Человек в информационном обществе

СТАТУС ЭТИЧЕСКИХ КОДЕКСОВ В ЭТИКЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Алексеев Александр Петрович

Доктор философских наук, профессор
Московский государственный университет имени М. В. Ломоносова, философский факультет, заведующий кафедрой философии гуманитарных факультетов
Москва, Российская Федерация
aleksandr.alekseev.1957@list.ru

Алексеева Ирина Юрьевна

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

Аннотация

Авторы статьи рассматривают разработку и обсуждение этических кодексов (рекомендаций) как необходимую часть этики искусственного интеллекта (ИИ) – нового междисциплинарного направления, формирующегося в XXI веке. Анализ структуры и содержания сводов этических рекомендаций в сфере ИИ, принятых к настоящему времени международными и национальными организациями, позволяет авторам статьи сделать вывод о предпочтительности сочетания компактных формулировок общеобязательных норм с конкретизацией применительно к родам деятельности людей, вовлеченных в жизненный цикл систем ИИ, с учетом типов систем ИИ и культурных особенностей стран.

Ключевые слова

искусственный интеллект; этика искусственного интеллекта; робоэтика; философия искусственного интеллекта; кодекс этики; рекомендации по этике искусственного интеллекта

Введение

В мае 2023 года представители компании OpenAI, создателя самой популярной нейросети для генерации текстов, заявили о необходимости организации некого аналога Международного агентства по атомной энергии (МАГАТЭ) для контроля за разработками в области искусственного интеллекта [16]. Позже появились сообщения о том, что OpenAI разрабатывает специальную систему ИИ, которая будет оценивать безопасность других систем. В современных дискуссиях по поводу контроля за искусственным интеллектом и учреждения для этого аналога МАГАТЭ высказываются опасения, связанные с перспективой монополизации контролирующих функций большими компаниями и навязыванием неоправданных ограничений остальным участникам рынка высоких технологий. Такие опасения нельзя считать беспочвенными, однако это не отменяет актуальности вопросов о способах регулирования деятельности в сфере искусственного интеллекта и оценке интеллектуальных технических систем с точки зрения их способности «вписываться» в человеческую цивилизацию. Осознание значимости подобных вопросов стимулировало разработку этических кодексов, принимаемых разными организациями и объединениями. Примерами могут служить принятое в 2018 году Экспертной группой Еврокомиссии «Этическое руководство по искусственному интеллекту, заслуживающему доверия» (“Ethics Guidelines for Trustworthy AI”), «Рекомендации по этике искусственного интеллекта» (“Recommendation on the Ethics of Artificial

https://doi.org/10.52605/16059921_2024_04_43

Intelligence") утвержденные в конце 2021 года Генеральной конференцией ЮНЕСКО, а также подписанный примерно в то же время рядом крупных российских компаний, включая Сбер, «Кодекс этики в сфере искусственного интеллекта». При этом предметом дискуссий становится и качество кодексов, и возможности их практического применения, и собственно этическая природа.

1 Ориентиры для людей и для технических систем

Сегодня формируется этика искусственного интеллекта как направление междисциплинарных исследований, в которых участвуют представители технических, физико-математических, биологических и других наук. Конечно же, в эти исследования вовлечена философия, одним из важнейших разделов которой с древних времен является этика. Этика искусственного интеллекта - еще один член в растущем семействе так называемых прикладных этик, которые, начиная с середины XX века, становятся все более заметным явлением интеллектуальной жизни в разных странах [8; 11;14; 13; 17]. Этика ИИ охватывает две большие группы вопросов. Во-первых, это вопросы профессионального поведения и самосознания людей, участвующих в создании и эксплуатации систем и технологий ИИ, в принятии решений относительно разработки и использования таких систем и технологий. Речь идет об ученых, инженерах, менеджерах, других работниках, вовлеченных в соответствующие процессы. Во-вторых, это вопросы «поведения» искусственных интеллектуальных систем (ИИС), интеграции ИИС в человеческое общество, взаимодействия ИИС с человеком и между собой, этический статус ИИС.

Упомянутые группы вопросов взаимозависимы и переплетены друг с другом. Если создаваемая ИИС (например, автономный робот с искусственным интеллектом) должна выполнять функции целеполагания, планирования, выбора и совершения действий, то разработчик в какой-то момент оказывается перед необходимостью аппроксимации моральных ориентиров и норм человеческого поведения, формирования определенного рода «машинной этики». Пример видения ситуации разработчиками ИИС - статья В. Э. Карпова, П. М. Готовцева и Г. В. Ройзензона «К вопросу об этике и системах искусственного интеллекта». Авторы пишут: «При этом выбор, осуществляемый системой, должен определяться некоторыми этическими императивами и нормами в самом широком смысле. Например, этические нормы могут трактоваться как некоторые эвристики, которыми руководствуется ИИС при совершении выбора того или иного действия, формирования системы оценок, целевых функций и прочего» [4, с. 86]. Существует «проблема рамки», связанная с трудностями отделения релевантной информации, необходимой для принятия решения, от нерелевантной в условиях, когда система получает из окружающей среды огромные массивы данных [9].

В этике как философии нравственности с давних пор различаются подходы к моральной оценке действия человека на основании последствий данного действия (консеквенционализм) и на основании соответствия того же действия обязанностям человека, долгу, установленным правилам (деонтология). Одна из важных проблем этики ИИ связана с неспособностью системы предвидеть последствия решений и действий – будь то действие в материальном мире, поиск информации или обнаружение патологии в организме больного. Человек тоже не всегда может знать, к чему приведет то или иное его решение или действие. Однако именно на человеке лежит ответственность за действия, включая совершенные с помощью техники или на основании результатов работы техники и технологий, в том числе технологий ИИ. В некоторых сферах деятельности – например, в медицине, применение программного обеспечения с технологией ИИ подлежит особому контролю со стороны государства. Темой, обсуждаемой в прессе, стало принятое в ноябре 2023 г. решение Росздравнадзора о приостановке работы одного из таких сервисов, предназначенного для распознавания патологий на снимках, и о проведении внепланового выборочного контроля в отношении производителя. Решение принято на основании вывода о «наличии угрозы причинения вреда жизни и здоровью граждан» при использовании программного обеспечения [10].

Значимость этической регуляции ИИ отмечается в официальных документах. В тексте утвержденной в 2019 году Указом Президента Российской Федерации «Национальной стратегии развития искусственного интеллекта на период до 2030 года» среди «основных направлений создания комплексной системы регулирования общественных отношений, возникающих в связи с развитием и внедрением технологий искусственного интеллекта» указана «разработка этических правил взаимодействия человека с искусственным интеллектом» [12, 49 ж]. В концепции развития искусственного интеллекта, представленной правительством Российской Федерации осенью 2023 г., предложено решать часть вопросов разработки и применения ИИ в рамках «Кодекса этики в

сфере искусственного интеллекта» [7]. Проблема, однако, заключается в том, обладает ли упомянутый кодекс – как и аналогичные своды этических правил – всеми теми свойствами, которые требуются для его успешного применения на практике.

2 Возможности кодекса этики: новые проблемы и давние дискуссии

Разработчики интеллектуальных роботов В. Э. Карпов и В. В. Леушина следующим образом сформулировали позицию, исходя из которой анализируют и оценивают этические кодексы в сфере ИИ: «Необходимо понимать, что этические вопросы, касающиеся ИИ, стоят перед всем мировым сообществом, а это означает, что необходимо разработать некую документальную, нормативную основу, которой смогут следовать все страны, чтобы на ее основе стало возможным сформулировать уточняющие стандарты или рекомендации, учитывающие собственные ценности, культурные традиции, моральные нормы различных стран» [6, с. 125]. В качестве одного из претендентов на создание общей основы подробно рассмотрены «Рекомендации по этике искусственного интеллекта» (“Recommendation on the Ethics of Artificial Intelligence”), принятые на Генеральной конференции ЮНЕСКО в 2021 г. [18]. Российские авторы положительно оценивают попытку определить «универсальную модель этичного ИИ», однако с сожалением отмечают слишком большое количество «конъюнктурных и сомнительных пассажей», содержащихся в данном документе.

Следует отметить объективную сложность разработки рекомендаций для всех «акторов ИИ», действующих на любой стадии жизненного цикла искусственной интеллектуальной системы – включая, в числе прочего, исследования, проектирование, продажи, финансирование, использование, демонтаж. Речь идет об субъектах и агентах, ученых, программистах, инженерах, предпринимателях, конечных пользователях, университетах, общественных организациях, о самых разных физических и юридических лицах. Очевидно, что все они, в зависимости от рода деятельности, сталкиваются со специфическими проблемами этического характера – например, существенная часть проблем, с которыми имеет дело продавец, отличается от тех, с которыми имеет дело ученый.

Одной из основных тем «компьютерной этики» – направления, оформившегося первоначально в США в 80-е годы прошедшего века, – стала тема информационно-технологического (компьютерного, цифрового) неравенства как нового вида неравенства, дополняющего «старое» неравенство в распределении материальных благ [3]. Эта тема звучит и в принятых более тридцати лет спустя рекомендациях ЮНЕСКО по этике искусственного интеллекта. Однако соединение в одном ряду самых видов неравенств и дискриминаций способно вызвать закономерное недоумение «акторов ИИ», живущих и работающих в странах с достаточно глубоко укоренившимися идеями равенства.

Например, в России, где равноправие мужчин и женщин было узаконено более ста лет назад, часто с недоумением воспринимают рекомендации, касающиеся соблюдения в жизненном цикле систем ИИ принципов гендерного равенства. Так, В. В. Леушина и В. Э. Карпов выражают сомнение в целесообразности продвижения «гендерно неспецифического языка» в целях расширения представленности женщин в области естественных и технических наук [6, с. 129–130]. Мы же считаем продвижение подобных языков вовсе нецелесообразным, будучи убеждены в том, что выбор девушкой той или иной профессии должен определяться интересами, вкусами и способностями данной девушки, а не императивами «гендерного равенства». И этот выбор реально осуществляется за пределами тех стран, где установлены обусловленные религиозными верованиями (а вовсе не языком) ограничения на образование и работу женщин.

Что касается российского «Кодекса этики в сфере искусственного интеллекта», то содержащиеся в нем основные установки и принципы сходны или совпадают с теми, что отражены в «Рекомендациях» ЮНЕСКО и других сводах правил подобного рода. Такое сходство вполне закономерно, однако национальный кодекс выгодно отличается отсутствием установок «позитивной дискриминации», а также (мы отмечали это в ранее опубликованных работах [1]) акцентированием необходимости «сохранения интеллектуальных способностей человека как самостоятельной ценности и системообразующего фактора современной цивилизации» [5].

Заключение

Дискуссии по поводу собственно этического статуса этических кодексов имеют давнюю историю. Одни участники таких дискуссий оценивают кодексы как выражение «коллективной мудрости» сообществ, другие – как своды «клубных правил», не имеющие отношения к моральной философии, предполагающей свободу субъекта [2]. Обсуждение содержания и проблем применения этических рекомендаций – законная часть этики ИИ как новой области междисциплинарных исследований. Принятие международными организациями глобально ориентированных рекомендаций отвечает реально существующим запросам, обусловленным вовлеченностью людей в разных странах в разработку и применение систем ИИ. Однако попытка учесть в одном обширном документе проблемы, имеющиеся в регионах мира с разными культурными традициями, и притом актуальные для людей разных профессий и родов деятельности, существенно затрудняет восприятие и практическое использование подобных документов. Было бы предпочтительно иметь компактные формулировки действительно общих установок («для всех»), дополняемые более конкретными рекомендациями с учетом типов систем ИИ, родов деятельности людей и культурных особенностей разных стран.

Литература

1. Алексеев А. П., Алексеева И. Ю. Естественный интеллект в условиях цифровых трансформаций // Информационное общество. 2022. № 1. С. 2–8. Извлечено от http://infosoc.iis.ru/article/view/702.
2. Алексеева И. Ю. Прикладная этика как культурная система // Научно-техническое развитие и прикладная этика. М., 2014. С. 60–82.
3. Алексеева И. Ю., Шклярик Е. Н. Что такое компьютерная этика? // Вопросы философии. 2007. № 9. С. 60-72.
4. Карпов В. Э., Готовцев П. М., Ройзензон Г. В. К вопросу об этике и системах искусственного интеллекта // Философия и общество. 2018. № 2. С. 84–105.
5. Кодекс этики в сфере ИИ // Альянс в сфере искусственного интеллекта. URL: https://ethics.a-ai.ru/
6. Леушина В. В., Карпов В. Э. Этика искусственного интеллекта в стандартах и рекомендациях // Философия и общество. 2022. №. 3. С. 124–140.
7. Правительство представило концепцию по развитию искусственного интеллекта. URL: https://www.economy.gov.ru/material/news/pravitelstvo_predstavilo_koncepciyu_po_razvitiyu_iskusstvennogo_intellekta.html
8. Разин А. В. Этика искусственного интеллекта // Философия и общество. 2019. № 1 С. 57–73.
9. Середкина Е. В. Этические аспекты социальной робототехники // Человек. 2020. Т. 31, № 4. С. 109–127.
10. Сообщение Росздравнадзора о медизделиях, являющихся программным обеспечением с технологией ИИ // Росздравнадзор. URL: https://t.me/roszdravnadzor_official/2029
11. Тихомиров Ю.А., Крысенкова Н.Б., Нанба С.Б., Маргушева Ж.А. Робот и человек: новое партнерство? // Журнал зарубежного законодательства и сравнительного правоведения. 2018. № 5. С. 5–10.
12. Указ Президента Российской Федерации от 10.10.2019 г. №490 О развитии искусственного интеллекта в Российской Федерации. URL: http://www.kremlin.ru/acts/bank/44731
13. Almeida de, P.G.R., Santas dos, C.D., Farias, J.S. Artificial Intelligence Regulation: a framework for governance // Ethics Inf Techno. 2021. Vol. 23. Pp. 505–525. URL: https://doi.org/10.1007/s10676-021-09593-zIssue DateSeptember 2021
14. Dubber M. D., Pasquale F., Das Es. (Eds.) The Oxford Handbook of Ethics of AI. Oxford University Press. 2020. 896 p.
15. Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence // European Commission. URL: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf

16. Guardian: в OpenAI призвали к созданию аналога МАГАТЭ для контроля за разработкой искусственного интеллекта (ИИ). URL: https://russian.rt.com/inotv/2023-05-25/Guardian-v-OpenAI-prizvali-k?ysclid=loefi635io389771023)
17. Karpov V.E. Can a robot be a moral agent? // Artificial Intelligence. RCAI 2020. Lecture Notes in Artificial Intelligence (LNAI), 2020. Vol. 12412. Pp. 61–70.
18. Recommendation on the Ethics of Artificial Intelligence //UNESCO. URL: https://unesdoc.unesco.org/ark:/48223/pf0000381137

THE STATUS OF ETHICAL CODES IN THE ETHICS OF ARTIFICIAL INTELLIGENCE

Alekseev, Aleksandr Petrovich

DSc in philosophy, professor
Lomonosov Moscow State University, Philosophical faculty, Department of philosophy for humanities, chairman
Moscow, Russiam Federation
aleksandr.alekseev.1957@list.ru

Alekseeva, Irina Yurievna

DSc in philosophy, associate professor
Institute of Philosophy, Russian Academy of Sciences, Department of philosophical problems of social sciences and humanities, leading researcher
Moscow, Russian Federation
ialexeev@inbox.ru

Abstract

The authors of the article consider the development and discussion of ethical codes (recommendations) as a necessary part of the ethics of artificial intelligence (AI), a new interdisciplinary field emerging in the 21st century. The analysis of the structure and content of the sets of ethical recommendations in the field of AI, adopted by international and national organizations, allows the authors of the article to conclude that it is preferable to combine compact formulations of generally binding norms with concretization in relation to the types of activities of people involved in the life cycle of AI systems, taking into account the types of AI systems and cultural characteristics of countries.

Keywords

artificial intelligence; ethics of artificial intelligence; philosophy of artificial intelligence; code of ethics; recommendations on ethics of artificial intelligence

References

1. Alekseev A. P., Alekseeva I. Yu. Estestvenny`j intellekt v usloviyax cifrovy`x transformacij // Informacionnoe obshhestvo. 2022. № 1. S. 2–8. Izvlecheno ot http://infosoc.iis.ru/article/view/702.
2. Alekseeva I. Yu. Prikladnaya e`tika kak kul`turnaya sistema // Nauchno-texnicheskoe razvitie i prikladnaya e`tika. M., 2014. S. 60–82.
3. Alekseeva I. Yu., Shklyarik E. N. Chto takoe komp`yuternaya e`tika? // Voprosy` filosofii. 2007. № 9. S. 60–82.
4. Karpov V. E`., Gotovcev P. M., Rojzenzon G. V. K voprosu ob e`tike i sistemax iskusstvennogo intellekta // Filosofiya i obshhestvo. 2018. № 2. S. 84–105.
5. Kodeks e`tiki v sfere II // Al`yans v sfere iskusstvennogo intellekta. https://ethics.a-ai.ru/
6. Leushina V. V., Karpov V. E`. E`tika iskusstvennogo intellekta v standartax i rekomendaciyax // Filosofiya i obshhestvo. 2022. №. 3. S. 124–140.
7. Pravitel`stvo predstavilo koncepciyu po razvitiyu iskusstvennogo intellekta. https://www.economy.gov.ru/material/news/pravitelstvo_predstavilo_koncepciyu_po_razvitiyu_iskusstvennogo_intellekta.html
8. Razin A. V. E`tika iskusstvennogo intellekta // Filosofiya i obshhestvo. 2019. № 1 S. 57–73.
9. Seredkina E. V. E`ticheskie aspekty` social`noj robototexniki // Chelovek. 2020. T. 31, № 4. S. 109–127.
10. Soobshhenie Roszdravnadzora o medizdeliyax, yavlyayushhixsya programmny`m obespecheniem s texnologiej II // Roszdravnadzor. https://t.me/roszdravnadzor_official/2029
11. Tikhomirov Yu.A., Kry`senkova N.B., Nanba S.B., Margusheva Zh.A. Robot i chelovek: novoe partnerstvo? // Zhurnal zarubezhnogo zakonodatel`stva i sravnitel`nogo pravovedeniya. 2018. № 5. S. 5–10.
12. Ukaz Prezidenta Rossijskoj Federacii ot 10.10.2019 g. №490 O razvitii iskusstvennogo intellekta v Rossijskoj Federacii. http://www.kremlin.ru/acts/bank/44731

13. Almeida de, P.G.R., Santos dos, C.D., Farias, J.S. Artificial Intelligence Regulation: a framework for governance // Ethics Inf Techno. 2021. Vol. 23. Pp. 505–525. https://doi.org/10.1007/s10676-021-09593-zIssue DateSeptember 2021
14. Dubber M. D., Pasquale F., Das Es. (Eds.) The Oxford Handbook of Ethics of AI. Oxford University Press. 2020. 896 p.
15. Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence // European Commission. https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf
16. Guardian: OpenAI prizvali k sozdaniyu analoga MAGATE` dlya kontrolya za razrabotkoj iskusstvennogo intellekta (II). https://russian.rt.com/inotv/2023-05-25/Guardian-v-OpenAI-prizvali-k?ysclid=loefi635io389771023)
17. Karpov V.E. Can a robot be a moral agent? // Artificial Intelligence. RCAI 2020. Lecture Notes in Artificial Intelligence (LNAI), 2020. Vol. 12412. Pp. 61–70.
18. Recommendation on the Ethics of Artificial Intelligence //UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137