

МАТЕРИАЛЫ КРУГЛОГО СТОЛА
«АКТУАЛЬНЫЕ ПРОБЛЕМЫ НЕЙРОЭТИКИ»
(30 ОКТЯБРЯ 2019 Г.)

DOI: 10.17323/2587-8719-2020-1-135-167.

30 октября 2019 г. в Московском кампусе НИУ ВШЭ состоялся круглый стол «Актуальные проблемы нейроэтики», организованный Школой философии НИУ ВШЭ и Научным советом при Президиуме РАН по методологии искусственного интеллекта (НСМИИ). Круглый стол был приурочен к 55-му заседанию Международного междисциплинарного семинара «Нейрофилософия».

УГЛЕВА АНАСТАСИЯ ВАЛЕРЬЕВНА¹:
ВСТУПИТЕЛЬНОЕ СЛОВО

Дорогие коллеги, данный круглый стол мы решили посвятить обсуждению предметного поля и актуальных вопросов нейроэтики — дисциплины, начавшей формироваться в зоне пересечения двух междисциплинарных исследовательских полей: с одной стороны, изучения и разработки этических стандартов научных исследований мозга человека (в этом смысле нейроэтика предстает разновидностью биоэтики); а с другой стороны, изучения мозговой деятельности как основы осуществления этических способностей человека (в таком качестве нейроэтика является направлением нейронауки). Как известно, нейронауки возникли в дополнение и в развитие классических исследований в области неврологии и нейрофизиологии. Не случайно Леон Касс определяет нейроэтику как дисциплину, которая «исследует этические последствия прогресса нейронаук и нейропсихиатрии»². Примечательно, что нейроэтика не является творением рук философов, она естественным образом возникла внутри самой нейронауки. Впервые этот термин употребил Вильям Сафир на конгрессе в Сан-Франциско в 2002 году. Однако за пару десятков лет до этого философы постепенно стали осмысливать новую область знания на стыке когнитивной науки

¹Кандидат философских наук, PhD, доцент, руководитель Школы философии НИУ «Высшая школа экономики», Москва.

²Session 3 : Neuroscience, Neuropsychiatry and Neuroethics: An Overview / The President's Council on Bioethics. — 2004. — URL: <https://info.publicintelligence.net/GlobalTrends2030.pdf/> (visited on Feb. 2, 2020).

и философии. Так, Патрисия Черчленд в 1986 г. опубликовала монографию «Neurophilosophy: Toward a Unified Science of the Mind-Brain», в которой ее предмет обозначила как классическую для философии проблему соотношения сознания и тела. Но уже в 1992 г. она же привносит в нее психонейрофармакологическую проблематику, что особенно четко прослеживается в другой коллективной монографии под общей редакцией Черчленд «Neurophilosophy and Alzheimer's Disease». Нейрофилософия в самых общих чертах занялась поиском естественных оснований моральности поведения человека, что постепенно и выделилось в самостоятельную сферу исследований — «мета-этическую биологию» или «нейро-этику»³. Что касается отечественных исследований психики и мозга, то уже с конца 1960-х гг. в предметное поле этики начали активно вплетаться теоретические обобщения генетики, математики, кибернетики, искусственного интеллекта, а с начала 2000-х гг. — и социально-психологические исследования феномена обмана, свободы воли, разработка алгоритмов обучения формальных нейронных сетей, т. е. алгоритмов машинного самообучения, дополняющиеся рефлексивными алгоритмами «самообмана», и многое другое. Такого рода исследования имеют особое значение, например, для развития интеллектуальной робототехники и банковского дела, в котором алгоритмы переобучения формальных нейронных сетей призваны обособивать, например, доверительный уровень кредитования. Технологии нейроинтерфейсов и нейромаркетинговые исследования подняли вопросы о повышении личной ответственности и правового регулирования не только сознательного, но и подсознательного поведения человека. Проекты «чтения мозга», сужающие границы человеческой приватности, выдвинули требования детальной спецификации нейротехнологической деятельности вместе с ужесточением этико-правовых запретов. Таким образом, нейроэтика не является исключительной прерогативой специалистов в сфере этики. Что обусловило наше желание привлечь обсуждению предметного поля нейроэтики не только философов, но и практиков. Мы предлагаем вам высказаться относительно того, что представляет собой нейроэтическое знание, каков предмет нейроэтики и можно ли считать ее наукой наравне с нейрофизиологией, нейробиологией, нейролингвистикой и т. п.

³ *Miles J. Global Trends 2030 : Alternative Worlds / National Intelligence Council. — 2012. — URL: <https://info.publicintelligence.net/GlobalTrends2030.pdf/> (visited on Feb. 2, 2020).*

РАЗИН АЛЕКСАНДР ВЛАДИМИРОВИЧ⁴:
ВОЗМОЖНОСТИ И ПРЕДЕЛЫ НЕЙРОЭТИКИ

В последнее время нейроэтика получает бурное развитие. Ее основными утверждениями, являются тезисы о том, что моральное поведение целиком зависит от строения мозга, гормональной регуляции (в частности — действия гормонов дофамина и окситоцина), способности к эмпатическому отношению, которую мы наследует от животного мира. Последнее иногда связывается с так называемыми зеркальными нейронами (Ризолатти). Что касается зависимости морального поведения от строения мозга, то здесь я могу сослаться на выступление Патриции Чёрчленд на круглом столе, посвященном нейрофилософии, на философском факультете 18 июня 2015 г. Она приводила пример с полевыми мышами, одна разновидность которых расстается сразу после спаривания, а другая — создает стабильные пары. Все это объяснялось через различия в строении полушарий их мозга. Но для того, чтобы любая способность, в том числе моральная, была эволюционно закреплена, подтверждена на уровне мозговых процессов в виде развития особых отделов мозга или специфических гормонов, необходимо предшествование повторяющегося поведения в ходе жизни многих поколений. Иначе говоря, чтобы выработался гормон, отвечающий, например, за заботу о потомстве, кто-то должен начать заботиться о своем детеныше на каком-то ином, но достаточно устойчивом основании. На мой вопрос Патриции Чёрчленд относительно основания повторяющегося устойчивого поведения заданный после ее выступления, она ответила, что это определяется генетически. Но в таком случае получается круг в рассуждении: сами мозговые структуры, обуславливающие альтруизм и заботу о потомстве фиксируются на генетическом уровне и, соответственно, генетически передаются от поколения к поколению. И генетически же обусловлено предшествующее такому эволюционному закреплению повторяющееся поведение. Понятно, что здесь требуется какое-то иное объяснение. На мой взгляд, эволюционно развиваются сначала не те структуры, которые отвечают за заботу о потомстве, или высоко развитые формы взаимной помощи, а другие психические механизмы, обусловленные задачами ориентации, практического приспособления к действительности в смысле предвосхищения ее будущих состояний, опережающего отражения. Речь идет об опережающем отражении на

⁴Доктор философских наук, профессор, заведующий кафедрой этики философского факультета МГУ имени М. В. Ломоносова, Москва.

основе идеальных образов. Способными к развитым формам опережающего отражения на основе идеального образа, а соответственно и к высшим формам эмпатии, оказываются в основном те животные, которые проходят так называемый зеркальный тест, т. е. узнают себя в зеркале. Это говорит о том, что они в какой-то степени обладают самосознанием. Для того чтобы понять этот механизм, необходимо кратко охарактеризовать природу идеального, в нашем случае — способа ориентации на основе идеального образа. Идеальное это отнюдь не копия действительности, а способ ее опережающего отражения, основанный на том, что в сигналах головного мозга разыгрывается возможное будущее состояние реальности, отнесенное к определенному моменту времени. Из этих возможностей выбирается наиболее вероятное, которое и становится основанием для осуществления практического действия. Например, если волк охотится за зайцем, он должен предположить возможные движения жертвы, оценить расстояние (оптимальное для осуществления прыжка) до зайца, при этом надо еще оценить рельеф местности, скорость движения в данных условиях и т. д. И как это ни парадоксально, для того чтобы осуществить подобную интерпретацию мира, необходимо поставить себя на место другого, т. е. в нашем примере — представить себя зайцем. Это сложный процесс работы психики, который не сводится к действию зеркальных нейронов. В докладе А. М. Сурина «Мозг культурный и некультурный: о культурах нервной ткани» (в Институте им. Сеченова 23.10.2019) говорилось о том, что процесс восприятия внешнего мира, в том числе реакция на другого выглядит куда сложнее работы зеркальных нейронов. Для переработки поступающей информации подключаются новые и новые нейронные сети, которые не являются постоянными. Возникает вопрос: причем здесь мораль? Разумеется, во время охоты, когда есть доминирующая цель ни о какой морали речи идти не может. Психика, также как и сознание по своей природе эгоистична. Но сознание не всегда находится в актуальном состоянии. Оно может быть приглушено во время отдыха или сна, и вот тут-то мозг независимо от желания особи начинает производить моральные представления. Из того, что ты на месте другого, ты начинаешь чувствовать как другой. Тем более это относится к ситуации, когда вполне практически обусловленные механизмы психики начинают распространяться на отношение к особям своего вида, своей группы.

Таким образом, мораль в своих исходных основаниях является как бы неразстворимым остатком вполне практического, эволюционно обусловленного эгоистического поведения. Это проявляется как в плане

эмпатического отношения, связанного с утешением, так и в более развитой форме, основанной на сострадании и программно-целевой помощи. Но далее, конечно, эволюционно могут закрепляться и такие структуры мозга и гормональной регуляции, которые отражают постоянство в отношениях с членами своей группы. Собственно говоря, попытки вывести мораль из каких-то биологических предпосылок не являются совершенно новыми. В работах П. А. Кропоткина «Этика», «Взаимопомощь как фактор эволюции», «Взаимопомощь среди животных и людей» приводятся примеры весьма примечательного поведения животных, при котором они демонстрируют подлинный альтруизм: сильные заботятся о слабых, попавшим в беду помогают другие и т. д. В современной этике подход, связанный с тем, что основы нашего социального поведения мы наследуем от животного мира получает значительное развитие. Аргумент Франса де Вааль заключается в том, что альтруистическое поведение доставляет приматам удовольствие. Поэтому надо пересмотреть всю концепцию морали, отказаться от ее принудительно-императивной трактовки (концепция лакировки) и рассматривать человека как существо сверхкооперирующееся, входящее в социальную жизнь без всяких усилий. Однако следует заметить, что такой подход, несмотря на всю его привлекательность, имеет свои ограничения. Они связаны с таким этапом культурного развития, когда человек начинает ставить перед собой сверхзадачи, когда он в каком-то смысле начинает насиловать свою собственную природу, вовлекая себя во все более и более развитые и все более интенсивные виды социальной деятельности. В результате он в некотором смысле начинает действовать против эволюции, ибо подобная интенсификация не требуется ни для производства полноценного потомства, ни для выживания в условиях традиционной природной среды. Индивид вовлекается в интенсивные виды деятельности в результате тенденций, проявляющих себя в культуре как новой самостоятельной сущности. Поэтому интенсивное развитие навязывается индивиду именно извне, в том числе — за счет императивно-принудительного воздействия морали. В эволюционном смысле это сказывается на биологической организации человека, способствует развитию более совершенной энергетики, потреблению меньших ресурсов, чем это имеет место в процессе жизнеобеспечения животных организмов, и запаса значительного количества так называемой свободной энергии, накапливаемой сверх того, что необходимо для выполнения конкретной мышечной работы (И. А. Аршавский). Это

определяет большую продолжительность жизни человека, по сравнению с жизнью близких ему по весу животных. Понятно, насколько это необходимо для обеспечения условий передачи социального опыта, как за счет этого развивается морально обусловленная забота о старшем поколении.

ДЕНИСОВ ЭДУАРД ИЛЬИЧ⁵:

НЕЙРОЭТИКА И ИНФОРМАЦИОННАЯ ГИГИЕНА

Указом Президента Российской Федерации от 10 октября 2019 г. №490 утверждена «Национальная стратегия развития искусственного интеллекта на период до 2030 года». В соответствии с ними будут внесены изменения в национальную программу «Цифровая экономика РФ».

Новые технологии несут блага человеку, но и новые факторы риска. Это делает актуальным анализ сущности цифровизации экономики и общества с позиций становления парадигмы информационной гигиены (ИГ), а также этики, правового и нормативно-технического регулирования как основы гигиенической регламентации кибер-физических систем (КФС) и технологий, основанных на искусственном интеллекте (ИИ).

Интересны воззрения корифеев естествознания и медицины по этико-философским вопросам общества, труда и гигиены. Отец медицины труда Б. Рамаццини (1700) среди трех причин болезней работников упомянул «внимание и применение ума»⁶. Дени Дидро, по свидетельству Ш. Леви, отмечал, что «любой вопрос морали есть также вопрос гигиены». Сам Ш. Леви заметил: «Гигиена есть клиника здорового человека». Русский врач А. Никитин (1847) писал: «Каждое усовершенствование в искусствах и промышленности было куплено ценою телесного здоровья»⁷. Основатель современной гигиены Ф. Ф. Эрисман (1872) считал целью гигиены «найти средства для смягчения действия всех неблагоприятных для организма человека условий со стороны природы и общества»⁸.

⁵ Доктор биологических наук, профессор, главный научный сотрудник ФГБНУ НИИ медицины труда имени академика Н. Ф. Измерова, Москва.

⁶ *Рамаццини Б.* О болезнях ремесленников : Рассуждения / пер. с лат. В. О. Горенштейна. — М. : Медгиз, 1961.

⁷ *Никитин, А. Н.* / Болезни рабочих с указанием предохранительных мер, описанные доктором медицины Александром Никитиным. — СПб. : Типография Эдуарда Праца, 1847.

⁸ *Эрисман, Ф. Ф.* / Руководство к гигиене, обработанное по лучшим и современным сочинениям д-ром мед. Ф. Эрисманом. В 3 т. Т. 1. — СПб. : Печатня Головина, 1872.

Русский философ Вл. Соловьев (1897) отмечал, что «этика есть гигиена, а не терапия духовной жизни»⁹. Его духовный последователь В. И. Вернадский (1926) считал, что «под влиянием научной мысли и человеческого труда биосфера переходит в новое состояние — ноосферу»¹⁰.

Таким образом, этика и гигиена важны для здоровой духовной жизни общества в эпоху цифровизации.

Из 23 Асиломарских принципов безопасной, продуктивной и нравственной разработки систем ИИ (2017) — 7 принципов могут служить основой для гигиенической оценки ИИ и КФС. Важны вопросы правового регулирования, в частности, проект ЮНЕСКО, резолюция Европарламента, а также проекты законов Великобритании, США и др. Росстандартом созданы ТК 164 Искусственный интеллект и ТК 194 Кибер-физические системы.

Еще в 2012 г. Российская академия медицинских наук провела сессию по нейронаукам для медицины. ФГБНУ НИИ МТ совместно с ИМАШ РАН проводит исследования по новому направлению профилактической медицины — ИГ¹¹. Разработаны глоссарий, методические рекомендации, интерактивный директорий-справочник «Профессиональный риск»¹².

Регулирования информации вплоть до запрещения важно для уязвимых групп населения (подростков, беременных и кормящих грудью женщин, хронических больных, мигрантов и др.)¹³ и особенно детей¹⁴. Наряду с разработкой ГОСТов по ИИ и КФС необходима гигиеническая регламентация для управления рисками здоровью работников и населения.

⁹Соловьев В. С. Оправдание добра / под ред. О. А. Платонова. — М. : Институт русской цивилизации, Алгоритм, 2012.

¹⁰Вернадский В. Научная мысль как планетное явление / под ред. А. Л. Яншина. — М. : Наука, 1991.

¹¹Бухтияров И. В., Денисов Э. И., Еремин А. Л. Основы информационной гигиены : концепции и проблемы инноваций // Гигиена и санитария. — 2014. — Т. 93, № 5. — С. 5–9; Денисов Э. И. Роботы, искусственный интеллект, дополненная и виртуальная реальность : этические, правовые и гигиенические проблемы // Гигиена и санитария. — 2019. — Т. 98, № 1. — С. 5–10.

¹²Профессиональный риск : директорий-справочник / Одобрено Бюро ОПМ РАМН. — 2011. — URL: <http://medtrud.com/> (дата обр. 2 февр. 2020).

¹³Денисов Э. И., Еремин А. Л., Сивочалова О. В. Информационная гигиена и регулирование информации для уязвимых групп населения // Гигиена и санитария. — 2014. — Т. 93, № 5. — С. 43–49.

¹⁴Кучма В. Р., Сухарева Л. М., Храмов П. И. Гигиеническая безопасность жизнедеятельности детей в цифровой среде // Здоровье населения и среда обитания. — 2016. — Т. 281, № 8. — С. 4–7.

Мир становится все более технологичным и все менее этичным и духовным, в нем хрематистика берет верх над экономикой (термины по Аристотелю), т. е. извлечение прибыли преобладает над удовлетворением общественных потребностей¹⁵.

Это побуждает гигиенистов к упреждающим действиям по выявлению и оценке новых рисков от бурно развивающихся технологий, что аналогично смене парадигмы в медико-биологических науках¹⁶: замене концепции гомеостаза новой концепцией — аллостаза, как модели предиктивной регуляции¹⁷.

Известно высказывание К. Маркса (1845) о том, что «философы лишь различным образом объясняли мир; но дело заключается в том, чтобы изменить его»¹⁸. Изложенное выше дает дополнительные медико-биологические инструменты сохранения и укрепления здоровья, в т. ч. душевного, в цифровую эпоху¹⁹.

ПЕТРОВА ЕКАТЕРИНА ВИКТОРОВНА²⁰:

РАЗВИТИЕ ЭМОЦИОНАЛЬНОГО ИНТЕЛЛЕКТА В ЦИФРОВОЙ СРЕДЕ
КАК ПРОБЛЕМА НЕЙРОЭТИКИ

Для меня, как исследователя проблемы взаимодействия человека и современной информационной (цифровой) среды, проблемные поля нейроэтики и нейрофизиологии представляют интерес, прежде всего, в связи с вопросами корреляции между активностью нашего мозга, различными типами обработки информации и отражением этой проблематики в культуре.

¹⁵Юдина Т. М. Осмысление цифровой экономики // Теоретическая экономика. — 2016. — № 3. — С. 12–16.

¹⁶Денисов Э. И., Пфаф В. Ф., Степанян И. В., Горохова С. Г. Сдвиг медико-биологической парадигмы : от гомеостаза к аллостазу // Нейрокомпьютеры : разработка и применение. — 2016. — № 2. — С. 16–21.

¹⁷Sterling P. Allostasis : A Model of Predictive Regulation // Physiology and Behavior. — 2012. — No. 106. — P. 5–15.

¹⁸Маркс К. Тезисы о Фейербахе // Собрание сочинений. В 50 т. / К. Маркс, Ф. Энгель ; пер. с нем. Ю. И. Айхенвальда. — М. : Государственное издательство политической литературы, 1955. — С. 1–4.

¹⁹Denisov E. I. Informational Hygiene as a Medical and Biological Tool for Health Preservation in the Digital Era // American Journal of Biomedical Science and Research. — 2019. — Vol. 4, no. 2. — P. 73–75.

²⁰Кандидат философских наук, ведущий научный сотрудник Института философии РАН, Москва.

Опираясь на концепцию Вяч. Вс. Иванова²¹, можно предположить, что у представителей древних дописьменных обществ было преимущественно развито образное мышление и, соответственно, правое полушарие (распространенная в таких обществах традиция устной передачи народного творчества свидетельствует о преобладании правополушарной памяти, организованной иным образом, чем память левого полушария). У человека дописьменного общества логическое мышление (прерогатива левого полушария) было менее развито, чем у современного человека. Появление фонетического письма стимулировало развитие логического мышления и, соответственно, активизацию левого полушария. С начала XX в. мы наблюдаем постоянное возрастание количества образной информации — кино, телевидение, видеофильмы, наконец, компьютер и Интернет. Информационный ряд компьютерных игр, которыми увлекаются и взрослые, и дети, почти полностью состоит из образов. Соцсети также предоставляют большой процент информации в виде образов. Возрастание объема образной информации может способствовать активизации правого полушария, что позволяет предположить: «цифровой» человек будет в большей степени «правополушарным». М. Маклюэн писал о том, что будущая всемирная цивилизация превратится в общество «гармоничной коммуникации» и «образного мышления»²².

При всей неоднозначности влияния на человека современной цифровой среды, несомненно одно: применение компьютерных технологий существенно преобразует когнитивную, мотивационную и эмоциональную составляющие человеческой личности. Скорее всего, многие когнитивные функции (способность хранить в памяти большие объемы информации, планирование, моделирование, «вычислительный» интеллект) будут все больше «отдаваться на откуп» технологиям.

Но пока человек остается человеком, он обладает способностями, которые, по крайней мере в настоящее время, машины заменить не могут. Я имею в виду, прежде всего, так называемый эмоциональный интеллект (emotional intelligence, EI). Впервые понятие «эмоциональный интеллект» ввели П. Сэловей и Дж. Майер в 1990 г. и определили его

²¹ *Иванов В. В.* Чет и нечет : Асимметрия мозга и знаковых систем. — М. : Советское радио, 1978.

²² *Маклюэн М.* Галактика Гутенберга : Становление человека печатающего / пер. с англ. И. О. Тюриной. — М. : Академический проект, 2018.

как совокупность ментальных способностей, обеспечивающих осознание и понимание собственных эмоций и эмоций окружающих людей²³.

Человек, обладающий высоким уровнем эмоционального интеллекта, зачастую более успешен в профессиональном плане и в плане построения межличностных отношений. Это обусловлено тем, что он лучше понимает эмоции, свои и окружающих, что помогает ему быть более адаптивным и эффективным в общении. Гармоничные отношения как с внешним миром, так и со своим внутренним, способствуют душевному здоровью и повышению уровня осознанности и самопознания. Даже если человек обладает не слишком высоким IQ, но при этом развитым эмоциональным интеллектом, он может быть очень успешен как профессионал и востребован как личность, так как умеет мотивировать и вдохновлять себя самого и окружающих, выстраивать хорошие отношения с коллегами, креативно решать задачи.

В структуре эмоционального интеллекта принято выделять три компонента — эмоциональный, когнитивный и поведенческий. При этом эмоциональный компонент выражается в эмпатии, в том, что человек способен свободно выражать свои эмоции, а не держать их при себе, он эмоционально отзывчив, правильно интерпретирует эмоции других людей, способен сопереживать их радости и горю.

В рамках беседы о нейроэтике мне представляется весьма интересным вопрос о корреляции между высоким уровнем эмоционального интеллекта и доминирующей деятельностью того или иного полушария человеческого мозга. Согласно данным, которые приводит нейрофизиолог В. Л. Деглин²⁴, активация правого полушария сопровождается усилением эмоциональной экспрессии.

Под эмоциональной экспрессией принято понимать внешнее выражение собственных эмоций и чуткость к выражению эмоций окружающих людей. Можно сделать вывод, что по крайней мере один компонент ЭИ — эмоциональный — можно связать с активизацией правого полушария.

Конечно, влияние преобладания того или иного полушария на эмоции человека, да и на другие проявления его личности, не нужно переоценивать. Сам Деглин характеризует проблему связи функциональной

²³Emotional Development and Emotional Intelligence : Educational Implications / ed. by P. Salovey, D. Sluyter. — New York : Perseus Books Group, 1997.

²⁴Деглин В. Л. Лекции о функциональной асимметрии мозга человека. — Амстердам, Киев : Ассоциация психиатров Украины, 1996.

асимметрии мозга и эмоций как очень сложную. Но мне кажется, что определенную осторожную корреляцию можно наметить.

В связи с развитием искусственного интеллекта, нейросетей, идеологии трансгуманизма в новом свете предстает проблема «расчеловечивания» («дегуманизации») человека и «порабощения» его машинами. Но французский философ Мишель Серр, напротив, видит в масштабных технологических и культурных преобразованиях цифровой эпохи «человечение», а не «расчеловечивание»²⁵. Как мне представляется, один из путей для человечества сохранить в себе «человеческое» — развивать такие способности, в которых машины, по крайней мере пока, не могут нас превзойти. Одной из таких способностей как раз и может стать эмоциональный интеллект, высокий уровень которого, связан, в том числе, и с развитием правого полушария. Логично предположить, что возрастающее количество образной информации в современной цифровой среде будет этому способствовать. Конечно, в проблеме межполушарной асимметрии мозга на сегодняшний день немало неоднозначных и не до конца изученных моментов, но, как мне кажется, ее дальнейшее развитие внесет существенный вклад в разработку проблемного поля нейроэтики.

ВВЕДЕНСКАЯ ЕЛЕНА ВАЛЕРЬЕВНА²⁶:

НЕЙРОХАКИНГ: ЭТИКО-ФИЛОСОФСКИЕ ПРОБЛЕМЫ

Нейрохакинг является одной из форм биохакинга, направленной на вмешательство в структуру или функцию нейронов для восстановления или улучшения работы мозга и ЦНС. Основной целью нейрохакинга является достижение и сохранение оптимального психического здоровья с помощью влияния на эмоциональную сферу, мышление, восприятие и реакции человека. Так же нейрохакинг может быть направлен на усиление когнитивных и сенситивных способностей.

Существует два вида нейрохакинга: инвазивный и неинвазивный²⁷. К инвазивному нейрохакингу относятся основанные на хирургических

²⁵ *Serr M.* Девочка с пальчик / пер. с фр. А. Соколинской. — М. : Ад Маргинем Пресс, 2016.

²⁶ Кандидат философских наук, доцент кафедры философии и старший преподаватель кафедры биоэтики лечебного факультета РНИМУ им. Н. И. Пирогова, ведущий научный сотрудник ИНИОН РАН, Москва.

²⁷ *Teunisse W., Youssef S., Schmidt M.* Human Enhancement Through the Lens of Experimental and Speculative Neurotechnologies // *Human Behavior and Emerging Technologies.* — 2019. — Vol. 1, no. 4. — P. 361–372.

вмешательствах технологии нейростимуляции, которые позволяют заменять или увеличивать функции организма (имплантаты, протезы). Неинвазивный нейрохакинг подразумевает использование технологий, которые модифицируют организм без таких вмешательств (определенный тип питания и комплекс физических упражнений, медитация, самогипноз, средства психофармакологии, технология взаимодействия мозг-машина, способная связывать людей и артефакты, транскраниальная электростимуляция мозга).

Одним из самых известных нейрохакеров является художник-киборг Нил Харбиссон. Рожденный дальтоником Харбиссон в 2004 году встроил с помощью доктора, пожелавшего остаться анонимным, в свой череп антенну, переводящую цвета в звуки, которые он слышит благодаря костной проводимости. Данное устройство частично восстанавливает его цветное зрение, но также позволяет ему «слышать» цвет, такой как инфракрасный и ультрафиолетовый²⁸, недоступные восприятию обычного человека. Таким образом, новое устройство дополняет и расширяет восприятие Харбиссона, позволяя ему слышать цвета в голове, открывая новые и незнакомые для человека ощущения.

В настоящее время имплантаты становятся более доступными, и сообщества киборгов, такие как Фонд Киборга, основанный Нилом Харбиссоном и Мун Рибас, начали продвигать имплантаты для нейротехнологий. Художница-киборг Мун Рибас использует имплантаты, которые связаны с онлайн-новыми сейсмографами и которые дают ей дополнительные ощущения, позволяя ей «видеть» движение за спиной и чувствовать землетрясения во всем мире²⁹.

Илон Маск разработал Neuralink: имплантируемый мозговой интерфейс для людей, соединяющий мозг и ИИ³⁰ Neuralink представляет собой вживленные провода и электроды через 8-миллиметровое отверстие в черепе, в то время как сигналы принимает расположенный за ухом небольшой блок, который передает данные на компьютер. Основная цель создания интерфейса в данный момент: излечение людей

²⁸ *Jeffries S.* Neil Harbisson : The World's First Cyborg Artist / The Guardian. — 2014. — URL: <https://www.theguardian.com/artanddesign/2014/may/06/neil-harbisson-worlds-first-cyborg-artist/> (visited on Feb. 2, 2020).

²⁹ *Moon Ribas* : The Cyborg Dancer Who Can Detect Earthquakes / CNN. — 2018. — URL: <https://edition.cnn.com/style/article/moon-ribas-cyborg-smart-creativity/index.html/> (visited on Feb. 2, 2020).

³⁰ *Musk E.* An Integrated Brain-Machine Interface Platform with Thousands of Channels / bioRxiv. — 2019. — URL: <https://doi.org/10.1101/703801/> (visited on Feb. 2, 2020).

с неврологическими расстройствами. В дальнейшем Neuralink может быть использован здоровыми людьми для расширения естественного интеллекта, благодаря интеграции с ИИ.

Сьюзан Шнайдер, исследователь из Тихоокеанского университета, утверждает, что такая интеграция опасна и может привести к случайному самоубийству. «Вы можете дополнить свой мозг чипами, но наступит момент, когда вы покончите с жизнью. Я называю это „утечкой мозгов“»³¹. В качестве примера она привела гипотетический сценарий, при котором во время рождения в мозг ребенка встраивается ИИ-девайс, который постоянно мониторит поведение и активность мозга, чтобы идеально копировать мысли и действия. В результате, к достижению совершеннолетия девайс создаст идеальную копию мозга, которая может думать и вести себя также, как человек. В этот момент биологический мозг становится не нужен и его замещают копией мозга. Возникают закономерные вопросы: что является истинным «Я» — биологический мозг или его идеальная копия? Где будет находиться оригинальное сознание? Исчезнет ли оно в момент удаления биологического мозга? Как считает С. Шнайдер, в момент отделения естественного мозга от его копии происходит самоубийство биологического «Я».

Нейрохакеры имеют различную мотивацию, которая включает: преодоление умственных или физических недостатков, самоконтроль, улучшение когнитивных способностей и расширение чувственных восприятий, самосовершенствование, улучшение самооценки. Практические причины нейрохакинга определяются целями, которые варьируются от общего улучшения человеческих способностей и интеллекта до необходимости прикладывать меньше усилий (например, управление светом или обогревателем в «умном доме»). Некоторые нейрохакеры стремятся к добавлению дополнительных чувств и ощущений, чтобы выйти за пределы человеческого восприятия и найти новые формы самовыражения, творчества и понимания мира в целом.

Особую группу представляют домашние пользователи нейротехнологий. Они изучают нейротехнологию, чтобы демократизировать инструменты науки, повысить результаты обучения³², адаптировать инстру-

³¹ *Schneider S.* Merging with AI Would Be Suicide for the Human Mind / The Financial Times. — 2019. — URL: <https://www.ft.com/content/0c4fac58-bd15-11e9-9381-78bab8a70848/> (visited on Feb. 2, 2020).

³² *Wexler A.* The Social Context of “Do-It-Yourself” Brain Stimulation : Neurohackers, Biohackers, and Lifehackers // *Frontiers in Human Neuroscience*. — 2019. — Vol. 11, no. 224. — P. 1–6.

менты к их конкретным потребностям и создавать приложения, не имея коммерческого интереса. Хотя, в основном, их намерения являются конструктивными, им часто не хватает профессиональных знаний и навыков, что может привести к непреднамеренному самоповреждению.

Технологии нейрохакинга имеют одну общую черту, состоящую в том, что первоначально разрабатываемые для медицинского применения, они имеют в себе потенциал использования здоровыми людьми для улучшения своих когнитивных и сенситивных способностей. В этом отношении технологии нейрохакинга мы можем назвать трансгуманистическими, поскольку они способствуют когнитивной и биологической эволюции человека в новый, доселе несуществовавший на земле, вид. Трансгуманистические технологии нейрохакинга также могут произвести социальную революцию, в то время как высокая стоимость многих нейротехнологий увеличит расслоение общества. Так, авторы доклада Национального совета по разведке США «Глобальные тренды — 2030» заявляют: «Через 10–15 лет технологии расширения человеческих возможностей будут доступны только тем, кто сможет за них заплатить. Это повлечет за собой построение двухуровневого общества, и неизбежно возникнут морально-этические проблемы»³³. К таким проблемам можно отнести доступность данных технологий всем людям, дискриминацию людей, не желающих или не имеющих возможности расширять свои когнитивные и сенситивные способности и как следствие социальную сегрегацию.

Бесконтрольное вторжение в тонкие структуры мозга обостряет поиск ответов на философские вопросы о сущности нашего сознания, воспоминаний и когнитивного опыта. Эти вопросы стали наиболее актуальными в связи с проведенными в последние годы опытами по имплантации фальшивых воспоминаний крысам и птицам.

Анализ развития технологий нейрохакинга и постоянное расширение областей его применения приводит к реальной опасности манипулирования поведением человека путем имплантации фальшивых воспоминаний, стирания воспоминаний и считывания мыслей. Стремясь к улучшению своих когнитивных и сенситивных способностей с помощью нейрохакинга, человек может сделать свой мозг уязвимым для внешнего контроля и управления им, что грозит ему потерей свободы воли, конфиденциальности и достоинства. Применение технологий нейрохакинга ставит

³³Global Trends 2030 : Alternative Worlds / National Intelligence Council. — 2012a. — URL: <https://info.publicintelligence.net/GlobalTrends2030.pdf/> (visited on Feb. 2, 2020).

перед нами вопрос о «норме» допустимого вторжения в мозг человека и «норме» допустимого улучшения его способностей. Также нельзя забывать о том, что нейротехнологии могут быть инструментами «двойного назначения», а это значит, что их можно применять не только для решения медицинских проблем или с целью улучшения способностей, но и в военных целях. «Прикрепленные к мозговой ткани компьютерные системы, которые позволяют парализованному пациенту силой мысли управлять роботизированными протезами, можно также использовать для управления бионическими солдатами и пилотируемыми летательными аппаратами. А те устройства, которые поддерживают дряхлеющий мозг, можно применять для внушения новых воспоминаний или удаления существующих — как у союзников, так и у врагов»³⁴.

Технологии нейрохакинга нуждаются в этико-философской оценке и регулировании, поскольку бесконтрольное и бездумное их использование несет в себе экзистенциальные риски, как для человека, так и для человечества.

ХАСАНОВ РУСТАМ ЮРЬЕВИЧ³⁵:

СОЦИОКУЛЬТУРНЫЕ ОСОБЕННОСТИ НЕЙРОЭТИКИ

Нейроэтика представляет собой отдельную ветвь биоэтики, призванную объединить нейробиологические знания с человеческими ценностями. Возросший интерес к нейроэтике сегодня обусловлен значительными успехами в области наук о мозге. Новые открытия позволяют пересмотреть традиционные философские интуиции и юридические нормы в отношении моральной ответственности человека. В рамках ознакомления широкой общественности с результатами работы проведенной в области нейронаук конгресс США объявил «десятилетие мозга» (1990–1999)³⁶ и «десятилетие контроля над болью» (2001–2011)³⁷, в фокусе которых формулировался современный нейроэтический дискурс. В 2006 при поддержке фонда Дана было создано международное нейроэтическое сообщество, способствующее поощрению диалога об

³⁴ *Рекварт Т.* Вот ваш мозг : а вот ваш мозг как оружие / Foreign Policy. — 2015. — URL: <https://inosmi.ru/world/20150922/230398000.html/> (дата обр. 2 февр. 2020).

³⁵ Аспирант философского факультета МГУ имени М. В. Ломоносова, Москва.

³⁶ *Tandon P. N.* The Decade of the Brain : A Brief Review // *Neurology India.* — 2000. — Vol. 48, no. 3. — P. 199–207.

³⁷ *Brennan F.* Congressional “Decade on Pain Control and Research” 2001–2011 : A Review // *Journal of Pain and Palliative Care Pharmacotherapy.* — 2015. — Vol. 29, no. 3. — P. 212–227.

ответственном использовании достижений наук о мозге. Сегодня работа нейроэтических комитетов осуществляется с подачи национальных исследовательских инициатив в области нейронауки³⁸. Уже 8 стран утвердили собственные национальные программы. Первая из них была разработана Евросоюзом — это известный «Human Brain Project»³⁹. Этические поиски в рамках этого проекта весьма амбициозны и затрагивают вопросы научного понимания сознания, свободы воли и отличительных особенностей естественного интеллекта. Кроме того, юридические аспекты, такие как важность конфиденциальности получаемых данных, и возможности, предоставленные для анализа таких данных, тоже оцениваются философами и этиками. В 2013 году Барак Обама подписал национальную программу «Brain Initiative»: в подготовительном обзорном докладе, посвященном нейроэтике, предлагается проводить исследования по следующим направлениям. (1) Исследования на стыке «неврология / этика», (2) клинические исследования в области нейронауки и связанные с ним риски для пациента, распространение знаний о новых клинических исследованиях пациентов, страдающих психическими заболеваниями; (3) поддержка информированности в области этических проблем, возникающих в результате исследований мозга, в идеале это интеграция этики и биологии; (4) и новые возможности для информационно-познавательного участия, направленного на привлечение политиков, журналистов, преподавателей, а также юристов для обсуждения социальных и этических последствий научных исследований⁴⁰. Если раньше исследования мозга были сосредоточены в Европе и Соединенных Штатах, то сегодня большое количество новых данных поступает из Китая и Индии, а также стран Южной Америки и Африки. О своих национальных программах заявили Япония, Израиль, Китай, Канада, Южная Корея, Австралия. Также о желании организовать собственные национальные программы заявили Россия, Иран, Куба⁴¹.

³⁸Salles A., Bjaalie J. G., Evers K. The Human Brain Project : Responsible Brain Research for the Benefit of Society // The Journal of Neuroscience. — 2018. — Vol. 38, no. 50. — P. 10583–10585.

³⁹Salles A., Bjaalie J. G., Evers K. The Human Brain Project : Responsible Brain Research for the Benefit of Society // Neuron. — 2019. — Vol. 101, no. 3. — P. 380–384.

⁴⁰Rommelfanger K. S., Jeong S. J. Global Neuroethics Summit Delegates : Neuroethics Questions to Guide Ethical Research in the International Brain Initiatives // Neuron. — 2018. — Vol. 100, no. 1. — P. 19–36.

⁴¹Rommelfanger K. S., Jeong S. J. Global Neuroethics Summit Delegates : Neuroethics Questions to Guide Ethical Research in the International Brain Initiatives // Neuron. — 2018. — Vol. 100, no. 1. — P. 19–36.

Такое количество новых проектов рождает огромную проблематику биотического и этического толка, в виду различия культурного наследия этих стран. Например, канадская программа «Canadian Brain Research Strategy» претендует на специфический «традиционный» канадский подход, к вопросам нейроэтики. Основное направление — это неврология и право. Идеология развития национального проекта выражена в 4-х положениях: (1) максимальное использование потенциала каждого человека; (2) здравоохранение, ориентированное на восстановление и предотвращение когнитивных нарушений; (3) выяснение возможностей внешнего воздействия на автономность работы мозга; (4) инновации и этика опасность новых технологий⁴². Китайские ученые и специалисты по этике выступают за интеграцию этических, правовых и социальных вопросов в планировании и дальнейшей реализации «China Brain Project». Есть проблема соответствия этического поиска в рамках переосмысления языкового и философского наследия и идеологии современной правящей партии, на что обращают внимание китайские этики. Трансплантация органов в Китае затрудняется постулатом о сыновьей почтительности, тело человека принадлежит, скорее, отцу и матери, чем ему самому, отсюда проблема донорства. Но все же, КНР стремится к реализации международных отношений в этическом диалоге, и эта жизненная необходимость для Китая не скрываясь определяется экономическими соглашениями с Западом⁴³. Корейский проект «Korea Brain Alliance» направлен в первую очередь на детальное картирование мозга человека и построение коннектома различных типов. Современная медицинская практика мало связана с культурным наследием, но моральный выбор донорства для Кореи, схож с китайским⁴⁴. Японская программа «Brain/Mind» в постановке своей основной задачи кроме трансляции знаний и правовых аспектов фокусируется на организации данных о здоровье мозга в единой системе. Конструирование таких данных позволит построить модели здорового и больного мозга. В итоге планируется построение «банка мозгов» содержащего персональные

⁴² *Illes J., Weiss S., Jeong S. J. Neuroethics Backbone for the Evolving Canadian Brain Research Strategy // Neuron. — 2019. — Vol. 101, no. 3. — P. 370–374.*

⁴³ *Wang Y. [et al.]. Responsibility and Sustainability in Brain Science, Technology, and Neuroethics in China—a Culture-Oriented Perspective // Neuron. — 2019. — Vol. 101, no. 3. — P. 375–379.*

⁴⁴ *Wang Y. [et al.]. Responsibility and Sustainability in Brain Science, Technology, and Neuroethics in China—a Culture-Oriented Perspective // Neuron. — 2019. — Vol. 101, no. 3. — P. 375–379.*

данные о фМРТ для всей Японии. Эти проекты — несмотря на языковые, социокультурные особенности и особенности, связанные с интуицией относительно природы собственного «я» — одинаково ориентированы в области юридических стандартов. Стремление к объединению этики науки и науки об этике указывает на желание установить нейроэтику в качестве новой этической теории моральной нормативности. Если научное знание предлагает овладение чем-либо, то современные инициативы в области нейронауки потенциально представляют колоссальную угрозу. Интерпретация обществом, в том числе его учеными, интуиций относительно морали и моральной ответственности в значительной степени влияет на выбор осуществляемых проектов и пути проведения исследований, а также определяет конечные области применения нейротехнологий. Поэтому в настоящее время так необходима этическая и философская инспекция работы проводимой в этой области.

ГАВРИЛИНА ЕЛЕНА АЛЕКСАНДРОВНА⁴⁵:

МОЖЕТ ЛИ РОБОТ БЫТЬ МОРАЛЬНЫМ АГЕНТОМ?

В современном мире роль техники и создаваемой с ее помощью реальности, определяющей не только параметры и контуры нашего вещного мира, но и регулятивы человеческой активности и взаимодействия, уже практически никем не оспаривается. Более того, достаточно активно эта реальность постулируется как гибридная, сочетая не только физический естественный мир, самого человека, но и множество нечеловеческих искусственных агентов, опосредующих как функционирование этой среды, так и поведение человека в ней. Типичным примером действующих нечеловеческих агентов выступают интеллектуальные автономные системы, в частности — роботы. Некоторые исследователи даже говорят о формировании нового типа социальности, в которой люди вступают во взаимодействие с искусственными автономными агентами (например, с голосовыми помощниками или агентами типа Replika), называя его «искусственной социальностью». Действительно, обратившись к анализу того, как происходит взаимодействие между людьми и роботами в сфере ассистивной роботехники (использование роботов-сиделок, роботов для обучения детей с РАС, роботов-партнеров и других), можно заметить (см. исследования в области Human-Robot Interaction), что у людей есть определенный набор ожиданий в отношении взаимодействий

⁴⁵Кандидат философских наук, доцент кафедры социологии и культурологии МГТУ им. Н. Э. Баумана, Москва.

с роботами, при этом происходит как бы «вочеловечивание роботов», и одновременно происходит некоторая «роботизация человека». Очень важными аспектами уже недалекого будущего станут вопросы, связанные с наделением роботов субъектностью, как минимум в правовом поле, с пересмотром способов регистрации отношений людей и роботов и другие. Краткая фиксация новых типов социальных взаимодействий, возникающих в наше время позволяет нам распространить на роботов и, шире, на интеллектуальные автономные системы в целом, категорию морального партнерства. Это понятие изначально развивалось в экологической этике и под ним понималось нечто, наделенное внутренней ценностью, вызывающее эмоциональный отклик и иногда обладающее способностью вступать в эмоциональные отношения. Примером морального партнерства могут быть домашние животные, природа в целом, какие-то памятные предметы, связанные с эмоционально значимым лицом и т. п.

Моральный партнер, по сути, это уже субъект морали, который в свою очередь может осознавать или не осознавать этичность или неэтичность своих поступков, но по отношению к которому человек наделяется определенным спектром прав и обязанностей. Моральным партнером может быть, например, маленький ребенок, или человек, ведущий в силу, каких-то обстоятельств, вегетативный образ жизни. Моральный партнер становится моральным агентом, если он способен строить моральные суждения и нести ответственность за свои действия, сконструированные на основе этих суждений. Если говорить более простым языком, то моральный агент — кто-то или что-то, способное совершать действие, связанное с необходимостью или способностью выбирать (в определенных параметрах) что делать. При этом необходимо уточнить, что даже взрослый дееспособный человек не всегда является моральным агентом. Очень характерный пример утраты агентности — объективация взрослого человека внутри медицинских процедур (см. работы Мишеля Фуко, Аннмари Молл, Хариса Томпсона и др.). Кроме того, мы достаточно часто сталкиваемся с объективацией другого человека, обращаясь лишь к его функциональной роли или, как Марина Абрамович в своем знаменитом перформансе *Rhythm 0*, представляя себя в качестве объекта для воздействия других. Это означает, что переход из состояния агентности в иное состояние людьми освоен. Теперь зададимся вопросом, возможно ли, чтобы объект, артефакт наделялся людьми агентностью? Агентность понимается нами здесь как способность действовать, вызывая при этом ответное действие другого, как

способность быть агентом. Для ответа на этот вопрос представляется логичным обратиться к ресурсам акторно-сетевой теории, в рамках которой важен принцип ирредукционизма — запрета на априорное различение человеческого и нечеловеческого, культурного и природного, микро и макро и т. п. Эта методологическая установка дает решающее преимущество в исследованиях процессов инноваций, где природа и масштаб акторов, правила взаимодействия, критерии оценки и т. д. крайне изменчивы и не определены. Таким образом, ANT наиболее сильна в эмпирическом изучении ситуаций с высоким уровнем неопределенности, поскольку предлагает очень гибкий методологический аппарат, позволяющий описывать любые действия и отношения акторов, не редуцируя их, и не меняя язык описания. В работе Тревора Пинча «Укрощение нечеловеков: некоторые рутинные онтологические исследования» описано с отсылкой к Латуру, как люди делегируют нечеловеку принципы морали, формально присущей только людям. Знак «снизьте скорость» заменяется скоростным барьером («лежачим полицейским»), хоть в этом случае взаимодействие человека с нечеловеками является опосредованным другим нечеловеком (автомобилем). Разумеется, в данной коммуникативной ситуации окончательное решение принимает человек, но оно обусловлено нечеловеческими агентами. В этом же контексте становятся значимыми рассуждения об основаниях принятия решения, например, автономным автомобилем, что приводит к восприятию последнего как морально действующего агента. Причем в полном соответствии с понятием моральной агентности. Скорее всего, нет никаких ограничений и для восприятия робота в качестве морального агента. Тем более в силу эксцентрической позициональности (в духе Плеснера) человека, для нас не составляет никакого труда подбирать такие антропоморфные причины поведения последних.

ШНУРЕНКО ИГОРЬ АНАТОЛЬЕВИЧ⁴⁶:

«ТЕМНАЯ ОНТОЛОГИЯ» КАК «СТАРАНИЯ МАЛЫХ»
И ПРОБЛЕМА ОТКАЗА ОТ ЧЕЛОВЕКОЦЕНТРИЗМА

В техническом использовании всего человеческого ресурса, вплоть до эмоций, мыслей и когнитивных способностей, видится основная цель «Четвертой промышленной революции». Обоснованием тезиса «человек-новая нефть» служат идеи таких представителей аналитической философии, как Леви Брайант, отрицающих сознание. Его отказ

⁴⁶Писатель, журналист, Москва.

от человекоцентричной позиции зиждется на опыте деконструкции, предпринятой структуралистами и постструктуралистами. Так физикалистский взгляд на сознание приводит к отрицанию самой возможности познания как выхода человека за свои пределы. Вот что говорит британский философ Гален Струосон в своем объяснении абсурдности иллюзионизма: «В двадцатом веке произошел самый впечатляющий эпизод в истории идей, да и во всей истории человеческой мысли. Целый ряд мыслителей стали отрицать существование того, о чем мы с уверенностью можем сказать, что оно существует: а именно, существование сознания». Правящие элиты на протяжении мировой истории никогда особо не церемонились с человеком. Прогресс требует жертв, и власти ценили жизнь своих подданных только тогда, когда их к этому вынуждали. Рабы в просвещенной Греции были «говорящими орудиями», русских крестьян покупали и продавали как скот, английская буржуазия составила свой первоначальный капитал на «огораживании», безжалостно сгоняя с земли соотечественников, живших на ней веками. В этом контексте можно вспомнить и работорговлю, которая в Новое Время велась не менее масштабно, чем в эпоху древних цивилизаций. Человек всегда использовался элитами как некий возобновляемый ресурс, поэтому девиз (пусть и неофициальный) четвертой промышленной революции «человек — это новая нефть» никого особенно не удивил. Открыто или негласно, человек всегда считался при этом частью материального мира, что позволяло обращаться с ним как с предметом, но все же это делалось с оговорками. Все же предполагалось, что в нем есть неисчисляемые, неалгоритмизируемые качества, такие как «сознание», «душа», «свобода воли». То, что свободу воли можно отобрать, сознание отнять, а душу купить или продать подразумевало, что они есть даже у раба. Непосредственно нематериальные понятия описывались не наукой, а философией, и при том, что именно науке общество обязано нынешним техническим взлетом. Философия также до самого последнего времени ценилась как некий кладезь мудрости, куда можно слазить за ответами на базовые, ключевые вопросы о бытии и сущем, об отношении тела к разуму и духу, о том, действительно ли мы видим то, что мы видим, или нам все это, в самом деле, только кажется, и о том, что такое «действительно», «в самом деле» и «кажется». Таким образом, человек напрямую и целенаправленно не являлся объектом воздействия техники и преобразования материального мира, хотя, конечно, не мог не испытывать на себе последствий собственных действий по подчинению и преобразованию природы.

В последнее время, однако, ситуация кардинально поменялась. Подчинив в огромной степени живой мир планеты и поставив под контроль ее геологические, энергетические, биологические и информационные ресурсы, технократическая элита целенаправленно обратилась к человеку как к ресурсу нового витка развития⁴⁷. При этом, несмотря на всю риторику, развитие на деле сужается до экономического роста, и более того, до набора финансовых индикаторов («ВВП» и подобные фетиши). В техническом использовании, сознательной манипуляции и экономической эксплуатации прежде всего человеческого ресурса, вплоть до эмоций, мыслей, видится основная цель «Четвертой промышленной революции». Сегодня речь идет уже не об изменении сознания, а о его отрицании. Это фашистскому режиму нужны были восторженные крики и слезы на глазах обывателей, слушающих фюрера, а в прекрасном новом мире технократов прекрасно обойдутся и без бутафории. Технократам не нужна личность с промытыми мозгами — им не нужна личность как таковая. Им не нужен контроль над сознанием, ибо сознание в любом случае несет риски. Рисков нужно избегать, поэтому лучше перевести субъекты в полностью контролируемые объекты, то есть всегда счастливые машины, не подверженные смерти и страданиям, которые несет сознание. «По мере того как растут неспособность мышления, нежелание осмысления и отвращение к вопрошанию и бессилие сущностного решения и взвешивания, величие истории оказывается в зависимости от старания малых», писал М. Хайдеггер⁴⁸. В качестве примера «впутывания малых в видимость величия» можно назвать модную сегодня среди технократов «философию темной онтологии» Леви Брайанта⁴⁹. Леви Брайант — аналитический философ из университета Остина (Техас), преподает «континентальную философию» под которой понимаются прежде всего структуралисты вроде Фуко и Деррида. Опыт тотальной деконструкции не мог не привести их последователей к отказу от человекоцентричной позиции, что оказалось весьма кстати для технократов, которые тяготеют к необходимости учитывать человека в своем «просчитывании будущего». Отвечая предложением на

⁴⁷ *Шваб К.* Четвертая промышленная революция / пер. с англ. К. Ахмедова. — М. : Бомбора, 2016.

⁴⁸ *Хайдеггер М.* Размышления VII–XI : Чёрные тетради 1938–1939 / пер. с нем. А. Б. Григорьева. — М. : Издательство Института Гайдара, 2018.

⁴⁹ *Briant L.* Axioms for a Dark Ontology / Larval Subjects. — 2013. — URL: <https://larvalsubjects.wordpress.com/2013/05/14/axioms-for-a-dark-ontology/> (visited on Feb. 2, 2020).

образовавшийся запрос, Брайант стал одним из пионеров онтикологии, то есть объекто-ориентированной онтологии⁵⁰. Брайант примерил на себя роль Николая Коперника, который 500 лет назад убрал человека из центра Вселенной. Теперь человек перестает быть центром метафизики, провозгласил философ. Метафизика без человека не нужна, поэтому все метафизические концепции объявляются ненужными. За исключением одной: метафизической концепции о том, что сознание может постулировать мир без сознания. В понимании Брайанта, мир состоит лишь из объектов, которые разделены между актуальностью и виртуальностью, а бытие состоит полностью из машин или процессов. Это машинно-ориентированная онтология — хотя «онтологией» ее назвать нельзя, потому что она описывает мир далеко не полностью и не исчерпывающе. Один из тезисов «темной онтологии» Брайанта гласит: «У сознания нет особого понимания ни того как действует тело, в котором оно возникло, ни понимания причин своих когнитивных и аффективных состояний. Как следствие, к свидетельствам, получаемым в результате самоанализа, следует относиться с осторожностью»⁵¹. Здесь полностью уходят все понятия, хоть как-то связанные с человеком — вернее, они замыкаются сами на себе. Так физикалистский взгляд на сознание и человека приводит к отрицанию самой возможности познания как выхода за пределы человека. Остается лишь сплошной *das Vorhandene* — наличное. Термин, обозначающий то сущее, которое просто пребывает в мире, удаленное от человека. Хотя и по его поводу человек не может знать ничего. Уходят такие понятия, *die Zeitlichkeit* временность, феноменальное основание или смысл бытия. Совершенно непонятно, что в таком мире вообще делает человек. Он возник совершенно случайно и становится средством, *der Zeug*, способом бытия подручной вещи в окружающем мире. Этот «*der Zeug*» собственно и является на деле предметом изучения «темной онтологии», которую правильнее было бы назвать «цойгологией». Согласно Брайанту, все смыслы за всю историю существования человечество, включая науку являются иллюзиями не в том смысле, что они несовершенны и будут со временем заменены новыми смыслами. Нет, это полная иллюзия,

⁵⁰ *Briant L.* Axioms for a Dark Ontology / Larval Subjects. — 2013. — URL: <https://larvalsubjects.wordpress.com/2013/05/14/axioms-for-a-dark-ontology/> (visited on Feb. 2, 2020).

⁵¹ *Ibid.*

наваждение случайного набора кусочков в калейдоскопе. То есть человек может быть лишь средством, инструментом, вещью в этом мире, полностью лишенном смыслов.

Многие из представителей интеллектуальной элиты активно продвигают внедрение когнитивных технологий, закрывая глаза на риски. Обоснованием этого и служат тезисы «темной онтологии».

БУГАКОВ ИГОРЬ АЛЕКСАНДРОВИЧ⁵²:

НЕЙРОЭТИКА И ДРУЖЕСТВЕННЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Последние данные различных, прежде всего имеющих отношение к «нейро», наук свидетельствуют о незаслуженном забвении суждения Канта об априорной предзаданности вызывающего «сильное удивление и благоговение» морального закона в каждом из нас и ошибочности превалирующего ныне мнения о том, что все нравственное в человеке формируется социумом. Форма и структура живых организмов априорно приспособлены к среде их существования. Причем не только «копыто лошади» или «плавник рыбы», но и «формы восприятия пространства и времени»⁵³. Основные структуры мозга человека соответствуют возможностям его тела (и наоборот) и заданы априорно. Априорно заданы и многие алгоритмы функционирования этих структур. В соответствии с принципом минимальности (ПМ) природа стремится сочетать в каждом живом организме «жесткое» филогенетическое и «гибкое» (под конкретную среду обитания и решаемые задачи) онтогенетическое⁵⁴.

⁵²Межрегиональное общественное учреждение «Институт инженерной физики», г. Серпухов Московской области.

⁵³*Лортец К.* Кантовская концепция а priori в свете современной биологии / пер. с нем. А. Б. Толстова // Эволюция. Язык. Познание / под ред. И. П. Меркулова. — М.: Языки русской культуры, 2000.

⁵⁴*Бугаков И. А.* Система принципов построения сложных технических систем и их элементов, подвергающихся экстремальным воздействиям в процессе функционирования, на основе антропологического подхода // Датчики и системы. — 2000. — № 10. — С. 67–71; *Бугаков И. А.* Принцип минимальности и деятельность мозга // Сборник научных трудов юбилейного симпозиума, посвященного 150-летию издания «Рефлексов головного мозга» И. М. Сеченова / под ред. А. Ю. Алексева, Ю. Ю. Петрунина, А. В. Савельева, Е. А. Янковской. — Москва: ИИнтеЛЛ, 2014с. — С. 291–316; *Бугаков И. А.* Минимальность как «кредо» мозга // Нейрокомпьютеры : разработка, применение. — 2014а. — № 4. — С. 13–14; *Бугаков И. А.* Принцип минимальности и деятельность мозга // Актуальные вопросы нейрофилософии — 2015. Ежегодник : материалы международного междисциплинарного семинара «Нейрофилософия» / под ред. А. Ю. Алексева, Д. И. Дубровского, В. Г. Кузнецова. — Москва : ИИнтеЛЛ, 2016. — С. 103–130; *Бугаков И. А.* Информация : появление, существование, восприятие // Датчики и системы. — 2001а. — № 1. — С. 61–68 и *Бугаков*

Отсюда «жесткость», априорная заданность филогенетических нейронных структур (из быстропроводящих сигнал возбуждающих нейронов с электрическими синапсами), оперативно реализующих важные генетически заданные алгоритмы, например, инстинктивные, и «гибкость», формирование по мере надобности (с последующим расформированием для иного использования нейронов при долгом незадействовании) онтогенетических нейронных структур (из возбуждающих и тормозящих нейронов с химическими синапсами), отражающих индивидуальный опыт. Такой подход, в частности, объясняет, почему у низших животных преобладают электрические синапсы. Наличие смешанных (частично химических, частично электрических) синапсов можно объяснить как незавершенностью (продолжением) эволюции вида (с накоплением некоторого порогового количества поколений, полезно использующих реализуемую конкретной структурой функцию, эта структура записывается в геном с переходом на «жесткую», полностью электрическую связь реализующих ее нейронов), так и расширением возможностей использования априорно электрически предпорогово связанных структур.

Базисные нравственные качества человека, способствующие выживанию вида, имеют эволюционное значение и потому должны задаваться генетически⁵⁵. Потенциальная степень проявления в человеке нравственного, как и иные способности, индивидуальна в пределах заданного природой для человека как вида диапазона возможных значений нравственности. Конечно, степень использования каждым заложенного зависит от среды. Но некоторых людей в любой среде невозможно склонить к безнравственным поступкам. Человеком нельзя «быть» от рождения, им можно только «становиться». Главное предназначение человека — (стремиться) стать настоящим человеком.

Вышеизложенное имеет непосредственное отношение к нейроэтике и проблеме обеспечения дружественности искусственного интеллекта (ИИ), который бы «испытывал к человечеству чувство привязанности и близости», не оказывая на него негативного влияния. Созданием такого ИИ на благо общества увлечены некоммерческие организации (OpenAI, MIRI и др.). Конечно, С. Хокинга, Б. Гейтса, И. Маска и др.

И. А. Информация : появление, существование, восприятие // Датчики и системы. — 2001b. — № 2. — С. 61–68.

⁵⁵О наличии в нас «морального инстинкта, позволяющего быстро и неосознанно выносить суждения о добре и зле», идет речь, например, в *Хаузер М.* Мораль и разум : как природа создавала наше универсальное чувство добра и зла / пер. с англ. Т. М. Марютиной, Ю. И. Александрова. — 2008.

можно отнести к алармистам, считая, что опасности от ИИ нет, в качестве «объяснения» твердя: «Этого не может быть, потому что этого не может быть никогда» (или, что точнее: «...потому, что я в это не верю» или «...я этого не понимаю»). Можно. Но от этого потенциальная опасность ИИ не исчезает. Такая опасность есть. И она велика. Понимание того, как в человеке (и высших животных) реализуется нравственное, позволит человечеству уменьшить возможное зло как от негативного применения ИИ человеком против человека (человечества), так и от отрицательных действий по отношению к человечеству «рожденного» человеком саморазвивающегося, универсального, «осознающего» себя ИИ, позволит предотвратить (во всяком случае, отсрочить) цивилизационную катастрофу. Если в человеке есть генетическая составляющая нравственности, то почти ясен путь создания дружественного ИИ, состоящий в сочетании в «настоящем» ИИ генетики и образования. Технологически правильно заданные генетические базисные моральные нормы никаким отрицательным «воспитанием» до конца не уничтожить и технически сложно изменить.

Казалось бы, наиболее полный ответ на вопрос о природе психического должны иметь науки о человеке, в частности, психология. Но, несмотря на появление в последнее время большого количества эмпирических данных, понимания того, как нервная система порождает психическое, по-прежнему нет. Конечно, это понимание — сверхсложная задача, ибо «законов такого порождения нет ни в отражаемом, ни в доступных анализу физиологических процессах» (А. Н. Леонтьев). Пока знания о психике по-прежнему представляют собой совокупность подходов, индивидуальных предпочтений, мнений и идей о состояниях, процессах и свойствах.

Еще меньше истинно научных достижений в этике как науке о морали (нравственности). Существует множество разделений этики по различным классификационным основаниям в соответствии с областью деятельности человека (профессиональная, корпоративная, социальная, светская, религиозная, ситуативная (?), универсальная (?) и пр.). Новые области деятельности и проблемы рождают новые «этики»: компьютерную, экологическую, [...], хакерскую (?!). Но понимания механизмов формирования (порождения, проявления) нравственности нет.

В докладе нейроэтику предлагается рассматривать в трех направлениях:

- ◊ Во-первых, как науку, объектом которой является нервная система человека и обладающих нравственным поведением животные;

предметом — процесс формирования и реализации с участием нервной системы нравственных оснований поведения; целью — понимание того, как материальное (процессы в нервной системе) рождает (или участвует в рождении) нравственные качества и переживания; цель — объяснить появление (проявление) нравственного и его задействование индивидом в условиях реальной жизни;

- ◇ во-вторых, как науку о нравственно допустимых (запрещенных) действиях при получении знаний о нейропроцессах и связанных с ними психических проявлениях и целенаправленного управления (как инвазивно, так и неинвазивно) ими (стимуляция мозга, чтение мыслей, стирание и перезапись памяти, формирование предпочтений и т. д.);
- ◇ в-третьих, как науку о применении знаний о нейропроцессах в живом для создания способных к автономному поведению объектов с ИИ, обладающих полезными для человека и носителя ИИ «нравственными» качествами.

Нейроэтика является разделом (продуктом) нейрофилософии. При этом саму нейрофилософию целесообразно рассматривать, во-первых — как философию нейронауки (по аналогии с другими «частными» философиями: науки и техники, управления, образования и т. д.⁵⁶). Иными словами — как философию науки о принципах и законах построения и функционирования нейронов и их сетей, имеющих следствием известные биологические, физиологические и (возможно) психические проявления. Во-вторых, как философию нейроподхода, т. е. как традиционную философию, особенность которой — базирование на знаниях о протекающих в живых системах нейропроцессах, реализуемых механизмах мышления и поведения, понимание и обобщение которых позволяет распространить их на человека, природу и общество, на создаваемые человеком объекты с ИИ.

⁵⁶ *Бугаков И. А.* Нейрофилософия как философия нейронауки // *Известия Института инженерной физики.* — 2014b. — № 2. — С. 43–45; *Бугаков И. А.* Еще раз о нейрофилософии : что это такое? // *Известия Института инженерной физики.* — 2015. — № 1. — С. 95–96; *Бугаков И. А.* Принцип минимальности и деятельность мозга // *Актуальные вопросы нейрофилософии* — 2015. Ежегодник : материалы международного междисциплинарного семинара «Нейрофилософия» / под ред. А. Ю. Алексеева, Д. И. Дубровского, В. Г. Кузнецова. — Москва : ИИнтелЛ, 2016. — С. 103–130; *Дубровский Д. И.* Информация, сознание, мозг. — М. : Высшая школа, 1980. ; *Дубровский Д. И.* Проблема «сознание и мозг» : теоретическое решение. — М. : Канон+, РООИ Реабилитация, 2015.

Дружественность ИИ в докладе предлагается рассматривать:

- ◊ во-первых, как гарантированную сохранность исходных базисных «нравственных» конструкторских настроек специализированного ИИ, защищенность его от сбоев и внешних целенаправленных негативных воздействий с целью изменить заложенные в него цели, программы и алгоритмы поведения (дружественность «слабого» ИИ);
- ◊ во-вторых, как исключение возникновения «желания» и возможности самостоятельного изменения целеполагания способного к автономному целенаправленному поведению универсального, саморазвивающегося, обладающего «сознанием» ИИ, что имело бы следствием причинение прямого или косвенного вреда человеку (дружественность «сильного» ИИ).

Очевидно, что на основе существующих знаний о нейропроцессах в живом и возможностях основного элемента ИИ — искусственных нейронных сетей (ИНС) — создание дружественного ИИ (как «сильного», но и «слабого») сегодня является невозможным. Присущие ИНС существенные недостатки:

- ◊ способность допускать грубые ошибки;
- ◊ легкость их целенаправленного обмана;
- ◊ «непрозрачность» принимаемых ими решений (непонятность для человека их обоснования);
- ◊ необходимость априорного наличия большого объема исходных данных для обучения;
- ◊ существенная зависимость эффективности обучения и последующей работы от качества обучающей выборки;
- ◊ склонность к переобучению;
- ◊ наличие «катастрофической забывчивости» (невозможность научить ИНС чему-то новому с сохранением прежнего опыта)

исключают делегирование им каких-либо прав самостоятельного принятия решений (поведения) в ответственных случаях. Более того, со всеми основаниями можно утверждать об отсутствии у современных ИНС «интеллекта» какого-то бы ни было уровня, как бы нас не убеждали в наличии у них даже «интуиции» некоторые популяризаторы науки. Современные ИНС — не более чем устройства для преобразования вектора входных данных в выходной вектор. «Настоящий» же интеллект — это то, что порождает эффективные алгоритмы (известное: «Интеллект — это сверхалгоритм по порождению алгоритмов»). Именно способность к порождению новых эффективных алгоритмов поведения

как условие успешного выживания в конкурентной борьбе должна быть главным признаком (критерием) наличия (проявления) интеллекта.

Основное предназначение интеллекта — предвидение (достоверное прогнозирование). Предвидение — основа принятия решения. Глубина предвидения — глубина интеллекта («управлять — значит предвидеть»). Ключевая операция — распознавание (на основе и во имя предвидения). Распознавание (объекта, состояния, процесса, будущего и т. д.) лежит в основе принятия решения и его дальнейшей реализации (распознавание текущих действий и их результатов, необходимых будущих действий, степени достижения цели и т. д.). Распознавание и предвидение невозможны без понимания, реализация которого требует наличия той или иной «модели мира» с соответствующим языком описания, органичное встраивание в которую свидетельствует о понимании воспринимающим воспринимаемого. Очевидно, что при этом не обойтись без иерархической категоризации, причем многофакторной (с одновременным отнесением воспринимаемого ко многим категориям области функционирования), что требует соответствующей семантики. Почти все живое чувствует и учитывает время, длительности процессов, последовательность следования событий. Наиболее развитые организмы понимают и учитывают причинную взаимосвязь наступления событий. Все это относится к опыту (генетическому и индивидуальному) и эффективно используется живым. Наконец, самые совершенные организмы обладают сознанием — важным, разрешающим проблему сложности, инструментом интеллекта, позволяющим глубже и эффективнее предвидеть, обеспечивая выживание в конкурентной борьбе с более быстрыми, но менее «умными» соперниками; с использованием механизмов внимания, эмоций, рефлексии формировать пополняющие «подсознание» новые автоматизмов поведения.

Некоторые необходимые действия для обеспечения дружелюбности ИИ (роботов с ИИ):

- ◇ «загрузка ценностей» в ИИ посредством сочетание «генетики» и «образования» (как единства ускоренного воспитания и обучения в среде, мотивирующей на дружелюбность). Генетическая составляющая — главное условие невосприимчивости к пытающимся внедриться в программное обеспечение ИИ «вирусам» недружелюбности (генетический код живого сложно изменить, потому что геном — в каждой клетке; технология блокчейн как некое приближение);
- ◇ использование сертифицированных средств разработчиками ИИ;

- ◇ законодательное ограничение на владение саморазвивающимся ИИ;
- ◇ введение ответственности разработчика («биологического родителя») и воспитателя («фактического родителя»-покупателя и «школы», где учится ИИ) за действия «растущего» ИИ;
- ◇ стандартизация образовательных программ для разработчиков ИИ, различных «видов» ИИ, различных «профессий» ИИ;
- ◇ периодическая диспансеризация ИИ и «антивирусные» прививки (обновления «антивирусных» программ).

Природный ПМ как один из основных принципов эволюции и естественного отбора, постулируя стремление к минимизации расходов основных природных ресурсов (энергии (массы), пространства, времени) на формирование структур и их функционирование, указывает направление, в котором следует искать понимание и искусственное воспроизведение функций интеллекта. Некоторые проявления (следствия) ПМ⁵⁷: сохранение (инерция), самоорганизация, интеграция и дифференциация, динамическое равновесие, динамическое восприятие, предвидение, универсальность, ассоциативность, инвариантность, базисность, относительность, фрактальность, диапазонность существования и пороговость взаимодействия (развития), квантованность, пластичность, фрактальность, полярность, приоритет крутизны (производные и градиенты), иерархически-сетевая структурированность, подражание (имитация), внимание, инстинкты, рефлексy, автоматизмы, эмоции, единство филогенетического и онтогенетического.

ГАРБУК СЕРГЕЙ ВЛАДИМИРОВИЧ⁵⁸:
ОСОБЕННОСТИ СТАНДАРТИЗАЦИИ ТЕХНОЛОГИЙ
ПРИКЛАДНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В последние годы особое внимание уделяется вопросам развития технологий искусственного интеллекта (ИИ). Связано это, прежде всего, с высоким уровнем ожиданий от внедрения существующих и перспективных технологий ИИ в самые разные отрасли экономики, в сферу

⁵⁷Бугаков И. А. Принцип минимальности и деятельность мозга // Сборник научных трудов юбилейного симпозиума, посвященного 150-летию издания «Рефлексов головного мозга» И. М. Сеченова / под ред. А. Ю. Алексеева, Ю. Ю. Петрунина, А. В. Савельева, Е. А. Янковской. — Москва : ИИнтелЛ, 2014с. — С. 291–316.

⁵⁸Кандидат технических наук, директор по научным проектам НИУ «Высшая школа экономики», Москва.

обороны и безопасности. История развития технологий ИИ насчитывает более шестидесяти лет. Динамичное развитие технологий ИИ в последние годы обусловлено, в основном, следующими технологическими факторами: совершенствованием инфраструктуры сбора и передачи больших объемов информации, необходимой для обучения систем ИИ; появлением новых и развитием известных ранее методов обработки информации, основанных на усвоении эмпирического опыта (обучении); совершенствованием средств вычислительной техники, обеспечивающих эффективную реализацию интеллектуальных алгоритмов обработки данных. По мере повышения уровня зрелости технологий ИИ, как и в случае с любыми другими технологиями, наступает момент, когда стандартизация и техническое регулирование становятся необходимыми. Это объясняется тем, что дальнейшее развитие и практическое использование технологий начинает сдерживаться отсутствием соответствующих стандартов. В случае с ИИ таким сдерживающими факторами являются: отсутствие доверия к системам ИИ, связанное с негарантированным качеством работы, непрозрачной логикой принятия решений, подверженностью «состязательным» атакам. Это приводит к ограничениям на применение систем ИИ при решении ответственных задач; метрологические сложности, связанные с отсутствием стандартных процедур сравнения функциональных возможностей систем ИИ друг с другом и с возможностями квалифицированного человека-оператора. В результате возникают проблемы оценки прикладного эффекта от применения систем ИИ; проблемы интероперабельности, обусловленные отсутствием стандартных форматов представления данных, требований к формированию обучающих и контрольных выборок, что приводит к проблемам эффективного встраивания систем ИИ в существующую и создаваемую информационно-коммуникационную инфраструктуру.

Учитывая необходимость стандартизации в области ИИ, на заседании проходившем во Владивостоке в октябре 2017 года Международной организации по стандартизации (ISO) было принято решение о создании специального подкомитета SC42 «Artificial Intelligence» в рамках объединенного технического комитета по стандартизации ISO и Международной электротехнической комиссии (IEC) JTC 1 «Information Technologies». На сегодняшний день проведено четыре заседания этого подкомитета. Заседание в апреле 2021 года планируется проводить в России.

Основными структурными единицами подкомитета SC42 являются его рабочие группы:

- ◇ WG 01 «Foundational Standards» (основополагающие стандарты);
- ◇ WG 02 «Big Data» (большие данные);
- ◇ WG 03 «Trustworthiness» (доверенность систем ИИ);
- ◇ WG 04 «Use Cases and Applications» (практические примеры и приложения);
- ◇ WG 05 «Computational Approaches and Computational Characteristics of AI Systems» (вычислительные методы и характеристики систем ИИ);
- ◇ Соответствующий («зеркальный» по отношению к SC42) национальный технический комитет ТК164 «Искусственный интеллект» был создан Росстандартом по инициативе Российской венчурной компании в июле 2019 года. Национальный ТК включает следующие рабочие группы: РГ01 «Основополагающие стандарты» (руководитель А. И. Агеев, организация, ответственная за ведение секретариата, Международный НИИ проблем управления);
- ◇ РГ02 «Большие данные» (Ю. Е. Хохлов, МГУ им. М. В. Ломоносова);
- ◇ РГ03 «Качество технологий ИИ» (М. В. Федоров, Университет «Сколтех»);
- ◇ РГ04 «Прикладные технологии ИИ» (А. В. Незнамов, ПАО «Сбербанк»);
- ◇ РГ05 «ИИ в образовании» (А. Н. Швиндт, НИУ ВШЭ).

На сегодняшний день в состав ТК164 входит около 90 организаций, включая профильные коммерческие компании, научно-исследовательские организации, вузы, общественные организации и заинтересованные федеральные органы исполнительной власти. Одним из вопросов технического регулирования, специфичных для систем ИИ, является проблема обеспечения этичности процессов создания и применения этих систем. Так, на прошедшей 8–9 ноября 2019 г. в Москве конференции «AI Journey» президент В. В. Путин особо отметил необходимость «формирования свода этических правил взаимодействия искусственного интеллекта и человека».

В рамках SC42 работы по исследованию этических вопросов в области ИИ проводятся, по крайней мере, при подготовке следующих трех документов по стандартизации: технический отчет TR 24027 AI— Bias in AI Systems and AI Aided Decision Making (вопросы справедливости, непредвзятости оценок систем ИИ); технический отчет PDTR 24028 AI—

Overview of trustworthiness in Artificial Intelligence (вопросы доверия к системам ИИ); технический отчет TR 24368 Overview of Ethical and Societal Concerns (этические и социальные вопросы применения систем ИИ).

Все вышеперечисленные документы разрабатываются в рамках рабочей группы WGo3 «Trustworthiness». Соответственно, в составе РГОз «Качество систем ИИ» национального ТК164 сформирована группа экспертов по вопросам этики искусственного интеллекта, представляющая российское научное сообщество при работе над документами. В состав группы экспертов вошли представители Университета «Сколтех», Российской ассоциации искусственного интеллекта, НИУ ВШЭ, НИЦ «Курчатовский институт», ПАО «Сбербанк», МФТИ, МНИИПУ, компании «Т-Платформы» и других организаций. При рассмотрении вопросов этики создания и применения систем ИИ будет учитываться то, что развитие этих систем и их широкое внедрение в человеческом обществе неизбежно приведет к изменению как самого человека, так и социального устройства. При этом необходимо исходить из того, что сформировавшийся эволюционно естественный интеллект человека является самостоятельной ценностью и системообразующим фактором современной цивилизации. Деградация и иные негативные изменения естественного интеллекта человека, возникающие вследствие создания и использования систем ИИ, должны быть полностью исключены. Для этого принятие значимых решений в области управления развитием технологий ИИ (принятие решений относительно финансирования научно-технических проектов в области ИИ и принятие регуляторных решений в данной области) должно сопровождаться возможностью прогнозирования социально-экономических последствий реализации принятых решений, включая, но не ограничиваясь прогнозированием: возможного снижения когнитивных способностей человека, возникающих при систематическом замещении естественного интеллекта искусственным в быту и при выполнении профессиональных обязанностей; последствий создания систем «сильного искусственного интеллекта», от которых ожидают «решения любых интеллектуальных задач, на которые способен человек». Разработка и создание таких систем должны сопровождаться тщательным обоснованием и публичным обсуждением поставленных целей, учитывая возможные изменения на рынке труда, сокращения вплоть до полного исчезновения спроса на отдельные профессии, необходимости переобучения и повторного трудоустройства работников.