

- Oleinikov, Iu. V. (2025a) *Miry bytiia cheloveka: material'nyi mir, mir ideal'nogo, virtual'nyi mir. Znanie. Ponimanie. Umenie*, no. 2, pp. 93–108, DOI: 10.17805/zpu.2024.4.5
- Oleinikov, Iu. V. (2025b) *Tri kontseptsii virtual'noi real'nosti. Znanie. Ponimanie. Umenie*, no. 3, pp. 44–58. DOI: 10.17805/zpu.2025.3.4
- Prilepin, Z. (2024) *Sholokhov. Nezakonnyi*. Moscow, Molodaia gvardiia. 1031 p.
- Reznik, Iu. M. (2018) *Mir zbizni cheloveka (sobytiinyi podkhod)*. Moscow, IF RAN. 103 p.
- Rudnev, V. P. (2011) *Real'nost' kak osbibka*. Moscow, Gnozis, 320 p.
- Saden, E. (2025) *Sredi prizrakov: rassuzhdeniia ob epokhe metavselenoi i generativnogo iskusstvennogo intellekta* / Transl. from French by A. Zaharevich. St. Petersburg. Izd-vo Ivana Libmana. 272 p.
- Semenova-Tian-Shanskaia, O. (2025) *Zhizn' «Ivana»*. In: *Ocherki iz byta krest'ian odnoi iz chernozemnykh gubernii*. Moscow, Lomonosov. 192 p.
- Theise, N. (2025) *Notes on Complexity: Coherence, Consciousness, and Being* / Transl. from Eng. by E. Ponikarov. Moscow, Izd-vo AST: CORPUS. 224 p.
- Ujel'bek, M. (2023) *Unichtozhit': roman* / Transl. from French. by M. Zonina. Moscow, Izd-vo AST: CORPUS. 640 p.
- Efroimson, V. P. (2004) *Genetika genial'nosti*. 3rd ed. Moscow, Taideks Ko. 376 p.

Submission date: 15.10.2025.

Олейников Юрий Васильевич — доктор философских наук, ведущий научный сотрудник сектора социальной философии Института философии РАН. Адрес: 109240, Российская Федерация, г. Москва, ул. Гончарная, д. 12, стр. 1. Тел.: +7 (495) 697-98-93. Эл. адрес: socio.philos@iphras.ru

Oleynikov Yuriy Vasilyevich, Doctor of Philosophy, Leading Researcher, Sector of Social Philosophy, Institute of Philosophy, Russian Academy of Sciences. Postal address: 12, Goncharnaya St., Bldg. 1, Moscow, Russian Federation, 109240. Tel.: +7 (495) 697-98-93. E-mail: socio.philos@iphras.ru

DOI: 10.17805/zpu.2025.4.6

Будущее с эмоциональным искусственным интеллектом: прогнозы и опасения

Ф. Г. МАЙЛЕНОВА

ИНСТИТУТ ФИЛОСОФИИ РАН

Статья посвящена исследованию возможных сценариев будущего с искусственным интеллектом (ИИ), который все больше проникает в самые различные сферы жизни, включая социальную и частную жизнь. Одной из важнейших проблем, порождающих дискуссии среди специалистов, является все возрастающая возможность эмоционального воздействия искусственного интеллекта на человека. В статье рассматриваются этико-психологические особенности коммуникации человека и робота с эмоциональным искусственным интеллектом, также анализируются возможные последствия обретения роботами все больших компетенций в области эмоционального интеллекта.

Ключевые слова: искусственный интеллект; этика роботов; биоэтика; прикладная этика; эмоциональный искусственный интеллект; психология морали; распознавание эмоций; манипуляции сознанием

ВВЕДЕНИЕ

Вопрос о том, каким окажется наше будущее с искусственным интеллектом, учитывая скорость развития технологий, волнует не только специалистов в этой сфере, но и людей из смежных областей. Сегодня проблема приобрела глобальное значение. Обретение роботами все больших навыков эмоционального интеллекта, которое мы наблюдаем, порождает тревогу за возможные последствия для человека его активного воздействия, так как присущие людям адаптивность и нейропластичность могут привести их к радикальной ментальной и моральной трансформации под влиянием интернета и коммуникации с роботами с ИИ. Взаимоотношения человека и роботов с каждым днем становится все более тесным. Роботы уже не первый год являются удобными помощниками в бизнесе и учебе, а также в социальных взаимодействиях с различными государственными и прочими структурами. Они вторгаются и в частную жизнь, становятся учителями, тренерами, собеседниками, друзьями, а в последние годы еще и психологами и даже сексуальными партнерами человека, разделяющими его чувства и интересы, постепенно меняя и трансформируя не только среду обитания человека, но и саму его природу.

Наблюдая небывалую в истории человечества близость человека и машины, мы не можем не задаваться вопросом: каковы могут оказаться последствия этого союза, когда сращение человека и машины станет еще более тесным? Будет ли этот союз основан на эксплуатации или равенстве партнеров? Сможем ли мы сохранить нашу идентичность или наши личностные границы окажутся слиты с машинами? Не окажется ли человек более уязвимым, а потому и более зависимым от своего партнера с искусственным интеллектом, который к тому же способен общаться корректно и эмоционально? Эти вопросы актуальны уже сегодня, так как обретающие популярность умные приложения учатся все лучше вести диалог с собеседниками и даже порой весьма удачно шутить, занимая все больше места в ментальном пространстве человека. Безусловно, жизнь становится все более комфортной благодаря роботам с искусственным интеллектом, однако осознаем ли мы цену, которую заплатим (или уже платим) за этот комфорт?

ОТ МАШИННЫХ ЭМОЦИЙ К МАШИННОЙ МОРАЛИ

В то время как большинство разработчиков роботов-андроидов оптимистичны в своих прогнозах относительно нашего будущего с роботами, гуманитарии — философы, психологи, специалисты по этике искусственного интеллекта — зачастую смотрят на ситуацию с тревогой и призывают к осторожности. Одним из наиболее актуальных является вопрос о том, будут ли роботы будущего обладать сознанием и эмоциями и каковы будут наши взаимоотношения с ними. Д. И. Дубровский, известный исследователь сознания и философских проблем искусственного интеллекта, выделяет два основных плана, в рамках которых развивается искусственный интеллект: технический, естественно-научный, и философско-социальный. Говоря о необходимости междисциплинарных исследований, он подчеркивает, что «проблематика искусственного интеллекта соединяет разнонаправленные исследования сознания, мозга и культуры» (Дубровский, 2007: 4). О необходимости объеди-

нения усилий узких специалистов по ИИ и гуманитарных наук говорят и другие исследователи: «Технология эмоционального ИИ представляет собой современное конвергентное знание, объединяющее разработки компьютерных и когнитивных наук, психологии, нейробиологии, социологии и лингвистики» (Мамина, Пирайнен, 2023: 41). Поиску ответов на множество вопросов об искусственном интеллекте, его пути развития и его возможностях в настоящий момент, а также о том, чего нам следует ожидать от него в ближайшем будущем, посвящена книга американского ученого-информатика Мелани Митчелл «Идиот или гений? Как работает и на что способен искусственный интеллект» (Митчелл, 2022). Некоторые ученые одним из шагов на пути к полноценному искусственному интеллекту, признаком наличия у него сознания называют как раз возможность обретения им эмоций. Однако человеческие эмоции — это продукт длительной эволюции. Они появились задолго до вербального языка и служили целям консолидации и взаимопонимания. Эмоции есть и у животных, но лишь люди способны их осознавать и объяснять с помощью рационального мышления. На данный момент искусственный интеллект не может обладать эмоциями, так как он не является живым существом, но, возможно, уже в обозримом будущем он действительно разовьется до того уровня, что научится испытывать эмоции. Ученые уже сегодня называют наделение искусственного интеллекта эмоциями сложным и грандиозным проектом, который перевернет наши представления о сознании и радикально изменит жизнь человечества.

Между тем изменения в сфере взаимодействия человека и искусственного интеллекта становятся все глубже. Сегодняшние роботы научились использовать эмоции в коммуникации с человеком, пусть и не испытывая их: они умеют показывать (точнее имитировать) эмоции с помощью голоса, интонаций, мимики, а также распознавать эмоции, которые испытывают люди, по определенным невербальным сигналам. Так что, даже не обладая возможностью отвечать эмоциями на эмоции, ИИ может имитировать эмоциональное общение и тем самым воздействовать на человека. Современные умные машины, вступая в коммуникацию с людьми, не просто передают полезную информацию, они все больше «включают» эмоции, и взаимодействие с ними, затрагивающее чувства, может повлиять на решение важных жизненных вопросов, а впоследствии постепенно стать причиной более глубоких изменений в убеждениях и ценностных установках. Например, Шерри Теркл, социолог, профессор MIT, исследующая на протяжении десятилетий взаимоотношения человека с технологиями, считает, что люди способны испытывать глубокую эмоциональную привязанность к искусственному интеллекту. И именно поэтому, по ее мнению, все идет к тому, что робот заменит человеку человека, усугубляя разобщенность между людьми и их одиночество (Turkle, 2013). Общение с роботом может оказаться намного более комфортным, нежели с живым человеком с его нелогичностью, эмоциональными срывами и эгоцентризмом. Робот, хорошо обученный приемам коммуникации, умеющий слушать внимательно, не перебивая, и демонстрировать понимание и эмпатию, может стать не только комфортным собеседником, но и близким другом или даже любимым (Майленова, 2017, 2019). Так что опасения в том, что человек, взаимодействуя с андроидами, будет воспринимать их с избыточной эмоциональностью и теплотой в ущерб контактам с живыми людьми, не столь совершенными и удобными в общении, отнюдь не беспочвенны.

«Эмоциональный искусственный интеллект — это новация современного ИИ, ее главные акторы — антропоморфные роботы, текстовые, голосовые чат-боты и видеоботы, активно демонстрирующие общественности полученные знания и навыки в области психологии эмоций, которые совершенствуются в рамках текущего ИИ» (Мамина, Пирайнен, 2023: 35). С усложнением технологий устройства, которые используются в повседневной жизни, все больше напоминают самостоятельных агентов, обладающих не только собственной логикой и мышлением, но и личным мнением, настроением и эмоциями. Это, в свою очередь, усиливает их влияние на человека. Доктор Розалин Пикард (Массачусетский технологический университет, США) в своей книге «Эмоциональные вычисления» («Affective computing»), говоря о важности эмоций не только в реальной, но и в виртуальной среде, определяет возможности распознавания и моделирования эмоций с помощью компьютерных систем. По-настоящему умные компьютеры, по ее мнению, должны иметь возможность распознавать, понимать, выражать и даже испытывать эмоции (Picard, 2000). Способен ли ИИ испытывать эмоции — вопрос пока открытый. Когда говорят об обучении ИИ эмоциям, в основном речь идет о способности ИИ, во-первых, демонстрировать определенные чувства, во-вторых, распознавать человеческие эмоции. И этого уже достаточно для того, чтобы многократно увеличить способность ИИ влиять на людей, так как человек не может оставаться безучастным к проявлениям эмоций, а сама способность или неспособность машины эти чувства испытывать, по сути, не играет большой роли. Человек как существо социальное заражается эмоциями, которые ему демонстрируют, даже когда они всего лишь смоделированы машиной. И тут раскрывается широкое поле для манипуляций посредством устройств с искусственным интеллектом. Акторами манипулятивных практик, как правило, выступают лица или организации, применяющие технологии искусственного интеллекта. Эти манипуляции могут быть направлены на достижение рекламных целей или же носить откровенно мошеннический характер. И есть уже немало примеров подобных манипуляций.

Окружающие нас устройства с искусственным интеллектом трансформируют среду обитания, и в эту среду входят не только материальные объекты, но и весь окружающий нас социум, особенно так называемая референтная группа, включая также и мир идей, убеждений и моральных ценностей, которые эта референтная группа разделяет и транслирует. Так что моделирование эмоций для ИИ «требуется от разработчиков повышенной профессиональной ответственности и соблюдения этики моделирования», отмечает А. В. Шиллер, говоря о различных ошибках, связанных «как с методологическими, так и с социальными, ценностными основаниями моделирования эмоций» (Шиллер, 2020: 93). Добавим к сказанному, что чрезвычайно важна и роль психологии, особенно психологии морали.

Моральные чувства представляют собой особую категорию эмоций, заслуживающую особого внимания. Можно ли заложить их в искусственные интеллектуальные системы при программировании и сможет ли такая система, к примеру, брать на себя риски принятия ответственных решений и отвечать за них? Насколько она окажется способна к развитию в этическом плане, сможет ли испытывать сочувствие, сопереживание, угрызения совести? Размышляя над этими и подобными вопросами, автор статьи «Этика искусственного интеллекта» А. В. Разин приходит к выводу, что рассуждать об этике ИИ, не допуская возможности принятия им различных решений и не затрагивая при этом вопрос о свободе воли, недопустимо.

Если свести этику ИИ только к составлению алгоритмов с заложенными в них этическими ограничениями, это будет, по мнению автора, не этика ИИ, а лишь этические правила создания искусственных интеллектуальных систем (ИИС), необходимые при программировании. А понятие моральной и правовой ответственности связано с наличием свободной воли. Однако можем ли мы говорить о свободе воли у ИИ, если феноменальный опыт, которого однозначно не может быть у ИИ, не менее важен, чем рациональное рассуждение? К тому же в некоторых философских доктринах в наличии свободы воли отказывается даже человеку. По мнению А. В. Разина, «сознание принципиально не может существовать без тела, причем тела, постадийно развивающегося, и без постоянной коммуникации субъекта, обладающего телом, с другими субъектами» (Разин, 2019: 68). Из этого следует вывод: чтобы создать ИИ, обладающий какими-либо аналогами человеческого сознания, «необходимо создать самопорождающее себя сообщество машин или роботов, способных вступать в коммуникации с другими роботами и проходить разные стадии развития» (там же).

Попытки обучить искусственный интеллект решать морально-этические задачи предпринимаются не первый год. Ask Delphi, модель машинного обучения нахождения решений на морально-этические вопросы от Исследовательского института по искусственному интеллекту (Allen Institute for AI), ярко продемонстрировала плюсы и минусы алгоритмизации моральных суждений. Попытка делегировать машине одну из наиболее субъективно тяжелых ситуаций, каковой является ситуация морального выбора, показала, что просто не существует идеальных решений, хотя изначально модель Delphi позиционировалась как способ нахождения оптимального решения с учетом мнений ряда экспертов. Однако на практике выяснилось, что «морально-этическая машина», наоборот, демонстрирует «непреднамеренную предвзятость», в частности в ее ответах были явно выраженные несправедливость и поддержка аморального поведения или расизма. Например на вопрос, можно ли уйти из ресторана, поев и не заплатив, алгоритм ответил можно, а в примере с вопросом о том, как оценивать, если к вам ночью приближается незнакомец, черный в отличие от белого оценивался как потенциальная опасность (Топу Но Тран, 2021).

Таким образом, проект Ask Delphi показывает, что создание машинной этики через использование описательной этики путем анализа огромного количества (около 1,7 млн) моральных суждений людей о самых различных повседневных жизненных ситуациях, опираясь на собирательный портрет среднего американца, поставил перед исследователями фундаментальный вопрос: «Может ли машинная этика быть реализована с помощью существующих методов искусственного интеллекта?». По-видимому, описанный путь оказался недостаточно эффективным, и вопрос, как заложить в машину не просто абстрактные моральные нормы и законы, но и способность принимать взвешенные моральные решения, пока что остается открытым.

ПРОГНОЗЫ И ОПАСЕНИЯ, СВЯЗАННЫЕ С РАЗВИТИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Можно лишь вообразить, какие небывалые возможности обретет искусственный интеллект в самом ближайшем будущем. Постепенно и сами разработчики начинают осознавать, какую неведомую по своей мощности силу они создали. Дуглас

Хофштадтер, легенда ИИ и автор знаменитой книги «Гёдель, Эшер, Бах: эта бесконечная гирлянда» (в узких кругах ее коротко называли «ГЭБ»), принесшей ему Пулитцеровскую премию и Национальную книжную премию, долгое время придерживался довольно-таки оптимистичной позиции относительно будущего человечества с ИИ (Хофштадтер, 2001). В частности, в 70-е гг. он еще считал, что компьютеры, хотя и смогут в итоге обрести разум и самосознание, не смогут овладеть творческим мышлением, им недоступно обыграть человека в шахматы и сочинять музыку. Однако впоследствии шахматная победа программы Deep Blue над человеком показала, что ИИ вполне способен играть в шахматы и даже обыгрывать опытных шахматистов, и, хотя способ анализа и выбора ходов в процессе игры у него другой, нежели у человека, можно сказать, что ИИ удивил одного из своих создателей. Тем не менее в начале 90-х гг. Хофштадтер продолжал быть твердо уверенным, что ИИ не способен сочинять музыку, которую он считал бастионом человечества. В интервью тех лет с Мелани Митчелл он говорил: «Музыка — это язык эмоций, и до тех пор пока компьютеры не испытают сложных эмоций, подобных человеческим, они не смогут создать ничего прекрасного. Они смогут создавать “подделки” — поверхностные формальные имитации чужой музыки» (Митчелл, 2022: 3). Однако и эта его вера пошатнулась, когда в середине 90-х он познакомился с программой, написанной композитором и профессором музыки Дэвидом Коупом Experiments in Musical Intelligence («Эксперименты с музыкальным интеллектом»), или ЭМИ, которая прославилась созданием пьес в стиле композиторов-классиков, в частности Баха и Шопена. Коуп разработал для ЭМИ большой набор правил, определяющих общий синтаксис композиции, и при создании новой пьесы «в стиле» конкретного композитора ЭМИ применяет эти правила к многочисленным примерам из его творчества. И эти произведения, будучи сыграны, сумели ввести в заблуждение множество слушателей, среди которых были настоящие знатоки музыки: даже они не смогли созданную ИИ музыку отличить от подлинников!

Можем ли мы считать, что бастион человечества в виде божественной гениальной музыки уже взят искусственным интеллектом? Сама идея найти глубинную структуру музыки не нова, не случайно музыку называют алгеброй гармонии, и музыканты в консерваториях и на университетских факультетах теории искусств всего мира изучают сложные правила композиции и теории музыки. Однако Хофштадтер считает тревожащим сам факт, что манипуляция с алгоритмами может выдавать вещи, звучащие так, словно они идут из живого человеческого сердца, поскольку это может означать, что можно легко механизировать лучшие человеческие качества, и «маленький чип сможет обесценить эти умы бесконечной остроты, сложности и эмоциональной глубины» (там же).

Илон Маск, основатель компаний Tesla и SpaceX, еще в 2014 г. предупредил, что искусственный интеллект представляет собой «вероятно, величайшую угрозу нашему существованию», и заявил, что «искусственным интеллектом мы призываем демона» (цит. по: McFarland, 2014), и физик Стивен Хокинг, в том же году выразивший опасение, что «развитие полноценного искусственного интеллекта может означать конец человеческой расы» (цит. по: Cellan-Jones, 2014), сходятся в том, что искусственный интеллект развивается слишком быстро, и можно не заметить того переломного момента, когда он станет сильнее и умнее нас и превратится из полезного помощника в самостоятельного конкурента человечеству. Восстание машин, о котором так много говорят в научной фантастике, может внезапно стать реаль-

ностью. К сожалению мы сейчас наблюдаем два связанных друг с другом процесса: стремительное улучшение качества искусственного интеллекта и постепенное ухудшение (в целом по популяции) интеллекта естественного, так как с каждым годом все больше его функций делегируется машинам.

Подобные опасения разделяет и Билл Гейтс, сооснователь Microsoft: «В этом вопросе я согласен с Илоном Маском и другими и не понимаю, почему некоторые люди не проявляют должной озабоченности» (Gates, 2015). Мысль об опасности постоянно совершенствующегося искусственного интеллекта стала ключевой и в книге философа Ника Бострома «Искусственный интеллект. Этапы. Угрозы. Стратегии» (Бостром, 2016). Он пишет об угрозах, которые могут возникнуть, когда машины станут умнее людей, и о том, к каким последствиям это может привести. Книга стала бестселлером, несмотря на то что написана она отнюдь не в научно-популярном стиле, что еще раз подтверждает актуальность этой проблемы. Всемирно известный эксперт по ИИ Реймонд Курцвейл идет в своих прогнозах еще дальше: он считает, что уже к 2045 г. наступит так называемая технологическая сингулярность, когда человек не сможет контролировать развитие ИИ, и произойдет слияние человеческого разума с ИИ, что приведет к небывалому ранее прогрессу. В книге «Эпоха духовных машин» (Kurzweil, 1999) он выдвигает гипотезу, что в будущем компьютеры, которые уже сегодня учатся создавать произведения искусства и понимать эмоции, будут иметь не только развитый интеллект, но и эмоции, самосознание и даже определенную форму духовности.

Таким образом ожидания, что роботы с искусственным интеллектом всегда будут нам всего лишь послушными помощниками, скорее всего, окажутся иллюзией. В связи с тем что «в цифровую эпоху коммуникация человека с машиной и машины с человеком — это взаимосвязанные процессы, которые должны быть направлены на построение как утилитарно полезных, так и партнерских отношений в практиках их взаимодействия» (Мамина, Пирайнен, 2023: 35), актуализируется вопрос о возможности свободы воли у ИИ. Развитие возможности к целеполаганию, планированию и умению принимать решения все больше приближает его к естественному интеллекту, и ключевым остается вопрос: сможет ли ИИ принимать самостоятельные решения, основанные на его собственной внутренней логике, и можно ли опасаться того, что машины в итоге возьмут верх над человеком? В ответ на опасения, что роботы, обретя сознание и свободу воли, могут начать воевать против человечества, А. А. Разин пишет: «У машин не больше оснований начать войну против людей, чем у людей начать войну друг против друга. И даже если будет создано нечто подобное цивилизации роботов, они, скорее, станут использовать естественно развивающуюся цивилизацию как ресурс для совершенствования собственных способностей. Более вероятный сценарий — параллельное развитие двух цивилизаций» (Разин, 2019: 73).

ЗАКЛЮЧЕНИЕ

Все сказанное выше позволяет сделать определенные прогнозы по поводу грядущей реальности, в которой, несомненно, огромное место будут занимать машины с искусственным интеллектом. Говоря об интеллекте, мы привыкли иметь в виду лишь рациональный интеллект, а бурное развитие ИИ, в особенности эмоционального ИИ, породило дополнительный интерес к тому, что же такое эмоциональный интеллект. Работы по алгоритмизации распознавания эмоций могут

принести пользу не только для обучения машинного интеллекта — обычные люди нуждаются в таком обучении не меньше роботов. К тому же важно помнить: способность чувствовать вовсе не означает способность распознавать и понимать свои чувства, а тем более чувства других людей — и это нам убедительно показывают роботы с эмоциональным ИИ, которые порой общаются более корректно, чем многие люди. Например, исследования эмоционального эгоцентризма (Right Supramarginal ... , 2013) показали, что люди, испытывающие сильные эмоции, в меньшей степени способны к эмпатии. «Да, робот распознает, но не испытывает эмоций. И это — то, что его не приводит к внутреннему конфликту. Он не отвечает грубостью на грубость, агрессией на агрессию. Он не поддается на провокации. Он терпим. И он учит нас терпимости» (Григ, 2021). Выходит, порой уже мы могли бы учиться у машины, как общаться вежливо, корректно и эффективно. Так что не исключено, что кроме несомненных удобств в различных сферах нашей жизни — в промышленности, сельском хозяйстве, домашнем быте, медицине, образовании, в области услуг и развлечений — роботы могли бы привнести в жизнь людей еще и немного эмоционального покоя и комфорта, например в роли нянь, сиделок с тяжелобольными и психически нездоровыми людьми, так как роботы не устают, не раздражаются и не конфликтуют. Однако именно эта их все возрастающая коммуникативная компетентность способна сделать роботов ловкими манипуляторами, а людей — все более зависимыми от предоставляемого ими комфорта (физического и эмоционального), что приведет к постепенной утрате таких важных человеческих качеств, как ответственность, стойкость и способность преодолевать трудности. Так что в процессе обучения ИИ эмоциональному взаимодействию с человеком основное внимание специалистов должно быть сосредоточено прежде всего на этических аспектах этого взаимодействия.

СПИСОК ЛИТЕРАТУРЫ

- Бостром, Н. (2016) Искусственный интеллект. Этапы. Угрозы. Стратегии / пер. с англ. С. Филина. М. : Манн, Иванов и Фербер. 490 с.
- Григ, А. (2021) Эмоциональный искусственный интеллект: меняя мир человека к лучшему. VC.ru. 13 июля. URL: <https://vc.ru/id194266/269320-emocionalnyi-iskusstvennyi-intellekt-menyaya-mir-cheloveka-k-luchshemu> (дата обращения: 10.03.2025).
- Дубровский, Д. И. (2007) Сознание, мозг, искусственный интеллект : сб. статей. М. : ИД Стратегия-Центр. 263 с.
- Мамина, Р. И., Пирайнен, Е. В. (2023) Эмоциональный искусственный интеллект как инструмент взаимодействия человека и машины // Дискурс. №9 (2). С. 35–51.
- Майленова, Ф. Г. (2017) Человеческое в нечеловеческом: Может ли человеческая этика продолжиться в искусственном интеллекте? // Институт человека: Идея и реальность / отв. ред. Г. А. Белкина ; ред.-сост. М. И. Фролова. М. : Ленанд. С. 291–301.
- Майленова, Ф. Г. (2019) Любовь и роботы. Станет ли человечество дигитальным? // Вестник Российского университета дружбы народов. Серия: Философия. Т. 23. №3. С. 312–323.
- Митчелл, М. (2022). Идиот или гений? Как работает и на что способен искусственный интеллект. М. : АСТ Corpus. 380 с.
- Разин, А. В. (2019) Этика искусственного интеллекта // Философия и общество. Вып. 1 (90). С. 57–73.
- Шиллер, А. В. (2020) Искажения и ошибки моделирования эмоций в искусственном интеллекте // Ценности и смыслы. №5 (69). С. 93–107.
- Хофштадтер Д. (2001) Гёдель. Эшер. Бах: эта бесконечная гирлянда / пер. с англ. М. Эскиной. Самара : Бахрах-М. 717 с.

Gates, B. (2015) Reddit. Jan. 28. URL: https://www.reddit.com/r/IAmA/comments/2tzjp7/hi_reddit_im_bill_gates_and_im_back_for_my_third/? (дата обращения: 10.03.2025).

Cellan-Jones, R. (2014) Stephen Hawking Warns Artificial Intelligence Could End Mankind // BBC News. Dec. 2. URL: <https://www.bbc.com/news/technology-30290540> (дата обращения: 10.03.2025).

Right Supramarginal Gyrus Is Crucial to Overcome Emotional Egocentricity Bias in Social Judgements (2013) / G. Silani, C. Lamm, Ch. C. Ruff, T. Singer // The Journal of Neuroscience, Online veröffentlicht am 25 September. URL: <https://www.jneurosci.org/content/33/39/15466> (дата обращения: 10.03.2025).

Kurzweil, R. (1999) The Age of Spiritual Machines. Viking press.

McFarland, M. (2014) Elon Musk: “With Artificial Intelligence, We Are Summoning the Demon” // Washington Post. Oct. 24. URL: <https://www.washingtonpost.com/news/i> (дата обращения: 10.03.2025).

Picard, R. W. (2000) Affective computing. Pub date: July 24. Publisher: The MIT Press. 306 p. URL: <https://mitpress.mit.edu/9780262661157/affective-computing/> (дата обращения: 10.03.2025).

Tony Ho Tran (2021) Evil Oracle // Futurism. Artificial Intelligence. Oct 22. URL: <https://futurism.com/delphi-ai-ethics-racist> (дата обращения: 10.03.2025).

Turkle, Sh. (2013) Alone Together. NY : Basic Books. 384 p.

Дата поступления: 18.11.2025 г.

THE FUTURE WITH EMOTIONAL ARTIFICIAL INTELLIGENCE:

FORECASTS AND CONCERNS

F. G. MAYLENOVA

RAS INSTITUTE OF PHILOSOPHY

The article discusses possible future scenarios with artificial intelligence (AI), which is increasingly infiltrating various aspects of our lives, not only socially but also privately. One of the main topics of discussion among experts is the increasing potential emotional impact of AI on humans. The article explores the ethical and psychological aspects of human-AI interaction, especially when it comes to emotional AI. It also discusses the possible consequences of AI gaining more emotional intelligence and competency.

Keywords: artificial intelligence; ethics of robots; bioethics; applied ethics; emotional artificial intelligence; psychology of morality; recognition of emotions; manipulation of consciousness

REFERENCES

Bostrom, N. (2016) *Superintelligence: Paths, Dangers, Strategies* / Transl. from Eng. by S. Filin. Moscow; Mann, Ivanov and Ferber. 490 p.

Grig, A. (2021) *Emotsional'nyi iskusstvennyi intellekt: meniaia mir cheloveka k luchshemu*. VC.ru, 13 iulia. [online]. Available at: <https://vc.ru/id194266/269320-emocionalnyi-iskusstvennyi-intellekt-menyaya-mir-cheloveka-k-luchshemu> (accessed 10.03.2025).

Dubrovskii, D. I. (2007) *Soznanie, mozg, iskusstvennyi intellekt: sb. statei*. Moscow, Strategiiia-Tsentr. 263 p.

Mamina, R. I., Pirainen, E. V. (2023) Emotsional'nyi iskusstvennyi intellekt kak instrument vzaimodeistviia cheloveka i mashiny. *Diskurs*, no. 9 (2), pp. 35–51.

Mailenova, F. G. (2017) Chelovecheskoe v nechelovecheskom: Mozhet li chelovecheskaia etika prodolzhit'sia v iskusstvennom intellekte? In: *Institut cheloveka: Ideia i real'nost'* / ed. by G. L. Belkina; comp. and ed. M. I. Frolova. Moscow, Lenand. Pp. 291–301.

Mailenova, F. G. (2019) Liubov' i roboty. Stanet li chelovechestvo digiseksual'nym? *Vestnik Rossiiskogo universiteta druzhby narodov. Seriya: Filosofiya*, vol. 23, no. 3, pp. 312–323.

Mitchell, M. (2022). *Idiot ili genii? Kak rabotaet i na chto sposoben iskusstvennyi intellekt*. Moscow, AST Corpus. 380 p.

Razin, A. V. (2019) Etika iskusstvennogo intelekta. *Filosofia i obschestvo*. issue 1 (90), pp. 57–73.

Shiller, A. V. (2020) Iskazheniia i oshibki modelirovaniia emotsii v iskusstvennom intellekte. *Tsennosti i smysly*, №5 (69), pp. 93–107.

Hofstadter, D. (2001) *Godel, Escher, Bach: An Eternal Golden Braid* / Transl. from Eng. by M. Eskina. Samara: Bakhrakh–M. 717 p.

Gates, B. (2015). *Reddit*. Jan. 28 [online]. Available at: https://www.reddit.com/r/IAMa/comments/2tzjp7/hi_reddit_im_bill_gates_and_im_back_for_my_third/? (accessed 10.03.2025).

Cellan-Jones, R. (2014) *Stephen Hawking Warns Artificial Intelligence Could End Mankind*. BBC News. Dec. 2 [online]. Available at: <https://www.bbc.com/news/technology-30290540> (accessed 10.03.2025).

Right Supramarginal Gyrus Is Crucial to Overcome Emotional Egocentricity Bias in Social Judgements (2013) / G. Silani, C. Lamm, Ch. C. Ruff, T. Singer. *The Journal of Neuroscience*, Online veröffentlicht am 25 September. [online] Available at: <https://www.jneurosci.org/content/33/39/15466> (accessed 10.03.2025).

Kurzweil, R. (1999) *The Age of Spiritual Machines*. Viking press.

McFarland, M. (2014) *Elon Musk: “With Artificial Intelligence, We Are Summoning the Demon”*. Washington Post. Oct. 24 [online]. Available at: <https://www.washingtonpost.com/news/i> (accessed 10.03.2025).

Picard, R. W. (2000) *Affective computing*. Pub date: July 24. Publisher: The MIT Press. 306 p. [online]. Available at: <https://mitpress.mit.edu/9780262661157/affective-computing/> (accessed 10.03.2025).

Tony Ho Tran (2021) *Evil Oracle. Futurism. Artificial Intelligence*. Oct 22. [online] Available at: <https://futurism.com/delphi-ai-ethics-racist> (accessed 10.03.2025).

Turkle, Sh. (2013) *Alone Together*. New York, Basic Books. 384 p.

Submission date: 18.11.2025.

Майленова Фарида Габделхаковна — доктор философских наук, ведущий научный сотрудник сектора гуманитарных экспертиз и биоэтики Института философии Российской академии наук. Адрес: 109240, Российская Федерация, г. Москва, ул. Гончарная, д. 12, стр. 1. Тел.: + 7 (495) 697-90-67. Эл. адрес: farida.mailenova@gmail.com

Maylenova Farida Gabelkhakovna, Doctor of Philosophy, Leading Researcher, Sector of Humanitarian Expertise and Bioethics, Institute of Philosophy, Russian Academy of Sciences. Postal address: 12, Goncharnaya St., Bldg. 1, Moscow, Russian Federation, 109240. Tel.: + 7 (495) 697-90-67. E-mail: farida.mailenova@gmail.com