего периода. На повестке дня развитие искусственного интеллекта. Помимо технических проблем, рассмотрение данной темы требует культурологического анализа. Один из поводов поговорить об этических проблемах искусственного интеллекта – скандально известный робот София¹, получивший гражданство Саудовской Аравии. Публичное заверение Софии, что в будущем она будет заботиться о людях и построит для них удобные «зоопарки», привел к ожесточенным спорам в академических кругах о человеческих перспективах в связи с дальнейшим развитием искус-

¹ Человекоподобный робот София в виде женщины, разработанный гонконгской компанией Hanson Robotics. 25 октября 2017 года на саммите «Инвестиционная инициатива будущего» в Эр-Рияде ей было предоставлено подданство Саудовской Аравии, и, таким образом, София стала первым роботом, который когда-либо имел гражданство

ственного интеллекта. Изучая конфуцианство, мы обнаружили возможность рассмотреть некоторые из имеющихся в нем идей, которые могут быть полезны для решения проблем взаимодействий человека и искусственного интеллекта, при помощи неевропейского взгляда на основные этические вопросы и подходы.

Проблемы определения статуса искусственного интеллекта.

Развитие искусственного интеллекта отличается от других известных нам интеллектуальных систем тем, что он строится по подобию человеческого. И судя по прогнозам разработчиков, весьма скоро он, опираясь на большие данные о жизни людей, может получить «самосознание» и стать подобным людям. Можно ли их определить как «людей» и могут ли они получить права, которыми живое существо должно обладать? Это сложный вопрос. Многих беспокоит, что искусственный интеллект может подорвать существующую социальную систему и образ жизни. Преимущества и недостатки его существования для человека, в настоящее время, неизвестны.

С точки зрения конфуцианства, «человек» представляет собой смесь биологического тела и социализированной идеальной личности. Китайский философ Мэн-цзы проводит различие между людьми и животными следующим образом:

«人之所以异于禽兽者几希，庶民去之，君子存之。舜明于庶物，察于人伦，由仁义行，非行仁义也». В книге «Мэн-цзы» читаем: между людьми и животными очень мало различий, основными являются «гуманность и справедливость». Благородный человек хранит их в своем сердце, а простые и низкие люди теряют их и могут дойти до животного состояния. Китайский философ Сюнь-цзы высказал свое мнение по этому поводу: он твердо верил, что «человек с рождения злой». Это противоречит мнению Мэн-цзы, о том, что «доброта изначально заложена в человеческой природе». Но к вопросу о том, какой человек действительно «гуманный» или «благородный», они оба придерживаются одинаковых взглядов. Спор между ними сводится лишь к тому, какой аспект наиболее важен при достижении такой цели – природный или приобретенный (вдохновение или образование). Другими словами, в глазах конфуцианцев причина, по которой человека можно назвать «Человеком», является то, что у людей существуют такие понятия, как «мораль и этика», а у животных их нет. Чтобы стать настоящим «человеком», есть лишь одна возможность: необходимо сдерживать свои животные инстинкты и требовать от себя «гуманности и справедливости» возвращая в себе «нравственное сознание».

Помимо человеческой сущности, конфуцианство также говорит о другом важном качестве, которым нужно обладать – социальность. В книге

«Да сюэ» есть замечательная фраза: «необходимо добиваться самосовершенствования, в порядке содержать семью, управлять государством и нести в Поднебесной мир». Упомянутые выше «гуманность и справедливость» – это уникальная природа человека, но этого недостаточно. Человек должен быть социален и стремиться к совершенствованию. Это представляет смысл человеческого существования для этого мира, который отличается от смысла существования животных и растений. Как известно, в китайской культуре, основанной на конфуцианстве, интересы семьи, нации и страны всегда выше личных интересов. Западные мыслители полагают, что человеческое поведение имеет характеристики «рационально мыслящего человека» и он всегда будет делать выбор, который выгоден для него самого, то есть в личных интересах. Но очевидно, что конфуцианство так не думает. Оно считает, что, столкнувшись с семейными и национальными интересами, люди не будут делать выбор в пользу себя, а будут делать выбор, который будет полезен для всего человечества, исходя из этических требований, которые суть его «совесть».

Существование совести основано на самосознании. Конечно, у животных и даже у некоторых растений тоже есть «сознание», но они не могут задавать такие вопросы, как «кто я», а также размышлять о сути и смысле жизни. Но искусственный интеллект может развить подобное «самосознание» в будущем. Исходя из того, что если вышеуказанные условия будут выполнены, то следует определить искусственный интеллект, как «человека» или условно называть «механическим человеком». Так, в отличие от Запада, который определяет людей по биологическим признакам, конфуцианство определяет людей не по их естественным признакам, а по социокультурным, и наличию у них этических качеств. Данное противоречие между конфуцианством и «современным западным обществом», может быть новой возможностью для определения вектора развития искусственного интеллекта. Но даже в этом случае, когда искусственный интеллект определяется как «человек», мы также сталкиваемся с большой проблемой – как обращаться с такими «людьми»? Если рассматривать его как «человека», который подчиняется и безоговорочно выполняет инструкции, становится ли он «рабом» другой модели? В конце концов, если мы думаем, что они «люди», тогда они должны иметь ту же власть, что и мы, иначе будет новый тип «общества рабов». Очевидно, что это не изначальное намерение человечества создать искусственный интеллект. Хотя?

Человек и искусственный интеллект: конкуренция и интеграция

Если искусственный интеллект не будет обладать «самосознанием» или «интеллектом», подобно человеческому, то теоретической поддержки

определения искусственного интеллекта, как «человека», больше не будет. Но это также означает, что мы активно ограничиваем развитие искусственного интеллекта категорией машин и «глупых роботов». И если мы хотим, чтобы искусственный интеллект сам нашел лучший ответ на данную проблему, то его коэффициент умственного развития (IQ) должен быть выше и лучше. Следовательно, вопрос «как его контролировать» будет одной из самых основных между людьми и искусственным интеллектом. Люди научились готовить еду на огне и убивать врагов оружием. Условия, по которой эти «инструменты» могут использоваться разумно и не причинят вреда пользователю – при соблюдении правил использования. Это означает, что люди могут эффективно контролировать их (с помощью правил использования). Отношения между искусственным интеллектом и людьми похожи на ребенка, играющего с бомбой. Невежественный ребенок не может понять и контролировать силу этой «игрушки». Еще больше пугает то, что, когда детям угрожает опасность, они могут обращаться за помощью к взрослым. Но когда люди сталкиваются с «бомбой» искусственного интеллекта, обратиться за помощью не к кому. Шведский философ Ник Бостром (Nick Bostrom) [2] предположил, что человек и искусственный интеллект преследуют пять основных целей:

1. Самосохранение;
2. Согласованность содержания и целей;
3. Когнитивное улучшение;
4. Техническое улучшение;
5. Приобретение ресурсов.

Среди них он считает, что пятый пункт является главной причиной его создания и одновременно опасности со стороны искусственного интеллекта по отношению к людям. Он считает, что ценность ресурса зависит от того, для чего его можно использовать. И это тоже зависит от технологии, которую можно использовать [2]. Если современные технологии позволяют, то время, пространство и так далее, также могут быть использованы в качестве бесплатных ресурсов, такие как ветер, дождь, гром и молния, для достижения почти любой цели. Когда искусственный интеллект обладает самосознанием аналогичным человеческому, то он будет использовать определенное количество ресурсов, чтобы увеличить скорость работы, улучшить резервное копирование системы и системы защиты для выживания. В этом процессе люди будут рассматриваться, как конкуренты. И, если люди не смогут эффективно контролировать искусственный интеллект, его существование становится для них потенциальной угрозой. Получается, что контролировать искусственный интеллект все же важно.

Внедрение в программу развития ИИ изначального этического кода, подобного ДНК, в определенной степени может решить эту проблему. Во всем мире известна знаменитая теория Азимова, но факты доказали нам, что эта теория имеет серьезный недостаток. Когда перед искусственным интеллектом стоит дилемма, робот не может сделать выбор, соблюдая эти правила. Конфуцианство предоставляет несколько иную «идеальную» этическую модель. В трактате «Чжун юн» написано, что «необходимо исправлять людей при помощи людей, пока они не изменятся 以人治人, 改而». Это означает, что мы не должны относиться к другим с точки зрения первого лица, нужно стараться думать о проблеме с точки зрения другого человека. Для искусственного интеллекта этические правила должны быть рассматриваемы не с точки зрения «что это такое» и «что он должен делать», а с точки зрения «кем и чем он не является».

Таким образом внедренные этические принципы, не предполагают, что искусственный интеллект будет подчиняться всем приказам людей и не будет контролироваться ими на 100%. Вместо этого можно использовать косвенные и нормативные методы, такие как руководство и эвристика, чтобы разрешить искусственному интеллекту в полной мере использовать свои когнитивные преимущества и предоставить людям лучший выбор. Хотя косвенные меры регулирования не могут полностью решить эту проблему (в конце концов, люди тоже не могут решить эту проблему), по крайней мере, они предоставят искусственному интеллекту большую свободу выбора, чтобы он мог полагаться на свои собственные преимущества для принятия наилучшего решения.

Что касается отношений между искусственным интеллектом и людьми, есть еще один важный момент – возможность сочетание человека и искусственного интеллекта. Если в соответствии с конфуцианством мы определяем кого-то как «человеческое существо» на основании того, что он обладает «самосознанием» и присущим человеку моралью и этикой. Что в этом случае препятствует интегрировать искусственный интеллект в тело, трансплантировать его реальным людям, как например, кардиостимулятор или протез. Конечно, современный кардиостимулятор – это всего лишь инструмент, обычно используемый для лечения сердечных заболеваний, но если он обладает самосознанием, то в определенной степени он может определять жизнь и смерть хозяина. Фантазия, определяющая направление технического прогресса, уже давно нарисовала будущее, в котором люди будут использовать технологии, чтобы изменить свой IQ или телосложение и даже вживлять интеллектуальные чипы в свой мозг, чтобы быстрее получать знания и изменять память.

Сочетание человека и машины – неизбежное явление, которого трудно избежать после развития человеческой цивилизации до определенного этапа. Конечно, в эпоху бурного развития науки и технологий человеческое тело также подвергнется революционной трансформации. С одной стороны, это позволит преодолеть физические ограничения традиционного тела. С другой стороны, это может улучшить функции и навыки организма. Американский ученый Кэтрин Хеллер (N. Katherine Hayles) считает, что тело это не что иное, как «протез», которым люди должны научиться управлять. Поэтому, использование его как внешнего оборудования или замены должно стать непрерывным процессом. В этом процессе нет существенной разницы между структурой взаимоотношений человека и машины и биологической организацией [3].

Очевидно, если рассматривать данный вопрос с этой точки зрения, то это захватывающая физическая революция, которая предоставила бы людям безграничные возможности. Следовательно, в нем есть историческая рациональность, основанная на сочетании человека и искусственного интеллекта. Поэтому, мы должны с радостью принять это направление развития. Но стоит отметить, что при такой модели развития существует управляемая идеологическая проблема. В настоящее время, мы твердо верим, что машины и искусственный интеллект должны обладать своего рода управляемостью, независимо от того, как они развиваются. Однако, люди игнорируют «антиуправляемость» искусственного интеллекта. Быстрое развитие этого вида «антиконтроля» побудит человеческое тело в значительной степени полагаться на машину. Это приведет к тому, машина будет властвовать над природой тела. Это как водитель, который привык пользоваться навигацией. Он не может самостоятельно добраться до пункта назначения без помощи навигации.

Кроме того, исторически, люди в основном унаследовали или получили классовый статус в результате наследственности или получения высококачественного образования. А развитие искусственного интеллекта может позволить людям обладать физическими или интеллектуальными способностями, которыми обладают с помощью технических средств. Благодаря этому, людям состоятельным и ученым, имеющим знания, права и деньги, будет легче получить такие технологические преимущества, тем самым закрепив свои преимущества в социальной иерархии. Это может привести к большему дисбалансу справедливости, чем во всех ранее существовавших социальных формациях. Концепции конфуцианства, такие как «всякое дело ничтожно, и только знание – путь к высокому 万般皆下品, 唯有读书高» и «решимость человека преодолевает даже небо 人定胜天», останутся только на бумаге. В то же время, это при-

ведет к потере ценности социальной интеграции через просвещение. Педагог Конфуций уделял большое внимание роли просвещения в построении общественного строя и считал, что люди должны понимать друг друга через образование и устанавливать общие базовые ценности. Но, как мы и предполагали, самое ожидаемое в отношении искусственного интеллекта – это то, что они могут легко получить все знания человечества, но вопрос стоит в следующем, будет ли такой же прогресс с точки зрения ценностей и этики, пока неизвестно.

Следует отметить, что для конфуцианства социальный статус человека зависит от его собственных усилий в борьбе за так называемое «станьте идеальным человеком через обучение». Оно подчеркивало изначальное равенство между людьми и наличие социальной роли, определяемой со дня рождения. Такое ролевое позиционирование сначала определяется на основе норм отношений между людьми, а затем распространяется на политику, общество и род занятий. Конечно, между разными ролями есть различия, которые естественным образом формируются и определяются их собственными ролями. Например, отец и сын, муж и жена, учителя и ученики и т. д., между ними будут различия в правах и обязанностях. Но социальный статус человека не определяется его ролью, за все другие социкультурные статусы, кроме императора, нужно бороться. Иными словами, с точки зрения конфуцианства, возникнет противостояние через наследование прав и богатства на получение приоритета к интеллектуальному ресурсу и социальному статусу.

Выводы

С древних времен различие между «естественными» и «искусственными» объектами служило основой человеческого познания мира. С развитием современных технологий эта дихотомия была восполнена различием между «организмами» и «не организмами» и далее была интегрирована в понимание «живого» и «неживого». Такое мировосприятие отражает настойчивость в самосознании людей, но также ограничивает человеческое познание. В настоящее время, когда люди стоят на пороге создания искусственного интеллекта, более глубокие размышления и анализ говорят нам, что эта граница может быть не такой четкой и твердой, как предполагалось. Что касается искусственного интеллекта, то с точки зрения конфуцианства нам необходимо сформировать открытое отношение к его приятию, чтобы понять и адаптироваться к новой реальности.

В современной конструкции этики искусственного интеллекта, конфуцианство предоставляет новый ракурс в понимании места и роли искусственного интеллекта в будущем обществе. Его вклад не в добавлении конкретных ценностей, а в необходимости развернуть поток деятельности

от движения к «себе» к движению «от себя» и «для других», что также совпадает и с буддийской установкой. Известные модели, которые прямо оговаривают этику искусственного интеллекта, такие как «законы о роботах Азимова», выявили свою ограниченность и неприменимость, но это не тупик. Конфуцианство и другие восточные учения предлагают иную идею для формулирования этики. Принцип косвенного регулирования, защищаемый конфуцианством, может быть важным вкладом конфуцианства в этику искусственного интеллекта.

Проблемы искусственного интеллекта в свете конфуцианского учения напоминают нам о рисках, с которыми мы можем столкнуться в будущем при развитии искусственного интеллекта, в том числе при гипотетической возможности вхождения его в стадию сингулярности и порождения радикальных изменений человеческой культуры в глобальном масштабе. Это также ставит перед нами вопрос: является ли мораль и аморальность искусственного интеллекта самой машины или общей проблемой для всех живых существ? Эти и многие другие вопросы и проблемы, несомненно, нуждаются в продолжении уже давно начатых дискуссий и исследований.

ЛИТЕРАТУРА

1. Asada M., MacDorman K. F., Kuniyoshi Y. Cognitive Developmental Robotics as a New Paradigm for the Design of Humanoid Robots // Robotics Autonomous Systems. 2001. No. 37. P. 185-193.
2. Bostrom, N. Superintelligence: Paths, Dangers, Strategies[M]. Oxford: Oxford University press, 2014 P. 109-113
3. N. Katherine Hayles. How We Became Posthuman [M]. Chicago: The University of Chicago Press, 1999. 350 p.

TRANSLIT

1. Asada M., MacDorman K. F., Kuniyoshi Y. Cognitive Developmental Robotics as a New Paradigm for the Design of Humanoid Robots // Robotics Autonomous Systems. 2001. No. 37. P. 185-193.
2. Bostrom, N. Superintelligence: Paths, Dangers, Strategies[M]. Oxford: Oxford University press, 2014 P. 109-113
3. N. Katherine Hayles. How We Became Posthuman [M]. Chicago: The University of Chicago Press, 1999. 350 p.