DOI: 10.30727/0235-1188-2022-65-1-124-143
Оригинальная исследовательская статья
Original research article

Ценностные ориентации технологий искусственного интеллекта в США и Китае: философский анализ*

А.М. Савельев
Институт философии РАН, Москва, Россия,
Аналитический центр при Правительстве Российской Федерации, Москва, Россия

Д.А. Журенков
Институт философии РАН, Москва, Россия,
Всероссийский научно-исследовательский институт «Центр», Москва, Россия

А.Е. Пойкин
Всероссийский научно-исследовательский институт «Центр», Москва, Россия

Аннотация

Искусственный интеллект (ИИ) в XXI веке уже перестал восприниматься как исключительно технологическое явление, все больше и больше приобретая черты социального и гуманитарного феномена, развивающегося в сложном контексте культурных, ценностных, мировоззренческих и морально-этических сторон жизни человека. Влияние технологий, связанных с ИИ, на современное общество пока еще сложно оценить в полной мере, что не мешает исследователям, энтузиастам и политическим лидерам делать попытки определить ценностные рамки, которые обеспечат использование ИИ в интересах развития общества. С ростом интереса к ИИ все больше технологически развитых стран мира создают свои стратегии по развитию и использованию этого технологического чуда XXI века. Эти новаторские документы часто кажутся расплывчатыми и неопределенными, но тем не менее они позволяют оценить, как политические и научно-технологические элиты этих стран видят ценностные ориентации развития технологий ИИ как на национальном, так и на международном уровне. В статье с позиций постнеклассической научной рациональности проведен философский анализ ценностных ориентаций развития

* Работа поддержана Российским научным фондом (РНФ), грант № 21-18-00184 «Социогуманитарные основания критериев оценки инноваций, использующих цифровые технологии и искусственный интеллект».

технологий искусственного интеллекта в США и Китае – современных научно-технических лидеров в этой сфере – на основе стратегических документов, определяющих развитие и применение ИИ в этих странах. Авторы статьи делают вывод, что на современном постнеклассическом этапе развития науки ценностная компонента не просто является одной из интегральных компонент научно-технической деятельности, но может носить определяющий характер в определении целей и задач развития высоких технологий на государственном уровне.

Ключевые слова: философия искусственного интеллекта, этика, социальная философия, постнеклассическая научная рациональность, ценности, философия техники, научно-технический прогресс.

Савельев Антон Максимович – аспирант Института философии РАН, ведущий советник Аналитического центра при Правительстве Российской Федерации.
anton.saveliev@gmail.com
https://orcid.org/0000-0002-3687-9147

Журенков Денис Александрович – аспирант Института философии РИН, руководитель Центра диверсификации организаций ОПК ФГУП «ВНИИ «Центр»
dzhurenkoff@mail.ru
https://orcid.org/0000-0002-3968-5815

Пойкин Артем Евгеньевич – заместитель начальника отдела Центра диверсификации организаций ОПК ФГУП «ВНИИ «Центр»
art-tem1@mail.ru
https://orcid.org/0000-0001-6185-8922

Для цитирования: *Савельев А.М., Журенков Д.А., Пойкин А.Е.* Ценностные ориентации технологий искусственного интеллекта в США и Китае: философский анализ // Философские науки. 2022. Т. 65. № 1. С. 124–143. DOI: 10.30727/0235-1188-2022-65-1-124-143

Value Orientations of Artificial Intelligence Technologies in USA and China: A Philosophical Analysis

A.M. Saveliev
Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia, Analytical Center under the Government of the Russian Federation, Moscow, Russia

* The work was supported by the Russian Science Foundation, grant no. 21-18-00184 “Social and humanitarian foundations for evaluation criteria for innovations based on digital technologies and artificial intelligence.”

D.A. Zhurenkov
Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia,
All-Russian Scientific Research Institute "Center", Moscow, Russia

A.E. Poikin
All-Russian Scientific Research Institute "Center", Moscow, Russia

Abstract

Artificial Intelligence (AI) in the 21st century is no longer perceived as a purely technological phenomenon, more and more becoming a social and humanitarian phenomenon that develops in a complex context of cultural, value, philosophical, and ethical aspects of human life. The impact of AI-related technologies on contemporary society is still difficult to assess fully, which does not prevent enthusiastic researchers and political leaders from attempting to define a value framework that will ensure the use of AI for societal development. As interest in AI grows, more and more technologically advanced countries in the world are creating their own strategies for the development and use of this technological marvel of the 21st century. These pioneering documents often seem vague and indefinite, but nevertheless they allow us to assess how the political and scientific and technological elites of these countries see the value orientations of AI technology development, both nationally and internationally. The article presents a philosophical analysis of the value orientations of AI technology development in the USA and China – modern scientific and technological leaders in this field – on the basis of strategic documents defining the development and application of AI in these countries from the position of post-non-classical scientific rationality. The authors of the article conclude that, at the contemporary post-non-classical stage of science development, the value component is not only one of the integral components of scientific and technological activities but may be decisive in determining the goals and objectives of high-tech development at the state level.

Keywords: philosophy of artificial intelligence, ethics, social philosophy, post-non-classical scientific rationality, values, philosophy of technology, scientific and technological progress.

Anton M. Saveliev – postgraduate student, Institute of Philosophy, Russian Academy of Sciences; Leading Adviser, Analytical Center under the Government of the Russian Federation.
anton.saveliev@gmail.com
https://orcid.org/0000-0002-3687-9147

Denis A. Zhurenkov – postgraduate student, Institute of Philosophy, Russian Academy of Sciences; Head of the Center for Diversification of

Defense Industry Organizations, All-Russian Scientific Research Institute "Center."
dzhurenkoff@mail.ru
https://orcid.org/0000-0002-3968-5815

Artem E. Poikin – Deputy Head of Department, Center for Diversification of Defense Industry Organizations, All-Russian Scientific Research Institute "Center."
art-tem1@mail.ru
https://orcid.org/0000-0001-6185-8922

For citation: Savelyev A.M., Zhurenkov D.A., & Poikin A.E. (2022) Value Orientations of Artificial Intelligence Technologies in the USA and China: A Philosophical Analysis. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 65, no. 1, pp. 124–143.
DOI: 10.30727/0235-1188-2022-65-1-124-143

Введение

Современная наука не отказывает искусственному интеллекту (ИИ) в праве иметь философское и ценностное измерение. Более того, ИИ сам зачастую понимается как обоюдоострое оружие прогресса – почти что наравне с технологией расщепления атомного ядра, чей разрушительный потенциал во многом определил контуры политического устройства мира во второй половине XX века. Такие выдающиеся ученые и общественные интеллектуалы, как Стивен Хокинг и Мартин Рис, а также инноватор Илон Маск и исследователь ИИ Стюарт Рассел, весьма красноречиво говорили о разрушительной силе, которой обладает ИИ, упоминая прежде всего о риске полного уничтожения человечества, если технологии сильного ИИ станут неуправляемыми или попадут в неумелые руки. В 2009 году на конференции в Асиломаре ведущие исследователи ИИ [Horvitz, Selman 2009] выразили свою растущую обеспокоенность ценностной и морально-этической стороной развития ИИ, что в итоге привело к подписанию открытого письма[1] и созданию Асиломарских принципов ИИ[2] – набора из

[1] An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence // Future of Life Institute. 2015. – URL: https://futureoflife.org/2015/10/27/ai-open-letter/

[2] Asilomar AI Principles // Future of Life Institute. 2017. – URL: https://futureoflife.org/2017/08/11/ai-principles/

23 руководящих принципов, описывающих ценностную, этическую и общественную проблематику развития ИИ и этические нормы для развития искусственного интеллекта во благо человечества. В свою очередь потенциальная опасность применения ИИ в качестве оружия открыто обсуждалась на ведущей конференции по ИИ – конференции Ассоциации по развитию искусственного интеллекта (*AAAI*) 2015 года – и на семинаре по ИИ и этике в рамках той же конференции. Международное сообщество также разделяет определенное недоверие к ИИ, но в целом согласно с тем, что разработать и формализовать какие-либо четкие ориентиры во избежание серьезных опасностей довольно трудно: технология ИИ слишком сложна [Floridi, Cowls 2022]. Тем не менее данный аспект не мешает технологически развитым странам создавать масштабные программы развития и технологий ИИ, чье социальное и гуманитарное обеспечение подкрепляется соответствующими документами стратегического характера. Зачастую данные стратегии носят крайне расплывчатый и общий характер, однако сам факт их существования, а также наличие в них ценностной и моральной составляющей, позволяет судить об ИИ как о сложном феномене, чьи границы уже давно распространились за рамки научно-технической парадигмы [Архипов, Наумов 2017а; Архипов, Наумов 2017б]. Авторы статьи предприняли попытку провести общий философский анализ стратегических документов, определяющих перспективы развития технологий ИИ в США и Китае на предмет выявления в них тех ключевых ценностных ориентаций, которые, с точки зрения руководства этих стран, должны определить развитие технологий искусственного интеллекта в ближайшем будущем. Выбор этих стран как объектов исследования отнюдь не случаен – американское и китайское руководство рассматривают ИИ не просто как изолированный набор передовых и необходимых технологических решений, но как инструмент комплексного, парадигмального преобразования общественных, экономических, управленческих и в конечном итоге мировоззренческих основ существующего миропорядка. Таким образом, философский анализ ИИ как преобразующего инструмента, провоцирующего парадигмальные сдвиги в обществе, выдвигает новые требования к методологии такого анализа, которая должна учитывать проблематику ИИ как феномена нового постнеклассического этапа развития науки, в условиях которого индивидуальный и коллективный субъект научного познания не

только решает сугубо деятельностные задачи научного поиска, но и осуществляет рефлексию над ценностными основаниями научной деятельности [Степин 2009, 249–295].

Искусственный интеллект на современном этапе его развития с уверенностью стоит рассматривать в контексте постнеклассической научной рациональности (постнеклассики), где он предстает в качестве сложной интегрированной системы [Лекторский 2016], включающей в себя явления и процессы технологического, инженерного, научного, социального, биологического, психологического и, как следствие, ценностного характера, где проблематика неизбежно смещается к парадигме «человек – машина» [Лекторский 2001]. Таким образом, ценностные ориентации развития технологий ИИ в контексте постнеклассики стоит рассматривать с позиций отражения этики, норм, ценностей и традиций коллективного субъекта инновационной деятельности (в контексте настоящего исследования – страны и нации) в подходах, как к самому научному исследованию, так и к предполагаемым его результатам, к внедрению технологий в социальную и культурную, экономическую и политическую ткань общества [Friedler, Scheidegger, Venkatasubramanian 2021].

Искусственный интеллект – этическое и ценностное измерение

Как уже отмечалось ранее – искусственный интеллект в условиях постнеклассики без сомнения является, в т.ч. и мировоззренческим феноменом, заслуживающим всестороннего анализа с позиций философии науки. В этой связи перед современным исследователем стоит задача поиска надлежащих исходных оснований для философско-методологического анализа сложного феномена ИИ. В настоящее время тема ценностного и этического осмысления ИИ как никогда актуальна, что заставляет представителей общественной, политической и научной элиты во всем мире разрабатывать соответствующие теоретические концепции, призванные, так или иначе, урегулировать ценностный и этический аспект разработки и применения технологий ИИ. Плюрализм подходов к этому вопросу, безусловно, обогащает научное знание, но не делает задачу философско-методологического анализа проще. Более того, многообразие ценностно-этических концепций существенно затрудняет выработку общепризнанных норм разработки и использования ИИ. К подобной мысли пришел и профессор философии и информационной этики Оксфордского

университета Л. Флориди (1964) – один из авторитетных исследователей в области современной философии техники, а именно – «философии информационных технологий» [Floridi 2002]. Флориди проанализировал шесть важнейших инициатив, направленных на совершенствование ценностно-этического обеспечения развития ИИ:

• «Асиломарские принципы ИИ», разработанные под эгидой Института будущего жизни в 2017 году;

• «Монреальская декларация ответственного ИИ», разработанная под эгидой Монреальского университета в 2017 году[3];

• Общие принципы, предложенные во втором издании книги «Этически обоснованный дизайн ИИ: взгляд на благополучие человека в автономных и интеллектуальных системах»[4]; этот документ является результатом совместных усилий 250 экспертов под патронажем Института инженеров по электротехнике и электронике (*IEEE*);

• Этические принципы, предложенные в докладе об искусственном интеллекте, робототехнике и «автономных» системах, опубликованном Европейской группой по этике в науке и новых технологиях Европейской комиссии в марте 2018 года;

• «Пять всеобъемлющих принципов для кодекса ИИ», предложенные в докладе Комитета по искусственному интеллекту Палаты лордов Великобритании в 2018 году;

• «Принципы Партнерства в ИИ» многосторонней организации *Partnership on AI*, состоящей из ученых, исследователей, общественных объединений, создающих и использующих технологии ИИ.

Л. Флориди совместно со своим коллегой Д. Коулзом проанализировали все эти документы и обнаружили, что в совокупности они содержат 47 основных принципов того, как ИИ может быть использован с пользой для общества. Согласно концепции, принятой вышеупомянутыми исследователями, все эти принципы имеют высокую степень когерентности с четырьмя основными принципами, обычно используемыми в биоэтике: *«делай благо»* (*beneficence*), *«не навреди»* (*non-maleficence*), *«уважение автономии субъекта»* (*autonomy*) и *«справедливость»* (*justice*) [Floridi, Cowls 2022; Beauchamp, Saghai 2012]. По мнению Фло-

[3] Montreal Declaration for a Responsible Development of Artificial Intelligence. – URL: https://www.montrealdeclaration-responsibleai.com/the-declaration

[4] Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version 2 // IEEE. – URL: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

риди, биоэтика больше всего похожа на информационную этику в вопросах средового взаимодействия субъектов в условиях новой сложной экосистемы современного цифрового общества [Floridi 2008; Floridi 2013; Floridi 2019].

Однако Флориди и Каулз утверждают, что дополнительно необходим новый принцип: принцип «*обоснованности*», который понимается одновременно как «понятность/объяснимость» ИИ для неспециалистов и его подотчетность пользователю [Floridi, Cowls 2022]. Приняв данный исходный посыл, основанный на работах Л. Флориди и Д. Коулза, можно сконструировать общую матрицу ценностных оснований технологий искусственного интеллекта и их онтологических особенностей (см. табл. 1).

Принцип ценностного основания	**Онтологические особенности**
«делай благо»	Содействие благополучию индивидуальных и коллективных субъектов, сохранение достоинства и поддержание гармонии в масштабе всего мира.
«не навреди»	Неприкосновенность жизни, безопасность и «осторожность в возможностях».
«уважение автономии субъекта»	Право субъекта принимать решение, право полностью или частично делегировать принятие решения иным субъектам.
«справедливость»	Содействие процветанию, солидарности, предотвращение несправедливости.
«обоснованность»	Обеспечение работы других принципов через понятность и подотчетность.

Табл. 1. Общая матрица ценностных оснований технологий искусственного интеллекта

1. ***«Делай благо»***: содействие благополучию, сохранение достоинства и поддержание гармонии в масштабе всего мира. Этот принцип подчеркивает основополагающее значение содействия благополучию людей в планетарном масштабе с помощью технологий ИИ, дальнейшего процветания человечества и сохранения социально ответственной окружающей среды для будущих поколений.

2. ***«Не навреди»***: неприкосновенность жизни, безопасность и «осторожность в возможностях» [Floridi, Cowls 2019]. Этот принцип предостерегает от различных негативных последствий

чрезмерного или неправильного использования технологий ИИ, особенно когда речь идет о неприкосновенности личной жизни и военном применении ИИ. При этом пока не ясно, как себя проявят угрозы, связанные с ИИ в будущем: будет ли речь идти преимущественно о неправомерном использовании ИИ самими людьми, или же опасность будет исходить от самих технологий как таковых.

3. ***«Уважение автономии субъекта»***: принимать решения и делегировать принятие решений. «Когда мы используем ИИ и его "умные" возможности, мы добровольно уступаем часть своих полномочий по принятию решений технологическим артефактам – искусственным агентам», – утверждает Флориди. Этот важный принцип говорит о балансе между правом принятия решений, которое люди сохраняют за собой, и правом, которое они делегируют искусственным цифровым агентам. Люди должны сохранять за собой право решать, как поступать: пользоваться свободой выбора там, где это необходимо, и уступать ее в тех случаях, когда на то есть веские причины, утверждает Флориди.

4. ***«Справедливость»***: содействие процветанию, солидарности, предотвращение несправедливости. Хотя справедливость может показаться довольно широким понятием, все энтузиасты ИИ и мыслители согласны с тем, что устранение несправедливой дискриминации, равно как и необходимость всеобщего процветания, должны стоять в центре использования технологий ИИ.

5. ***«Объяснимость/обоснованность»***: обеспечение работы других принципов через понятность и подотчетность. Проще говоря, этот принцип должен ответить на вопрос: является ли человечество «пациентом», получающим «лечение» в виде «горькой пилюли искусственного интеллекта», «врачом», назначающим его, или возможно, и тем, и другим. Таким образом, принцип объяснимости должен включать в себя понимание того, как работает ИИ, равно как и понимание меры ответственности тех, кто работает с ИИ [Floridi, Cowls 2022].

Данная матрица ценностных оснований технологий искусственного интеллекта, безусловно, не является бесспорной – она имеет довольно общий и размытый характер. Однако эти недостатки оборачиваются достоинствами в рамках наших исследовательских задач: стратегические документы, посвященные применению технологий ИИ в государственном масштабе, сами по себе носят

расплывчатый и общий характер. Таким образом, данная матрица может быть временно принята в качестве потенциально приемлемого инструмента оценки того, как ценностные и этические ориентации подразумеваются и реализуются в стратегиях развития ИИ. Чтобы немного облегчить эту задачу, авторы хотели бы предложить следующую визуализацию выбранного инструмента оценки (см. табл. 2).

Оценка	**Шкала основных принципов ценностной ориентации**	**Шкала принципа объяснимости**
0	принцип не упомянут вообще	не упомянуты поддерживающие механизмы и отсутствует дальнейшее представление
1	принцип упомянут хотя бы один раз в документе	поддерживающие механизмы упомянуты, но носят чисто декларативный характер
2	принцип раскрыт, по крайней мере, в одной из стратегических целей документа	поддерживающие механизмы подробно изложены и подкреплены юридическими инициативами
3	принцип представлен в многочисленных стратегических целях документа	поддерживающие механизмы четко прописаны, подкреплены законодательными инициативами и имеют соответствующие программы, поддерживаемые государством

Табл. 2. Шкала ценностных ориентаций ИИ в анализируемых документах

В рамках данной статьи оценка представленности тех или иных ценностных оснований в анализируемых документах будет проводиться по двум шкалам – шкале основных принципов ценностной ориентации (*«делай благо», «не навреди», «автономия», «справедливость»*) и шкале принципа *«обоснованности»*. Окончательный вердикт будет вынесен на основе простого семантического анализа, в котором выбранные выше единицы оценки могут встретить или не встретить свое семантическое представление в анализируемом тексте. При этом авторы данной статьи не ставят перед собой задачу определить точную частоту таких встреч. Вместо этого мы попытаются исследовать базовую семантическую корреляцию между принципами, упомянутыми в матрице ценностных оснований, и положениями анализируемых документов.

Данный методологический подход может показаться слишком расплывчатым, но авторы опасаются, что любой другой более специализированный подход может оказаться слишком узким для документов, которые кажутся расплывчатыми как по букве, так и по духу.

Ценностные ориентации технологий ИИ в США

Отличительной особенностью США является отсутствие как системного правового регулирования в сфере развития ИИ, так и общенациональной государственной стратегии в этой сфере. Ближе всего к национальной стратегии в области искусственного интеллекта Соединенных Штатов по своей сути соответствует «Исполнительный приказ по ИИ», подписанный президентом Д. Трампом 11 февраля 2019 года[5]. В данном документе провозглашена т.н. «американская инициатива по ИИ» – рамочная национальная стратегия Соединенных Штатов в области искусственного интеллекта. Эта стратегия предусматривает согласованные усилия по продвижению и защите американских технологий и инноваций в области ИИ в рамках пяти приоритетов: 1) устойчивое инвестирование в НИОКР в области ИИ, 2) использование средств федерального правительства для развития ИИ, 3) устранение барьеров на пути инноваций в области ИИ, 4) расширение возможностей американских работников с помощью образования, ориентированного на ИИ, и возможностей обучения, и 5) содействие созданию международной среды, поддерживающей американские инновации в области ИИ и их ответственное использование.

«Исполнительный приказ по ИИ» во многом дополняется приоритетами Национального стратегического плана НИОКР в области ИИ (*National Artificial Intelligence Research and Development Strategic Plan*) (далее – Стратегический план):

• осуществлять долгосрочные инвестиции в исследования ИИ, в приоритетном порядке инвестировать в следующее поколение ИИ, которые будут способствовать открытиям и проницательности и позволят Соединенным Штатам оставаться мировым лидером в области ИИ;

---

[5] Maintaining American Leadership in Artificial Intelligence // Executive Office of the President. E.O. 13859. February 11, 2019. – URL: https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence

• разработать эффективные методы взаимодействия человека и ИИ, создать системы ИИ, которые эффективно дополняют и расширяют человеческие возможности;

• учитывать этические, правовые и общественные последствия применения ИИ; разработать системы ИИ, которые способны решать этические, правовые и общественные проблемы с помощью технологических инструментов;

• обеспечить безопасность и надежность систем ИИ – они должны быть надежными, безотказными, безопасными и заслуживающими доверия;

• создать общедоступные массивы данных и среды для обучения и тестирования ИИ; разработать и обеспечить доступ к высококачественным базам данных и средам обучения ИИ, а также к ресурсам для тестирования технологий ИИ и подготовки профильных специалистов;

• разработать широкий спектр методов оценки ИИ, включая технические стандарты и контрольные показатели;

• учитывать и удовлетворять государственные и общественные потребности в кадрах для НИОКР в области ИИ, повысить эффективность подготовки кадров в сфере НИОКР в области ИИ;

• повысить эффективность государственно-частных партнерств для ускорения прогресса в области ИИ, продвигать возможности для постоянных инвестиций в НИОКР в области ИИ и перехода достижений в практические возможности в сотрудничестве с научными кругами, промышленностью, международными партнерами и другими нефедеральными структурами.

Рассмотрим то, как ценностные основания ИИ трактуются в данном документе.

1. ***«Делай благо»***, – пожалуй, является наиболее ярко выраженным принципом, поскольку все восемь стратегических приоритетов упомянуты как приносящие потенциальную пользу «почти всем аспектам общества, включая экономику, здравоохранение, безопасность, право, транспорт и даже саму технологию». Преимущества для общества от всестороннего внедрения ИИ – ускорение восстановления нормальной жизни людей после стихийных бедствий, улучшение медицинской диагностики (например, в выявлении особо опасных видов рака), появление новых рабочих мест на рынке и т.д. По базовой шкале этот принцип представлен на оценку **«3»**. Тем не менее Стратегический план не предусматривает правовых инициатив и планов, которые могли бы оказать

поддержку реализации принципа благодеяния. Таким образом, по шкале обоснованности можно выставить только оценку **«1,5».**

2. ***«Не навреди»*** – косвенно упоминается в одной из целей Стратегического плана: «осознать и подробно рассмотреть те этические, правовые и общественные последствия, которые влечет за собой внедрение ИИ». Американские стратегические документы применительно к этой цели призывают к этичному и безопасному внедрению ИИ для предотвращения любого возможного вреда людям. Кроме того, «не навреди» косвенно подразумевается в другой цели этого документа: «Обеспечить безопасность и надежность систем ИИ». Утверждается, что ИИ должен быть «безопасным по [изначальному] замыслу», где безопасность обеспечивается на протяжении всего жизненного цикла технологии ИИ. По базовой шкале мы можем присвоить оценку **«2,5»**, но по шкале обоснованности возможна только оценка **«1»**, поскольку все средства, поддерживающие данный ценный ценностный принцип, носят исключительно декларативный характер.

3. ***«Уважение автономии субъектов»*** – представлено в цели «разработать эффективные методы взаимодействия человека и ИИ», где говорится, что развертывание систем ИИ должно рассматриваться «как один из вариантов исходного дизайна ИИ, который позволяет его операторам решать, стоит ли им вообще прибегать к запуску системы с ИИ или нет». По основной шкале можно смело присвоить оценку **«1»**, но полное отсутствие вспомогательных средств и дальнейшего юридического представления позволяет выставить только оценку **«0»** по шкале обоснованности.

4. ***«Справедливость»*** – хорошо представлена как в значении справедливого и честного использования ИИ (о чем достаточно пространно говорится в Стратегическом плане с позиции, преимущественно, общественных и государственных субъектов), так и в контексте «социальной справедливости» (цель – «лучше понять национальные потребности в кадрах для НИОКР в области ИИ»). В последнем случае речь идет о необходимости поддержки американских исследователей в области ИИ, а также студентов старших курсов. По основной шкале нами поставлена оценка **«3»**, но по шкале обоснованности – только **«1,5»**: в документе заявлены многочисленные средства поддержки, хотя и без какого-либо юридического и финансового обеспечения.

Ценностные ориентации технологий ИИ в Китае

Китай является одним из самых активных игроков в сфере ИИ, что находит свое отражение не только в технологиях и рыночной

инфраструктуре, но и в сфере законодательного регулирования. В настоящее время Китай обладает самой разветвленной системой законодательных актов и государственных планов, декларирующих приоритеты развития ИИ-технологий, но это многообразие компенсируется достаточно размытым характером этих документов [Савельев, Журенков 2019].

Искусственный интеллект в приоритетах китайского руководства на ближайшие годы (до 2025 года) рассматривается не только в контексте повышения национальной конкурентоспособности и технологической независимости на стратегическом уровне, но и как «средство совершенствования социального управления и развития». Так, «План развития искусственного интеллекта нового поколения», принятый Государственным советом КНР в 2017 году, – основополагающий стратегический документ Китая в области ИИ – утверждает, что искусственный интеллект «способен своевременно распознавать групповые когнитивные и психологические изменения, а также поможет [ответственным лицам] проявлять инициативу в принятии общественно важных решений»[6].

В плане указаны шесть ключевых задач, которые необходимо решить для достижения вышеупомянутых целей, в частности: 1) создание открытой и дружественной инновационной системы технологий ИИ; 2) развитие высокотехнологичной и высокоэффективной интеллектуальной экономики; 3) создание безопасного и благоприятного интеллектуального общества; 4) усиление военно-гражданской интеграции в области ИИ; 5) создание всеобъемлющей, безопасной и эффективной интеллектуальной инфраструктуры; и 6) перспективное планирование нового поколения крупных проектов, связанных с ИИ.

Несмотря на свою детальность и последовательность, китайский стратегический план является, пожалуй, самым трудным документом для анализа на предмет ценностных ориентаций. Он носит декларативный характер, что сильно его роднит с аналогичными стратегическими документами США, однако ориентация на преимущественно экономические и технологические приоритеты ставит в один ряд с бизнес-планами, а не с государственными стратегическими документами.

---

[6] A Next Generation Artificial Intelligence Development Plan // State Council of the People's Republic of China . 2017. – URL: https://d1y8sb8igg2f8e.cloudfront.net/documents/translation-fulltext-8.1.17.pdf

1. ***«Делай благо»*** – принцип, который редко упоминается, но явно подразумевается в китайском стратегическом плане. Практически, все приоритеты и задачи стратегического плана призваны принести стране экономическое процветание КНР на годы вперед, а также установить новое качество жизни для китайских граждан. Это включает в себя применение инновационных технологий ИИ в образовании, здравоохранении, пенсионном обеспечении и других первостепенных нуждах общества. В документе напрямую говорится о том, что человека нужно ставить «на первое место», «следовать общечеловеческим ценностям», «уважать права человека», соблюдать «национальную и региональную этику». В плане есть несколько ключевых приоритетных областей всестороннего внедрения ИИ во благо китайского общества, а именно: интеллектуальное образование, интеллектуальное медицинское обслуживание, интеллектуальные системы здравоохранения и ухода за престарелыми. Тем не менее план не предусматривает никаких средств для реализации этих благородных целей и приоритетов. Нет ни программ, ни проектов, ни законодательных инициатив для их поддержки. По основной шкале можно поставить оценку **«3»**. По шкале обоснованности – только **«1»**.

2. ***«Не навреди»*** – этот ценностный принцип является одним из главных приоритетов стратегического плана: «Разработать законы, правила и этические нормы, способствующие развитию ИИ, что требует безопасного и этичного использования ИИ». С этой целью документ выступает за разработку правовых основ ИИ, а также за исследования в области науки о поведенческих паттернах ИИ. Тем не менее стратегический план никак не раскрывает смысловое содержание этих положений и не предлагает никаких мер по их реализации. По базовой шкале мы можем поставить оценку **«1»**, но по шкале обоснованности возможна только оценка **«0»**.

3. Принцип ***«уважения автономии субъекта»*** – вообще не представлен. В рассматриваемой стратегии не упоминаются какие-либо модели принятия решений и делегирования полномочий для пользователей ИИ. Утверждается, что ИИ может существенно упростить процесс принятия решений и повысить его эффективность, но из документа не ясно, каким образом предполагается этого достичь. Мы присваиваем **«0»** по обеим шкалам.

4. ***«Справедливость»*** представлена в основном в понимании ответственного и законного использования технологий ИИ, о чем гласит один из приоритетов стратегического плана – «разработка

законов, правил и этических норм, способствующих развитию ИИ», – предполагающий создание этической и моральной системы рамочного взаимодействия людей и машин. Отметим, что в плане прописываются и принципы законности при обработке личной информации, защита частной жизни и безопасности личных данных. Таким образом, нельзя делать вывод о том, что частное и индивидуальное в КНР считается менее важным и ценным, чем общественное благо – в официальных документах данной проблематике уделяется достаточно много внимания, вопреки распространённым стереотипам. Есть и национальная специфика – призывается строить «сообщество единой судьбы», «поощрять социальную справедливость» и в разделе «ответственности» отметим «создание механизма подотчетности ИИ». В то же время теоретические контуры этической и моральной системы рамочного взаимодействия людей и машин описаны приблизительно, а способы ее создания и укоренения в обществе вообще не названы. Мы присваиваем оценку **«1»** по основной шкале и оценку **«0»** по шкале обоснованности.

Заключение

Рассмотренные нами стратегические документы США и Китая в области технологий ИИ позволяют заключить, что его ценностное измерение сейчас находится в центре внимания технологически развитых государств, стремящихся упрочить свое положение в мире при помощи новых цифровых технологий. ИИ в Китае и США рассматривается в первую очередь как инструмент ускорения (или даже провокации) благоприятных социальных перемен.

Стоит отметить, что ценностные ориентации обеих анализируемых стран не демонстрируют существенного антагонизма практически по всем вопросам. Разница в выставленных нами оценках вызвана по большей части крайне размытым представлением тех или иных ценностных ориентиров в анализируемых документах.

	«Не навреди»	**«Делай благо»**	**«Не навреди»**	**«Принцип уважения автономии субъекта»**	**«Принцип справедливости»**
США	**3**	**2,5**	**1**	**3**
КНР	**3**	**1**	**0**	**1**

Табл. 3. Сравнительная матрица ценностных ориентаций в технологиях ИИ в США и Китае

Существует мнение, что основное отличие подходов Китая и Запада (США в настоящем исследовании) лежат в различии ценностей – коллективных и индивидуальных. Это закладывает различные этические подходы к разработке ИИ, разное понимание справедливости, безопасности, конфиденциальности. При этом, американские стратегические документы в своих целях в области ИИ зачастую оперируют именно коллективными ценностными ориентациями, такими как «общественное благо», «социальная справедливость», «ответственность перед обществом», уделяя особое внимание защите интересов коллективных субъектов – профессиональных, научных и культурных сообществ, а также социально незащищенных слоев. Не так однозначна и ориентация КНР на «коллективные ценности» – в рамках стратегических инициатив китайского правительства создаются меры регулирования алгоритмов (например, рекламы и продаж), которые обеспечивают безопасность личных данных пользователей, прозрачность использования данных в Интернете и т.д. Что действительно отличает ценностные подходы к технологиям ИИ в Китае, так это включение идеологических концептов в этику разработки и внедрения ИИ. Руководящая идеология прописывается в законодательстве, в алгоритмах как экологическая норма прописано «продвижение социалистических ценностей»[7]. Вторым важным отличием китайского подхода является включение национальной культурной парадигмы в ценностные ориентации технологического развития (что полностью игнорируется в американских стратегических документах). Например, инженеры в области ИИ и высоких технологий в Китае должны руководствоваться такими ценностями как «процветание, демократия, вежливость и гармония», «свобода, равенство, справедливость и верховенство закона», «патриотизм, преданность, честность, дружба». Особо выделяются принципы ответственности, «предшествующей свободе», обязательств, «предшествующих правам», коллективное, «предшествующее индивидуальному», «гармония, предшествующая конфликту»[8].

---

[7] См.: Руководящие заключения по укреплению комплексного управления алгоритмами предоставления информационных услуг в Интернете // Министерство промышленности и информатизации КНР. 2021. – URL: https://wap.miit.gov.cn/xwdt/gxdt/art/2021/art_a8af2b48620b4905b365fc73cd81a1ec.html

[8] См.: Национальный учебник по инженерной этике для выпускников ВУЗов КНР (2019). – URL: http://www.tup.tsinghua.edu.cn/booksCenter/book_06831902.html

Что действительно объединяет оба подхода к ценностным ориентациям ИИ в США и Китае, так это излишне декларативный характер, как самих ценностных ориентаций, так и тех методов, которыми они должны быть воплощены в жизнь. Американские и китайские эксперты в области ИИ в полной мере осознают угрозы, которые несет ИИ, равно как и те потенциальные блага, которые он может дать обществу. Однако у них отсутствует понимание тех конкретных мер и шагов, при помощи которых можно защитить человечество от неправомерного использования ИИ, а зачастую и понимание того, что считать неправомерным использованием данных передовых технологий. Этот печальный факт лишь подчеркивает важность принципа «обоснованности», введенного Л. Флориди. Любые, даже самые замечательные и ясно сформулированные, ценностные ориентации становятся лишь благими пожеланиями без конкретных механизмов их внедрения и обеспечения. В условиях нарастающей конвергенции биологических и цифровых субъектов, вызванной практически повсеместным внедрением технологий ИИ, прежние подходы утрачивают свое значение, ведь ИИ уже давно не ограничен стенами научно-исследовательских институтов и вычислительных центров. Представитель современного общества все больше и больше приобретает черты инфорга – гибридного субъекта, существующего и развивающегося одновременно в реальном и цифровом мире. В этих принципиально новых условиях уже мало создать некий набор тех или иных ценностных и морально-этических правил для пользования технологией во благо [Umpleby, Medvedeva, Lepskiy 2019]. Напротив, речь должна идти о новой области знаний – социогуманитарной эргономике технологий ИИ – дисциплине, которая бы учитывала биологическую, психологическую социальную, когнитивную и духовную природу субъекта, частью жизнедеятельности которого являются технологии ИИ.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Архипов, Наумов 2017а – *Архипов В.В., Наумов В.Б.* Искусственный интеллект и автономные устройства в контексте права: о разработке первого в России закона о робототехнике // Труды СПИИРАН. 2017. № 6 (55). С. 46–62.

Архипов, Наумов 2017б – *Архипов В.В., Наумов В.Б.* О некоторых вопросах теоретических оснований развития законодательства о ро-

бототехнике: аспекты воли и правосубъектности // Закон. 2017. № 5. С. 157–170.

Лекторский 2001 – *Лекторский В.А.* Эпистемология классическая и неклассическая. – М.: Эдиториал УРСС, 2001.

Лепский 2016 – *Лепский В.Е.* Технологии управления в информационных войнах (от классики к постнеклассике) – М.: Когито-Центр, 2016.

Савельев, Журенков 2019 – *Савельев А.М., Журенков Д.А.* Национальные стратегии развития систем искусственного интеллекта: опыт стран-лидеров и ситуация в России // Научный вестник оборонно-промышленного комплекса России. 2019. № 3. С. 75–82.

Степин 2003 – *Степин В.С.* Теоретическое знание. – М.: Прогресс-Традиция, 2003.

Beauchamp, Saghai 2012 – *Beauchamp T.L., Saghai Y.* The Historical Foundations of the Research-Practice Distinction in Bioethics // Theoretical Medicine and Bioethics. 2012. Vol. 33. No. 1. P. 45–56.

Floridi 2002 – *Floridi L.* What is the Philosophy of Information? // Metaphilosophy. Vol. 33. No. 1–2. P. 123–145.

Floridi 2008 – *Floridi L.* Foundations of Information Ethics // The Handbook of Information and Computer Ethics / ed by K.E. Himma, H.T. Tavani. – Hoboken: Wiley, 2008. P. 3–24.

Floridi 2013 – *Floridi L.* The Ethics of Information. – Oxford: Oxford University Press, 2013.

Floridi 2019 – *Floridi L.* What the Near Future of Artificial Intelligence Could Be // Philosophy & Technology. Vol. 32. No. 1. P. 1–16.

Floridi, Cowls 2022 – *Floridi L., Cowls J.* A Unified Framework of Five Principles for AI in Society // Machine Learning and the City: Applications in Architecture and Urban Design. – Hoboken: Wiley, 2022. P. 535–545.

Friedler, Scheidegger, Venkatasubramanian 2021 – *Friedler S.A., Scheidegger C., Venkatasubramanian S.* The (Im)Possibility of Fairness: Different Value Systems Require Different Mechanisms for Fair Decision Making // Communications of the ACM. 2021. Vol. 64. No. 4. P. 136–143.

Umpleby, Medvedeva, Lepskiy 2019 – *Umpleby S.A., Medvedeva T.A., Lepskiy V.* Recent Developments in Cybernetics, from Cognition to Social Systems // Cybernetics and Systems. 2019. Vol. 50. No. 4. P. 367–382.

REFERENCES

Arkhipov V.V. & Naumov V.B. (2017a) Artificial Intelligence and Autonomous Devices in Legal. *SPIIRAS Proceedings.* No. 6, pp. 46–62 (in Russian).

Arkhipov V.V. & Naumov V.B. (2017b) On Some Issues of the Theoretical Basis for the Development. *Zakon.* No. 5, pp. 157–170 (in Russian).

Beauchamp T.L. & Saghai Y. (2012) The Historical Foundations of the Research-Practice Distinction in Bioethics. *Theoretical Medicine and Bioethics.* Vol. 33, no. 1, pp. 45–56.

Floridi L. (2002) What is the Philosophy of Information? *Metaphilosophy.* Vol. 33, no. 1–2, pp. 123–145.

Floridi L. (2008) Foundations of Information Ethics. In: Himma K.E. & Tavani H.T. (Eds.) *The Handbook of Information and Computer Ethics* (pp. 3–24). Hoboken: Wiley.

Floridi L. (2013) *The Ethics of Information.* Oxford: Oxford University Press.

Floridi L. (2019) What the Near Future of Artificial Intelligence Could Be. *Philosophy & Technology.* Vol. 32, no. 1, pp. 1–16.

Floridi L. & Cowls J. (2022) A Unified Framework of Five Principles for AI in Society. In: Carta S. (Ed.) *Machine Learning and the City: Applications in Architecture and Urban Design* (pp. 535–545). Hoboken: Wiley.

Friedler S.A., Scheidegger C., & Venkatasubramanian S. (2021) The (Im) Possibility of Fairness: Different Value Systems Require Different Mechanisms for Fair Decision Making. *Communications of the ACM.* Vol. 64, no. 4, pp. 136–143.

Lektorsky V.A. (2001) *Classical and Nonclassical Epistemology.* Moscow: Editorial URSS (in Russian).

Lepskiy V.E. (2016) *The Technology of Management and Information Wars (From Classical to Post-Non-Classicalneklassike).* Moscow: Kogito-Tsentr (in Russian).

Savelyev A.M. & Zhurenkov D.A. (2019) National Strategies for the Development of Artificial Intelligence Systems: The Experience of the Leading Countries and the Situation in Russia. *Scientific Bulletin of the Military Industrial Complex of Russia.* No. 3, pp. 75–82 (in Russian).

Stepin V.S. (2003) *Theoretical Knowledge*. Moscow: Progress-Traditsiya (in Russian).

Umpleby S.A., Medvedeva T.A., & Lepskiy V.E. (2019) Recent Developments in Cybernetics, from Cognition to Social Systems. *Cybernetics and Systems.* Vol. 50, no. 4, pp. 367–382.