

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: СУБЪЕКТ И ОБЪЕКТ

Ключевые слова: искусственный интеллект, этика, телесность, технологии

Данная статья – этико-культурологический анализ взаимодействия человека с искусственным интеллектом (ИИ). Авторы рассматривают ИИ как “умного помощника” и как потенциально автономную личность. В первом случае речь идет о таких угрозах для человека, как усиленный контроль над частной жизнью, незащищенность личных данных, создание фейковой информации, сокращение рабочих мест, постепенное купирование человеческих способностей за счет технологической поддержки. Во втором – об угрозе порабощения человечества искусственным интеллектом или гибели человечества в результате того, что ИИ будет прогрессировать в темпе, недостижимом для человеческого мозга. В статье рассматриваются варианты отношений с ИИ в зависимости от того, будет ли он признан объектом или субъектом морали; анализируются и сравниваются интересы и ценности человека и ИИ, выявляются варианты морали и этики, обусловленные спецификой технологического и биологического развития личности. Исходя из того, что существует два основных мнения об истоках человеческой морали – базирующееся на осознанном выборе и детерминированное биологией, – авторы делают вывод, что главной проблемой на пути к единству интересов и целей человека и ИИ будет биологическая “бестелесность” последнего, поскольку большинство человеческих интересов ограничивается рамками телесности и смертности. Предполагается также, что ИИ не будет антропоцентричным и что его этика может быть ИИ-центричной или вневидовой

Существование человечества в привычных для нас биологических и социальных рамках кажется уже невозможным. Более того, многие эксперты видят угрозу для нашего дальнейшего развития. Причина этого – растущие риски, вызванные милитаризацией, экономическим неравенством и ухудшением экологической обстановки, но прежде всего связанные с развитием НБИК-технологий и разработками в сфере искусственного интеллекта (ИИ).

ИИ уже вышел из сферы мечтаний о будущем. Это уже наша реальность, хотя многие еще не осознают этого факта. Как пользователи интернета, клиенты банка,

Магдалена Кожевникова | <https://orcid.org/0000-0001-7484-9638> | kmagdalenayandex.ru | к. филос. н. | Институт этнологии и культурной антропологии Варшавского университета (Żurawia 4, 00-503 Warszawa)

Светлана Васильевна Карпова | <https://orcid.org/0000-0002-4488-8131> | SVKarpova@fa.ru | д. э. н., профессор | Финансовый университет при Правительстве РФ (Ленинградский пр. 49, Москва, 125993, Россия)

Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств в рамках государственного задания Финансовому университету в 2019 г.

покупатели и пациенты мы постоянно сталкиваемся с результатами работы самообучающихся нейронных сетей, для которых мы – источник новых данных. Сегодня ИИ для нас – это прежде всего “умный помощник”, который незаметно облегчает наши повседневные задачи на работе, помогает делать покупки, выбирать варианты отдыха. Существуют опасения, что это приведет не только к закрытию рабочих мест на рынке труда, изменению характера межлических контактов, но также ограничит наши способности – например, мы перестанем ориентироваться в пространстве без помощи GPS-навигации.

Некоторые, напротив, считают, что “умный помощник” – это технологическое расширение наших интеллектуальных способностей, некое усовершенствование (*enhancement*) человечества и неизбежный логический путь развития. Существуют опасения, что ИИ перерастет в самостоятельное, сознательное, автономное существо, которое, возможно, выйдет из-под контроля людей и станет угрозой для всего человечества. Однако существует также мнение, что даже в случае, если ИИ перерастет нас, он продолжит мирно к нам относиться.

Данная статья вписывается в русло этико-культурологического обсуждения взаимодействия человека с ИИ и связанных с этим рисков. Мы представим искусственный интеллект – каким он есть сегодня и каким он может стать завтра, – постараемся ответить на вопрос, возможно ли, чтобы ИИ руководствовался человеческими моральными нормами, а также задумаемся над тем, зачем людям вообще создавать ИИ.

История и метафизика

История исследований ИИ началась в середине прошлого века с развития кибернетики и компьютеров; целью этого развития было создание машины, решающей логические операции. Чаще всего появление размышлений об ИИ ассоциируют с “тестом Тьюринга” (эмпирический тест, предложенный Аланом Тьюрингом в статье “Computing Machinery and Intelligence”, опубликованной в 1950 г. в философском журнале *Mind*). Суть его в следующем: человек взаимодействует с компьютером и человеком, в ходе взаимодействия он должен определить на основании ответов на задаваемые вопросы, с кем разговаривает – с человеком или компьютерной программой. Стоит отметить, что в 2014 г. тест Тьюринга прошел чат-бот Женя Густман (*Eugene Goostman*), который представился 13-летним мальчиком из Одессы: 33 % судей приняли его за человека. Однако, несмотря на растущую схожесть поведения или, вернее, подражания человеку, до создания автономной и сознательной машины, которую можно было бы назвать “личностью”, все еще далеко.

Как это часто бывает с техническим прогрессом, большой скачок произошел во времена холодной войны: гонка вооружений и армейские разработки способствовали развитию новых технологий. Тренд актуален и сегодня и реализуется в создании автономного оружия и боевых роботов. Военным нужны машины, которые действовали бы максимально самостоятельно в боевых условиях и принимали решения почти без участия человека.

Термин “искусственный интеллект” (*Artificial Intelligence, AI*) ввел выдающийся информатик Джон Маккарти на Дартмутском семинаре в 1956 г. Семинар подразумевал двухмесячную работу группы ученых и профильных специалистов в сфере разработки вычислительных машин, его целью было помочь исследователям решить вопросы автоматизации и самообучения машин, использования ими языка, нейронных сетей и др. Участвующие в нем ученые согласились с тем, что мышление состоит из множества операций, которые можно провести вне биологического мозга, т.е. мышление не обязательно связано с биохимией.

В 1975 г. Хилари Патнэм представил функционалистский подход к теории разума, суть которого заключается в том, что интеллект может быть реализован на базе разных физических субстанций, если эти субстанции имеют определенную структуру. Согласно

теории функционализма, то, что “делает нечто ментальным состоянием определенного типа, зависит не от своего внутреннего устройства, а от того, как оно функционирует, какую роль играет в системе, частью которой является” (Левин 2013). Исходя из этой теории и рассматривалась возможность возникновения сознания в роботах или создание искусственного интеллекта. Современные разработчики предпочитают проектировать схемы, аналогичные схемам обработки сигналов биологическим мозгом (т.е. нейроморфные) как более эффективные.

Кшиштоф Гавковский приводит три самых популярных среди специалистов определения ИИ: “1) область исследований, которая пытается объяснить интеллектуальное поведение и подражать ему в терминах вычислительных процессов, 2) решение проблем методами, смоделированными на основе естественной человеческой деятельности и когнитивных процессов с помощью компьютерных программ, их имитирующих, и 3) попытки моделировать аспекты человеческого рассуждения (мышления) и решать такие проблемы, которые человек может решить быстрее при помощи компьютера” (Gawkowski 2018: 246). В общем, ИИ должен имитировать человеческое мышление и действовать в определенных ситуациях как человек. Преимущество ИИ перед человеком — его способность обрабатывать огромное число данных в максимально короткое время. В силу биологических и технических факторов человеческие способности сильно проигрывают возможностям компьютеров.

Сегодня искусственный интеллект — это способные решать задачи только определенного спектра узкоспециализированные программы, которые, однако, проявляются в виде “умного помощника” во множестве областей: в финансовой и банковской сферах, в управлении человеческими ресурсами, в службах поддержки клиентов и телекоммуникациях, в творчестве, в индустрии развлечений и игр, в СМИ и журналистике, в медицине и уходе за больными и престарелыми, в тяжелой промышленности и военных отраслях, на транспорте и в торговле, в криминалистике и др. Все эти применения приносят бесспорную пользу, иногда мы даже не замечаем участия умных технологий либо принимаем их как желаемое удобство и облегчение процесса. Однако в качестве факторов, усиливающих контроль над гражданами, ограничивающих личную свободу и постепенно вытесняющих людей с рынка труда, они вызывают также много вопросов социального и этического характера. Об этих рисках речь пойдет ниже.

Учитывая, что логическое мышление, т.е. одна из центральных характеристик ИИ, приписывается в основном (а во многих кругах — исключительно) человеку, можно сказать, что одной из важнейших задач разработок ИИ всегда было создание технологического продолжения человека, вплоть до “разумного Другого”. Мацей Ожуг пишет в этом контексте:

Классический тренд, связанный с ИИ, также называемый strong AI, тесно связан с философией и кибернетикой, сосредоточен на изучении природы человеческого интеллекта и поиске возможности создания искусственных устройств, способных моделировать его функционирование. В этом отношении его можно считать современным воплощением вечной мечты человечества о создании искусственного человека (Ożóg 2010: 50).

У человечества присутствует метафизическая тяга к поискам другого вида, с которым возможно было бы общение “на равных”. По нашему мнению, она может вытекать из чувства одиночества: люди ищут разумного Другого на Земле и в космическом пространстве, в мире материальном и нематериальном, среди богов, инопланетян и животных (мотив общения с животными на одном языке характерен как для детских сказок, напр., в серии Хью Лофтинга о докторе Дулиттле или в сказке К. Чуковского “Доктор Айболит”, так и для научных исследований по обучению человекообразных обезьян, дельфинов и других видов языку глухонемых, пиктограммам и т.д.). Люди

всегда мечтали и о создании “паралюдей”, в зависимости от эпохи и технонаучной среды – гомункулов, Голема из глины, роботов и андроидов.

Не все согласятся с тезисом об “одиночестве” в качестве основной мотивации человечества для создания других разумных существ. Например, Магдалена Радковска-Валькович, которая проанализировала образ “искусственного человека” в культуре, считает, что “отношения человек–машина никогда не бывают чистыми, бесконфликтными и нейтральными отношениями. Чаще всего мы сталкиваемся с отношениями власти: потребность царствования – это одна из основных причин создания искусственных людей” (*Radkowska-Walkowicz* 2008: 171). Бесспорно, такой аспект тоже присутствует и отражается как в заявлениях создателей “умных помощников” (облегчение повседневных задач человека), так и в научно-фантастических текстах культуры. Однако голоса, призывающие дать права роботам, и исследования в области, например, *Friendly AI*, свидетельствуют о глубоком желании включить потенциальные новые “личности” в наш моральный порядок.

Александра Пжегалинска – философ, специализирующийся в области ИИ, считает, что желание создать искусственный интеллект может рассматриваться как

выражение желания превзойти реальность и продлить биологическую жизнь через технологическую жизнь. Эта “вторая жизнь”, которую можно назвать также кибержизнью, подчиняется правилам запрограммированной рациональности, она предполагает умножение человеческой силы, главным образом в сфере умственной, но также и физической, и, прежде всего, не обременена биологическими ограничениями и перспективой смерти (*Przegalińska* 2016: 83).

Такое мнение близко трансгуманистическому проекту “апгрейда человечества”.

Тело и бестелесность ИИ

Мир “искусственных людей”, технологических существ издавна привлекал философов, например Декарта, Лейбница или Офре де Ламетри. Как пишет Пжегалинска, ключевой для развития ИИ была мысль Декарта о том, что разум не зависит от тела. “И несмотря на то что через сто лет появилась гораздо более механистическая, материалистическая и монистическая концепция Жюльена Офре де Ламетри, декартовская дуалистическая перспектива оказалась гораздо более плодотворной для развития технологий” (*Przegalińska* 2016: 29). Философская концепция человека-машины также уходит корнями во времена Декарта, который первым сравнил человеческое тело с машиной. В XX столетии, как замечает Радковска-Валькович, данное сравнение обрело другой смысл: человек-автомат – это человек дефектный, которому не хватает эмоциональности и индивидуализма. Однако на пороге XXI в. декартовское сравнение снова приобретает новый оттенок, и его оценка опять меняется: технологически улучшенные люди, например киборги, – это не только желаемый образ развития человечества в будущем (трансгуманисты), но и практические разработки, все чаще применяемые в медицинской, военной и других отраслях. Человек, совершенствуя себя, не бросает, однако, попыток создания Другого, с надеждой, что окажется творцом более удачливым, чем бог – создатель страдающего от болезней, старения, смерти и моральных пороков человечества. “В создание искусственного человека вписано недовольство формой человеческого бытия. Это также выражение своего рода культурного табу: разочарования и претензии к Богу за то, как выглядит наша жизнь”, – считает Радковска-Валькович (*Радковска-Валькович, Кожевникова* 2011). И еще: “Продукты технологии <...> должны быть совершеннее своих творцов – в противоположность божьему творению. Мы не хотим верить, что наше создание неполноценное. Мы предпочитаем себя обманывать” (*Radkowska-Walkowicz* 2008: 84). Это частично объясняет энтузиазм, который сопровождает новейшие достижения техники, даже вопреки многим голосам аргументированной критики из кругов видных интеллектуалов и ученых.

Значимой кажется мысль, что такие важные аспекты “разумного Другого”, как логическое мышление и язык, могут существовать вне тела — и не только биологического, но и любого техно-тела, т.е. в качестве виртуальной программы, бота. И хотя существуют многие воплощения ИИ — голограммы, аватары, разного вида машины и роботы, — он остается прежде всего виртуальным, трансцендентным проектом, что порождает прямые коннотации с божественностью. Более того, отсутствие тела с его сенсорными способностями и биологическими ограничениями, как мы постараемся доказать ниже, может иметь веские этические последствия.

Пжегалинска констатирует, что вместе с развитием кибернетики пришло мышление о машинах как о субъектах. На наш взгляд, такая идея всегда присутствовала в мышлении о машинах, хотя и не всегда была доминирующей. Свидетельством этого могут быть попытки создания антропоморфных машин, уходящие корнями в XVIII в. Тогда ключевым для размышления об их субъектности было подражание человеку, т.е. именно внешнее сходство (движения глаз, грудной клетки и т.д.), сегодня же субъектность заключается в другом — во внутреннем сходстве мыслительных и коммуникативных процессов, в интеллекте и — потенциально — в сознании. Для проявления такого сходства тело не нужно. Кинематографическим примером может послужить операционная программа Саманта из фильма “Она” (США, 2013). “Искусственный интеллект становится метафизическим проектом”, — делает вывод Пжегалинска (*Przegalińska* 2016: 98).

С одной стороны, отказ от антропоморфизма указывает на важные перемены в самом человечестве, несомненно, связанные с тем, что наш мир становится все более нематериальным, виртуальным, а виртуальная реальность во многих кругах считается уже не “вторичной”, а полноценной реальностью. С другой стороны, это приводит нас к некоей божественности и трансцендентности ИИ, который вполне может занять место в человеческом пантеоне богов, а это может иметь значительные последствия правовой и этической природы. Прежде чем развить эту мысль, мы должны проанализировать, что такое ИИ сегодня, как выглядит наше с ним взаимодействие и во что оно может перерасти в будущем.

ИИ сегодня и завтра

Проблемы с определением, что такое ИИ, начинаются уже на уровне терминологии.

Укоренившиеся в классической кибернетике (и в то же время являющиеся одним из главных драйверов цифровой революции), исследования в области искусственного интеллекта — это одна из самых увлекательных, сложных, противоречивых и в то же время влиятельных областей современной науки. Масштаб и разнообразие решаемых в них проблем заставляет нас думать о термине *artificial intelligence* как об общем названии, охватывающем ряд конкурирующих или даже противоречивых тенденций (*Ozóg* 2010: 48),

— пишет Ожуг, который предпочитает делить ИИ на *strong AI*, связанный с исследованием природы человеческого интеллекта и, таким образом, близкий философии, и на *Nouvelle AI* — прагматическое течение, концентрирующееся на отдельных проявлениях человеческого интеллекта и возможности подражания им и их внедрения. К *Nouvelle AI* он относит интеллектуальных агентов и интеллектуальные экспертные системы, которые присутствуют уже на глобальном уровне, а их влияние имеет как положительные, так и отрицательные стороны. Положительные связаны с прогрессом технологий и облегчением работы во многих сферах, отрицательные — с расширяющимся контролем и ограничением гражданских свобод.

Несмотря на некоторые различия, это разделение во многом перекликается с известной теорией “слабого” и “сильного” ИИ: первый исполняет служебную по отношению к человеку, инструментальную роль, в то время как второй имеет все признаки

личности — самостоятельное мышление и сознание. На данный момент все существующие ИИ принадлежат к “слабому” типу, однако именно создание “сильного” является основной проблемой в интересующем нас русле этики взаимодействия с искусственным интеллектом. Подходы к созданию “сильного” ИИ делятся в основном на функционалистские (*top-down*) и редукционистские (*bottom-up*) — симулирующие биологический мозг (физиологию нейронов и т.п.). По мнению многих специалистов, например известного польского когнитивиста Влодислава Духа, именно подражание нейроанатомии человеческого мозга, т.е. стремление к созданию нейроморфного ИИ, будет доминировать в этой области в будущем.

Термин “искусственный интеллект” чрезвычайно объемлен и многозначен, его толкования различаются не только на уровне специалистов и неспециалистов, но и внутри этих групп. То, что представляют в своих произведениях художники, — это суперинтеллект, чаще всего полностью автономный, сознательный и принимающий собственные решения. Большинство этических дебатов посвящено как раз этому виду ИИ, угрозам и вызовам, которые он несет в себе, несмотря на то что пройдет еще много десятилетий до его появления, если вообще он будет когда-либо создан.

Сегодня мы имеем дело прежде всего с тем, что специалисты иногда называют *Augmented Intelligence (AI)*, т.е. расширенным ИИ. Его главная задача — усилить способности человека, расширить его потенциал, облегчить и ускорить выполняемые им задачи или же вообще заменить простой, рутинный или опасный человеческий труд. Это часть ведущейся во многих сферах роботизации, в первую очередь в финансовом и банковском секторах, а также в области коммуникаций. Как пишет Анджей Собчак, основываясь на докладе 2017 г. компании “McKinsey” “Artificial intelligence: The next digital frontier?” (*Sobczak 2017*), сегодня именно эти направления — лидеры во внедрении ИИ. На противоположном крае находятся туризм, строительство, образование и большая часть сферы обслуживания. Умеренные успехи в области роботизации и внедрения ИИ достигнуты в торговле и медицине. Быстрыми темпами прогрессируют такие способности ИИ, как распознавание речи, изображений, перевод естественных языков, а также обработка и классификация данных. Именно поэтому многие профессии, например финансовый аналитик или переводчик, уже сегодня находятся под угрозой исчезновения.

Глобальное применение ИИ в различных сферах будет наращиваться; все больше стран, даже не являющихся “поставщиками” ИИ, задумывается над национальными стратегиями его развития, поскольку растет понимание необходимости таких стратегий в будущем. В связи с этим все чаще появляются вопросы взаимодействия людей и “разумных” машин/систем, этики, безопасности, ответственности последних, и даже их прав (напр., аналогично правам животных).

Критики и оптимисты — резюме плюсов и минусов

Безусловно, вокруг ИИ выросло много мифов, а создатели научно-фантастических текстов чаще всего показывают утрированный, крайний образ развития взаимодействия людей и ИИ в будущем, чем вызывают еще больше опасений у неспециалистов. За немногочисленными исключениями (напр., фильм “Она” 2013 г. или фильм “Когда йогурт захватил мир” из сериала “Любовь, смерть и роботы” 2019 г.) ИИ представлен как угроза всему человечеству. Иногда вина за такой поворот событий ложится на самого человека, но его создание — техноинтеллект — разрушает привычный нам мир. Скорее всего, именно поэтому в общественном сознании укрепились позиции скептиков или даже ярых противников ИИ.

Тем не менее среди специалистов достаточно много тех, кто не опасается прогресса в развитии ИИ и ожидает от него положительных изменений для мирового сообщества. К апологетам (и разработчикам) ИИ принадлежит Кайфу Ли, он считает технологии инструментом в руках людей, в итоге несущих ответственность

за результат действия “умных” машин и систем. Это мнение основано на том, что ИИ, применяя “глубокое обучение” (*deep learning*), использует данные, произведенные людьми. Здесь в качестве примера можно вспомнить чат-бота Тау компании “Microsoft”, запущенного в 2016 г., который, общаясь с пользователями интернета, быстро научился ругаться матом и высказывать расистские суждения, что вынудило “Microsoft” отключить его. Это иллюстрация ответственности создателя за свое создание и тест на нашу, человеческую, мораль. Нередко машина оказывается зеркалом, в котором можно увидеть собственные недостатки и пороки.

Для Кайфу Ли ИИ — это именно “умный помощник”, который вытеснит людей для их же блага из многих профессий. Во-первых, заменит их там, где работа несет угрозу жизни или здоровью человека, во-вторых, поможет там, где нужна большая точность, высокое техническое качество и быстрый анализ данных. Таким образом, как утверждает Кайфу Ли, ИИ “заберет у нас монотонный труд”, поскольку “ИИ может оптимизировать, но не может творить” (*Kai-Fu Lee 2018*). Людям принесет пользу то, что они будут частично заменены машинами на рабочих местах. (Такого же мнения придерживается один из ведущих разработчиков ИИ Себастьян Трун, который считает, что ИИ — это очередной инструмент в ходе эволюции, облегчающий наш повседневный монотонный труд.) Кайфу Ли утверждает, что благодаря ИИ люди перестанут быть трудоголиками, начнут наслаждаться жизнью, у них будет больше времени для хобби и творчества, а также для создания и укрепления социальных связей. Более того, останутся профессии, в которых люди будут незаменимы — те, которые требуют от нас творческого подхода или эмпатии.

На такую оценку воздействия ИИ на рынок труда в будущем, безусловно, повлиял тот факт, что Кайфу Ли — китаец (тайванец), а на его родине у большинства людей действительно остается крайне мало времени на частную жизнь. Это яркий пример того, как национальные и культурные различия сказываются на оценке взаимодействия человека с ИИ и сопутствующих рисков. И это происходит даже несмотря на то, что ИИ считается глобальным феноменом и над его разработкой нередко трудятся международные коллективы.

В духе Кайфу Ли высказывается и Том Грубер — разработчик ИИ (один из создателей *Siri*), который считает, что цель развития искусственного интеллекта заключается в наделении человека интеллектом компьютера и облегчении пользования вещами, а это также можно рассматривать в контексте практик технологического “улучшения человека” (*human enhancement*). Грубер воспринимает ИИ не как потенциально автономное существо, а лишь как инструмент в руках людей, который может быть на подчиненной или равной позиции в команде человек–машина. У такой команды появляются супервозможности, например, в медицинской диагностике или инженерном проектировании.

По мнению Грубера, ИИ — это технологическое усовершенствование (*enhancement*) как тела, так и интеллекта не только здоровых, но и больных людей. ИИ может помочь людям с ограниченными возможностями и даже страдающим деменцией в повседневной жизни. Грубер подчеркивает также роль ИИ в улучшении межчеловеческой коммуникации. Однако в отличие от Кайфу Ли, считающего, что ИИ вернет людям время для общения, Грубер видит в нем посредника в коммуникации, который восполнит некоторые недостающие функции. Таким образом, по мнению Грубера, ИИ будет играть преимущественно позитивную роль в жизни как здоровых, так и больных людей и окажет положительное влияние на социальную сферу.

К скептикам ИИ принадлежит известный философ Ник Бостром; он опасается того, что где-то в середине нашего столетия ИИ может превзойти аналитические способности человека, и тогда остановить его будет уже не в наших силах:

...предел скорости обработки информации машинами лежит далеко за пределами возможностей биологической ткани. Все сводится к физике. Биологический нейрон срабатывает примерно 200 раз в секунду, 200 герц. Но даже современный транзистор работает на гигагерцах. Нейроны медленно двигаются в аксонах, максимум 100 метров в секунду. А в компьютерах сигналы могут путешествовать со скоростью света. Есть также ограничения по размеру. Человеческий мозг обязан уместиться внутри черепа, а компьютер может быть размером со склад или даже больше (*Bostrom* 2015).

Таким образом, человечество перестанет быть главенствующим видом на Земле и безвозвратно утратит доминирующую позицию. Возможно, машинный интеллект будет “последним изобретением человечества”.

Более того, Бостром не говорит о “разумном” ИИ, аналогичном разумному человеку, не отсылает к возможному эмоциям или морали ИИ. Наоборот, он утверждает, что надо отбросить антропоморфизацию в мышлении об ИИ и понимать его как процесс оптимизации, настроенный на решение задач и достижение цели. При этом человечеству надо очень тщательно формулировать цели, чтобы не было различий между тем, что подразумевает человек, и тем, что понимает ИИ. Бостром приводит пример из мифологии: король Мидас пожелал, чтобы все, чего он касается, обращалось в золото, но не учел, что превратит в золото еду и даже своих близких. По мнению Бострома, эта история — урок для человечества, собирающегося создать сверхинтеллект.

Философ считает, что можно решить эту проблему. Даже несмотря на то что Бостром видит в ИИ угрозу, он не упускает из вида и выгоды от его создания, понимая, что развитие ИИ, скорее всего, уже не остановить. Бостром предлагает создать ограничения, которые гарантировали бы, что ИИ будет разделять наши ценности и в любой ситуации останется безопасным для человечества. Здесь (по аналогии со случаем Кайфу Ли) отчетливо прослеживается, что ученые и мыслители забывают, как неоднородно человеческое сообщество, когда говорят об общечеловеческих ценностях и целях, значительно различающихся в разных мировых культурах и обществах. На наш взгляд, именно эти культурные различия могут создать проблемы при реализации задачи, которую поставил перед разработчиками и специалистами по этике Бостром.

Более радикальную позицию занимает Сэм Харрис, который считает, что “в какой-то момент мы создадим устройства, которые будут умнее нас, и когда такие устройства появятся, они начнут улучшать себя сами. И тогда появится <...> риск утраты нами контроля над процессом” (*Harris* 2016). При этом не нужно, чтобы ИИ ненавидел людей или боролся с ними. По мнению Харриса, он будет относиться к людям примерно так же, как люди относятся к муравьям, поскольку его интеллектуальные способности намного выше человеческих и вскоре после создания ИИ разница между ним и нами будет непреодолимой:

Электронные схемы примерно в миллион раз быстрее биохимических, поэтому машина будет думать в миллион раз быстрее, чем создавший ее разум. Вы запускаете ее на неделю, и она выполняет 20 000-летнюю работу интеллекта человеческого уровня неделю за неделей. Как мы можем понять, не говоря о том, чтобы сдерживать, ум, работающий с подобной скоростью?

— задает вопрос Харрис (*Ibid.*). Он не пугает “бунтом машин”, а просто считает наши миры несовместимыми, при этом, в случае если человек ставит барьер для развития ИИ, ИИ просто сдвинет его, подобно тому, как мы строим дома, не обращая внимания на муравейники.

Кроме угрозы, исходящей от самого ИИ, существуют и другие, связанные с процессом его создания, источник которых — пороки человечества. Возрастающая автоматизация и замена человека на рабочих местах приведут к растущему экономическому неравенству и безработице, увеличению свободного времени и, как следствие, к потере смысла жизни и утрате интеллектуальных способностей многими людьми,

которым нечем будет заняться (здесь Харрис рассуждает совершенно иначе, чем оптимисты ИИ). Более того, если ИИ станет национальным проектом какого-нибудь одного государства, другие могут инициировать войну за доступ к ИИ, рассматриваемому как ценный ресурс.

Харрис считает, что процесс создания сверхинтеллекта уже не остановить и что он вполне реализуем. У него нет конкретных предложений, как справиться с задачей построения “безопасного” ИИ, он ограничивается общей фразой “создать его так, чтобы это совпало с нашими интересами”, где — опять же — интересы всего человечества понимаются как единые.

Прагматичную позицию по отношению к ИИ занимает Пжегалинска, которая считает, что технологии не бывают хорошими или плохими, они — инструменты в руках создателей и пользователей. Тем не менее для нее ИИ — это больше, чем только инструмент и “умный помощник”, это своего рода виртуальные существа, которые имеют “способность существовать на человеческий лад” (Przegalińska 2016: 203), т.е. подражают поведению человека и таким образом входят в интеракции с ним. В процессе этого взаимодействия со стороны человека возникает особенное психологическое и эмоциональное отношение.

Исходя из этого, Пжегалинска считает, что “намного важнее знания о том, как технология может служить человечеству или обществу, то, как она воспринимается” (Ibid.: 183). Исследовательницу волнуют не столько вопросы чисто (*stricte*) этического характера, а философский вопрос о *condicio humana* в отношении технологий. Для исключения злоупотребления технологиями со стороны государства, компаний или групп Пжегалинска предлагает сделать ИИ максимально демократичным и массовым.

На основе приведенных выше мнений экспертов можно утверждать, что ИИ несет в себе определенные риски для человека как в случае появления искусственного интеллекта в качестве “умного помощника”, так и в качестве “автономной личности”. В первом случае речь идет о таких угрозах, как усиленный контроль над частной жизнью, незащищенность личных данных, создание фейковых материалов, сокращение рабочих мест, постепенное купирование человеческих способностей за счет технологической поддержки. Во втором прогнозы скептиков более радикальны, вплоть до угрозы порабощения или гибели человечества в результате того, что созданный нами ИИ будет прогрессировать в темпе, недостижимом для нашего мозга.

Не существует согласия относительно принципиальной возможности появления когда-либо автономного, независимого и разумного ИИ. Тем не менее в соответствии с принципом, что этика технологий должна опережать вызовы, а не только реагировать на них, мнения скептиков нужно учитывать и анализировать, чтобы быть подготовленным к возникновению потенциальных угроз. Есть еще одна проблема этического характера: вполне возможно, что человечеству удастся создать что-то вроде разумного ИИ, который, однако, будет ограниченным и подконтрольным человеку в силу каких-либо технических условий. Этот вариант рассматривают специалисты по этике, занимающиеся, например, разработкой “прав роботов”, отчасти аналогичных правам животных.

Последствия бестелесности

Поскольку риски, связанные с ИИ в качестве “умного помощника” (объекта), уже активно обсуждаются, достаточно хорошо распознаны и классифицированы, а что еще важнее, поддаются регулятивным решениям, мы остановимся на тех угрозах, которые несет для человечества — пока только теоретически — автономный и независимый ИИ (субъект). В данном случае любопытно будет проанализировать и сравнить интересы и ценности человека и искусственного интеллекта, выявить связанные с этими интересами и ценностями варианты морали и этики, обусловленные спецификой технологического и биологического развития личности.

Сильно утрируя, можно сказать, что мораль либо основывается на осознанном выборе, т.е. выборе, сделанном на базе рационального мышления и различения добра и зла, либо же опирается на биологический “моральный инстинкт”. Последний, в свою очередь, выражается в эмпатии, обусловленной присутствием зеркальных нейронов и сенсорности – способности к ощущениям. Из этого следует, что есть два основных мнения об истоках человеческой морали: базирующееся на осознанном выборе и детерминированное биологией.

Наличие некоего “морального инстинкта”, доказанное исследованиями в области биологии, этологии и нейронаук, указывает на значение телесности для развития морали. То, что эмпатия носит нейробиологический характер, случайно открыл в 1992 г. Джакомо Риццолатти во время экспериментов на макаках. (Ученый регистрировал активность нейронов обезьяны в ответ на ее движения; внезапно обезьяна отреагировала на движение экспериментатора, как на свое собственное.) Зеркальные нейроны все еще продолжают изучать, однако уже известно, что их важнейшей функцией является распознавание настроений, мотиваций и чувств других.

Говоря языком нейробиологии, мы активируем в собственном мозгу нейронную презентацию моторной активности, аналогичной той, что видим или ожидаем у других. <...> Считалось, что мы принимаем решение проявить эмпатию и делаем это на основе сознательной симуляции: мы будто умышленно ставим себя на место другого человека и представляем, как чувствовали бы себя в его положении. Эмпатию рассматривали как когнитивный навык. Сегодня мы знаем, что этот процесс и проще, и автономнее. <...> Эмпатия появляется на основе бессознательных телесных связей, в которых задействованы лица, голоса и эмоции. Человек не может решить, быть ему эмпатичным или не быть; он эмпатичен по определению (*Вааль де 2014*),

– утверждает приматолог де Вааль.

В свою очередь, сенсорность можно определить как способность к ощущениям и положительного, и отрицательного характера (что может соответствовать квалиа [*qualia*] в когнитивистике, если понимать их вслед за Саймоном Блэкберном и Александрой Пжегалинской как осязаемые или феноменальные свойства сенсорных переживаний). Сенсорность, обусловленная чувствами, т.е. ощущениями тела, посредничает в познании мира и определяет характер взаимодействий с другими, что формирует нашу аффективность. Она связана с определенными участками мозга, без которых мы не можем получить представление об ощущениях – не можем чувствовать, как другой, имеющий эти ощущения. Например, на межвидовом уровне существует взаимная недоступность смысловых ощущений, информации и эмоций, которые они несут. Люди никогда не поймут, какие ощущения вызывает чувство эхолокации, видение в расширенном или, наоборот, в сокращенном световом диапазоне или четвероножность. Мы мыслим своим телом и весь мир воспринимаем в масштабах своего тела, его возможностей и ограничений. В итоге есть причины опасаться, что мы будем не в состоянии понимать небиологическую жизнь, пусть даже сознательную и разумную, в силу взаимной недоступности ощущений. Бестелесность ИИ (отсутствие сенсорности, опыта старения, заболевания и биологической смерти) ставит вопрос о его познавательных ограничениях, которые, среди прочего, перечеркивают способность к пониманию переживаний других.

Можно ли симулировать ощущения для систем, лишенных телесности, и как это повлияет на их восприятие? Поскольку это теоретически не исключено, то возникают очередные вопросы о подлинности и ценности таких ощущений, а следовательно, об их значении для развития “морального инстинкта”, т.е. среди прочего о способности со-чувствовать. Более того, если мы хотим единства ощущений с ИИ, мы должны проследить, чтобы его ощущения были “человеческими”, чтобы ИИ обладал чувствами и сенсорностью, типичными для *homo sapiens*.

Таким образом, мы можем утверждать, что главной проблемой на пути к единству интересов и целей человека и ИИ будет биологическая “бестелесность” последнего, поскольку большинство человеческих интересов ограничивается рамками нашей телесности и смертности.

Высший разум – высшая мораль?

Если, однако, мы решим, что ведущую роль в развитии морального поведения имеет осознанный выбор, то ИИ, как существо разумное, теоретически может быть также существом моральным. Более того, сверхинтеллект теоретически может означать сверхморальность. К сожалению для человечества, это не так просто. Прежде всего надо признать, что не существует общечеловеческой, универсальной этики: мировые культуры отличаются друг от друга иерархией ценностей, подходом к полам и возрастным группам, к меньшинствам, к работе и досугу, к правам индивида и общества. Опустим здесь правила, устанавливаемые разными мировыми религиями. Таким образом, то, что разработчики готовы внедрить в качестве моральных ориентиров в “разумные” программы в США, Китае или ОАЭ, будет сильно отличаться.

Более того, хотя философы спорят о природе добра, следуя утилитаристам и даже неокантианцам (Кристин Корсгиард), можно утверждать, что добро – не объективно и всегда “привязано” к кому-то, для кого имеет значение, т.е. нет “никому не принадлежащего” добра. Мы всегда смотрим на добро и интересы с чьей-то точки зрения: собственного я, своих семьи, народа, государства, религии и т.д. Сложно ожидать, что ИИ будет принимать во внимание все эти точки зрения хотя бы потому, что во многих случаях они взаимоисключаются. В связи с этим появляется проблема: как ИИ должен решить, чье добро приоритетнее. Опять же, ориентиры для такого решения, которые дадут ему разработчики, будут в значительной степени культурно обусловлены. Эта ситуация отчасти схожа с выбором критериев для принятия решения автономными “машинами без водителя”, где в качестве таковых рассматриваются, например, пол, возраст или даже экономический статус участников ДТП.

Нельзя забывать о том, что ИИ, как автономная личность, также будет иметь “свое добро”, т.е. свои собственные стремления и интересы, которым захочет следовать. До тех пор, пока ИИ ставит целью служение человеку, он не должен быть ориентирован на достижение своего блага. С перспективы человеческой этики, однако, это справедливо только до тех пор, пока ИИ не будет признан автономной личностью – субъектом морали. В качестве субъекта он имеет право действовать ради своего блага, отказывать ему в этом было бы насилием и дискриминацией. Риск возникает, когда интересы ИИ и человека (или человечества в целом) расходятся. То есть решением этой проблемы могло бы стать создание такого кодекса прав и обязанностей, который исключал бы ущемляющие действия с обеих сторон. Попытка создания такого кодекса, отдающего, однако, приоритет интересам человека, – это три закона робототехники Айзека Азимова – обязательные правила поведения для роботов, впервые сформулированные им в рассказе “Хоровод” (1942): 1) Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинен вред, 2) Робот должен повиноваться всем приказам, которые дает человек, кроме тех случаев, когда эти приказы противоречат Первому Закону, 3) Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму Закону.

Сложно поверить, что суперинтеллект будет оставаться подконтрольным человеку, имея доступ ко всемирному знанию и возможности самостоятельного действия. Кто может стать бенефициаром этики искусственного интеллекта? Люди, он сам, планета в целом? Можно предположить, что ИИ не будет антропоцентричным, что его этика будет, например, ИИ-центричной или вневидовой.

Варианты отношений с ИИ

В зависимости от того, будем ли мы считать ИИ объектом или субъектом, могут возникнуть различные этические проблемы нашего взаимодействия. Если ИИ будет иметь статус объекта, то этика ИИ будет концентрироваться на взаимодействии с людьми (клиентами, сотрудниками, пациентами и т.д.) и источниками информации (финансовыми, личными, секретными данными и др.), на видах знаний, загруженных в систему, и их обработке и применении: изоляции нежеланного (кем? по каким критериям?) знания, умения игнорировать нежеланное знание в процессе решения задачи. Появляются также вопросы ответственности за ошибки (самый обсуждаемый пример – автономный транспорт) и авторских прав (напр., в случае проектов, созданных алгоритмами). Данные этические проблемы будут касаться в первую очередь разработчиков ИИ и созданных ими алгоритмов.

В случае если ИИ будет иметь статус субъекта, проблемы будут концентрироваться на взаимном понимании, наделении правами и защите интересов – во что “втянуто” будет уже все человечество. В последние десятилетия расширяется понятие субъекта: это уже не только человек, но все чаще животное, киборг или гибрид, робот или искусственный интеллект. В западной философии существует сильное стремление к наделению правами и моральным статусом животных как не-человеческих субъектов. Все эти движения, относящиеся к биологическим, гибридным или технологическим существам, имеют много общего.

Если мы представим, что в будущем будет создан ИИ, обладающий полной автономностью, способный принимать самостоятельные решения и иметь собственные цели и интересы, которые могут не совпадать с нашими, то можно предположить различные варианты развития взаимодействия ИИ с человечеством – как дружеские, так и враждебные или нейтральные. Во-первых, это могут быть отношения, основанные на иерархии и подчинении, где главенствует либо человек, либо ИИ. Такая ситуация будет характеризоваться нестабильностью, поскольку подчиненные могут стремиться к изменению своего положения. Во-вторых, это могут быть отношения мирного сотрудничества и относительного равенства и, в-третьих – мирного сосуществования без сотрудничества, когда нет столкновения интересов человечества и ИИ.

Второй и третий варианты можно отнести к тому, что называют “дружественным ИИ” (*Friendly AI*). Данное понятие ввел Елиезер Юджовски, считающий, что ИИ не будет думать, как человек, и при этом сможет достичь всех поставленных задач, поэтому надо стремиться запрограммировать его таким образом, чтобы развивающийся и создающий новые “умные” программы ИИ реплицировал заложенную в нем дружественность к *homo sapiens*. Этот подход может показаться немного наивным, поскольку не учитывает как минимум два важнейших фактора, делающие эту задачу нереализуемой: во-первых, (как было уже многократно замечено выше) неоднородность человеческой морали и ценностей, а во-вторых, несовершенство разработчиков ИИ (как и всех остальных людей) в отношении моральности и дружественности. Между тем взгляды и предубеждения разработчиков, скорее всего, будут заложены в ИИ.

* * *

Перед лицом появления новых форм жизни – киборгов, умных машин, гибридов и т.п. – люди будут вынуждены разработать новый кодекс поведения по отношению к ним. Возможно, в нем будут прописаны неравные отношения, появится элемент власти и подчинения либо патриархальной зависимости. Возможно, мировой опыт продвижения прав животных окажется полезным для развития этики отношений человека и робота. Стоит прислушаться к интуиции художников, которые предостерегают нас от потери нашей гуманности. Тот факт, что в текстах культуры

не-люди все чаще оказываются более гуманными, чем представители *homo sapiens*, может свидетельствовать о том, что общество начинает задумываться о них, как о равноценных нам субъектах. В последнее время в кино и литературе особенно сильно прослеживается тенденция представлять человека как “бесчеловечного”, лишённого эмпатии творца, а его создания (андроидов, гибридов, живых машин, клонов и даже усовершенствованных животных) — со все более человеческими чертами. Примерами таких произведений могут быть фильмы: “Из машины” (*Ex Machina*), где создатель ИИ порабощает, насилует и убивает свои творения; “Не отпускай меня” (*Never Let Me Go*), снятый по одноименной книге, и “Остров” (*The Island*), в которых люди выращивают клонов, обеспечивая себя органами для трансплантации; “Искусственный разум” (*A.I. Artificial Intelligence*), где главный герой — робот-ребенок, нуждающийся в любви; “Я, робот” (*I, Robot*) — о машине, в которой проснулось сознание, и многие другие. Все эти произведения показывают нам, что мораль не должна ограничиваться узким кругом “своих”, она не должна быть моралью племенного типа. Современные люди, живущие в высокотехнологизированном и глобализированном мире, нуждаются в “апгрейде” своей этики до глобального уровня. В новой технонаучной ситуации, в которой человек оказывается не единственным разумным и чувствующим существом, наша этика устарела, и мы просто обязаны выйти за рамки видовых интересов. И действительно, это постепенно происходит.

И хотя на вопрос, можно ли чувствовать солидарность с существом, не принадлежащим к *homo sapiens*, по нашему убеждению, можно смело ответить “да”, то как быть с существами, которые способны погубить нас? Все мы грешим видовым шовинизмом, и сопротивление ему — это красивый с точки зрения этики и нравственности ход, если предметом нашей заботы являются виды слабее нашего (т.е. все, живущее ныне). Стоит ли пренебрегать видовой солидарностью в случае взаимодействия с опасными существами, например, с враждебно настроенным ИИ? А если он захочет ограничить свободу людей ради спасения экологии планеты?

Необходимый шаг на пути выстраивания отношений человека и ИИ — создание правил, уважающих интересы и потребности всех сторон. Включение ИИ в круг моральных субъектов потянет за собой также наделение их правами и обязанностями, которые в определенной степени могут выступать гарантом взаимного уважения и безопасности. Конечно, хорошим решением было бы заложение в ИИ “нашей” иерархии ценностей, однако для этого необходимо прийти к согласию относительно того, какие ценности мирового сообщества действительно могут считаться универсальными или хотя бы приемлемыми для всех.

Проблема отсутствия тела и связанных с ним наслаждений и ограничений, по нашему мнению, будет одной из главных помех в достижении общих целей, а значит, и взаимопонимания людей и искусственного интеллекта.

Примечания

¹ Здесь необходимо отметить, что три (а если включать Нулевой закон, сформулированный Азимовым позже, даже четыре) закона робототехники не имеют практического значения для развития ИИ и, несмотря на свою успешную “карьеру”, они много критиковались. Их применению на практике мешает несколько факторов, которые отметили, кроме самого Айзека Азимова, Дэвид Лэнгфорд, Роберт Сойер, Марк Тильден и другие. Прежде всего, роботы могут причинить вред человеку, не осознавая этого, могут захватить власть ради “спасения” человека от него самого; они могут иметь проблемы с определением “человека”, вдобавок ко всему их создатели и главные заказчики — армия или большие компании, которые не будут заинтересованы в соблюдении законов, даже наоборот. Кроме того, сегодня важной проблемой применения законов является заложенное в них подчинение потенциального морального субъекта другому субъекту — человеку.

Источники и материалы

- Левин 2013 – Левин Ж. Функционализм. Стэнфордская энциклопедия философии (версия осени 2013 года) / Ред. Э.Н. Залта; пер. с англ. А.В. Кузнецова. <https://philosophy.3-core.ru/sep/functionalist> (дата обращения: 10.04.2019).
- Bostrom 2015 – Bostrom N. What Happens when Our Computers Get Smarter than We Are? // TED. March 2015. https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are
- Harris 2016 – Harris S. Can We Build AI without Losing Control Over It? // TEDSummit. 2016. https://www.ted.com/talks/sam_harris_can_we_build_ai_without_losing_control_over_it/transcript#t-165096
- Kai-Fu Lee 2018 – Kai-Fu Lee. How AI Can Save Our Humanity // TED. April 2018. https://www.ted.com/talks/kai_fu_lee_how_ai_can_save_our_humanity
- Sobczak 2017 – Sobczak A. Branże o najwyższym stopniu wykorzystania AI – stan obecny i prognoza na 3 lata // Robonomika. 19 Listopad 2017. <https://robonomika.pl/branze-o-najwyzszym-stopniu-wykorzystania-ai-stan-obecny-i-prognoza-na-3-lata>

Научная литература

- Вааль де Ф. Истоки морали. В поисках человеческого у приматов. М.: Альпина нон-фикшн, 2014.
- Радковска-Валькович М., Кожевникова М. Искусственный человек в образах литературы и кино (интервью) // Медицинская антропология и биоэтика. 2011. № 1 (3). http://www.medanthro.ru/?page_id=1076
- Gawkowski K. Cyberkolonializm: Poznaj świat cyfrowych przyjaciół i wrogów. Gliwice: Helion, 2018.
- Ożóg M. Nauka w (krzywym) zwierciadle sztuki: problematyka sztucznej inteligencji w twórczości Kena Feingolda i Rafaela Lozano-Hemmera // Digitalne dotknięcia. Teoria w praktyce/Praktyka w teorii / Red. P. Zawojski. Szczecin, 2010. S. 43–63.
- Przegalińska A.K. Istoty wirtualne. Jak fenomenologia zmieniała sztuczną inteligencję. Kraków: Universitas Kraków, 2016.
- Radkowska-Walkowicz M. Od Golema do Terminatora. Wizerunki sztucznego człowieka w kulturze. Warszawa: WaiP, 2008.

Research Article

Kozhevnikova, M., and S.V. Karpova. Artificial Intelligence: Subject and Object [Iskusstvennyi Intellect: sub'ekt i ob'ekt]. *Etnograficheskoe obozrenie*, 2020, no. 1, pp. 80–94. <https://doi.org/10.31857/S086954150008759-4> ISSN 0869-5415 © Russian Academy of Sciences © Institute of Ethnology and Anthropology RAS

Magdalena Kozhevnikova | <https://orcid.org/0000-0001-7484-9638> | kmagdalena@yandex.ru | Institute of Ethnology and Cultural Anthropology, University of Warsaw (4 Żurawia ul., 00-503 Warszawa, Poland)

Svetlana Vasil'evna Karpova | <https://orcid.org/0000-0002-4488-8131> | SVKarpova@fa.ru | Financial University under the Government of the Russian Federation (49 Leningradsky Prospekt, Moscow, 125993, Russia)

Keywords

Artificial Intelligence, AI, ethics, corporality, technology

Abstract

This article is an ethical-cultural analysis of human interaction with Artificial Intelligence (AI). We treat AI as a “smart assistant” and as a potential autonomous personality. In the first case, we are talking about such threats to humans as tightened control over privacy, insecurity of personal data, fake news propagation, job cuts, and gradual devaluation of human abilities caused by the technological

progress. In the second case, we are talking about the threat of enslaving humanity by AI or death of humanity resulting from the AI progress at a pace unattainable for the human brain. The article discusses options of relations with AI, depending on whether it is going to be recognized as an object or subject of morality. The article analyzes and compares interests and values of a person and AI, identifies variants of morality and ethics associated with them, determined by the specifics of technological and biological development of a personality. Proceeding from the fact that there are two major opinions on the origins of human morality – one based on conscious choice, and the other determined by biology – we argue that the main problem on the path to humans' and AI's unity of interests and goals will be AI's biological "incorporeality", since most human interests are limited to physicality and mortality. We also assume that AI is not going to be anthropocentric, and its ethics may be AI-centric or extra-specific.

References

- Gawkowski, K. 2018. *Cyberkolonializm: Poznaj świat cyfrowych przyjaciół i wrogów* [Cybercolonialism: Discover the World of Digital Friends and Enemies]. Gliwice: Helion.
- Ożóg, M. 2010. Nauka w (krzywym) zwierciadle sztuki: problematyka sztucznej inteligencji w twórczości Kena Feingolda i Rafaela Lozano-Hemmera [Science in the (Crooked) Mirror of Art: Problems of Artificial Intelligence in the Work of Ken Feingold and Rafael Lozano-Hemmer]. In *Digitalne dotknięcia. Teoria w praktyce/Praktyka w teorii* [Digitally Touching: Theory in Practice/Practice in Theory], edited by P. Zawojcki, 43–63. Szczecin.
- Przegalińska, A.K. 2016. *Istoty wirtualne. Jak fenomenologia zmieniła sztuczną inteligencję*, [Virtual Beings: How Phenomenology Changed Artificial Intelligence]. Kraków: Universitas Kraków.
- Radkowska-Walkowicz, M., and M. Kozhevnikova. 2011. Iskusstvennyi chelovek v obrazah literatury i kino (interv'iu) [Artificial Man in Literature and Cinema Characters (Interview)]. *Medycinskaia antropologiya i bioetika* 1 (3). http://www.medanthro.ru/?page_id=1076
- Radkowska-Walkowicz, M. 2008. *Od Golema do Terminatora. Wizerunki sztucznego człowieka w kulturze* [From Golem to Terminator: Images of Artificial Man in Culture]. Warszawa: WaiP.
- Waal de, F. 2014. *Istoki moralii. V poiskah chelovecheskogo u primatov* [The Origins of Morality. Finding Human in Primates]. Moscow: Al'pina non-fikshn.