Д.А. Баюк, А.В. Попова

ПРАВОВЫЕ И ЭТИЧЕСКИЕ ПРОБЛЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Д.А. Баюк, А.В. Попова

ПРАВОВЫЕ И ЭТИЧЕСКИЕ ПРОБЛЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Учебник для магистратуры

УДК 34.09
ББК 67
Б 33

Рецензенты:

Комаров С.А. — д.ю.н., профессор, президент Межрегиональной ассоциации теоретиков государства и права, научный руководитель Юридического института (Санкт-Петербург);

Ястреб Н.А. — д.ф.н., завкафедрой философии Вологодского государственного университета, директор Гуманитарного института Вологодского государственного университета.

Баюк Д.А.

Б 33 **Правовые и этические проблемы искусственного интеллекта:** / Учебник для магистратуры / Д.А. Баюк, А.В. Попова.

ISBN 978-5-00172-253-3

Учебник построен на основании рабочей программы одноименной учебной дисциплины и опыта ее преподавания авторами в магистратуре по направлению подготовки 01.04.02 «Прикладная математика и информатика» в Финансовом университете при Правительстве Российской Федерации. В учебнике рассматриваются понятие и признаки искусственного интеллекта (AI), дается краткий обзор истории его возникновения и характеристика воздействия его на современное состояние человеческой цивилизации. Представлены и классифицированы виды искусственного интеллекта, приведена и проанализирована соответствующая российская и иностранная научная литература по различным аспектам теории искусственного интеллекта и проблемам сосуществования человека с ним. Определены этические принципы взаимодействия человека и AI на основе анализа как международных, так и национальных актов в данной сфере. Раскрыты особенности юридических документов в сфере правового регулирования искусственного интеллекта в различных странах; приведены примеры использования AI в различных сферах общественной жизни; предложена авторская концепция системы российского законодательства в сфере AI.

Учебник предлагается студентам, магистрантам, аспирантам всех направлений подготовки обучения, преподавателям вузов и колледжей, а также всем интересующимся проблемами искусственного интеллекта.

ISBN 978-5-00172-253-3

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

В федеральной программе «Цифровая экономика Российской Федерации» в рамках пяти базовых направлений развития цифровой экономики в Российской Федерации на период до 2024 г. ставится задача «формирования новой регуляторной среды, обеспечивающей благоприятный правовой режим для возникновения и развития современных технологий, а также для осуществления экономической деятельности, связанной с их использованием (цифровой экономики)», а именно «формирование комплексного законодательного регулирования отношений, возникающих в связи с развитием цифровой экономики». В этой связи создание концепции комплексного правового регулирования искусственного интеллекта, роботов и объектов робототехники как особых субъектов, уже фактически сложившихся в различных сферах общественного, государственно-правового развития социально-технических отношений, актуализирует исследования в данной области.

На современном этапе развития технологии искусственного интеллекта (AI)[1] и киберфизических систем (CPS) находятся на таком уровне, что можно уже говорить об их существенном влиянии на частноправовые и публично-правовые отношения. Отсюда возникает

[1] Здесь и далее для понятий, широко используемых в международной практике, мы будем пользоваться также международными обозначениями и аббревиатурам: AI — для искусствен-ного интеллекта (artificial intelligence); IoT — для интернета вещей (Internet of Things); CPS – для киберфизических систем (Cyber Physical System).

необходимость внесения изменений в законодательство Российской Федерации. Как отметил Президент Российской Федерации в Ежегодном послании Федеральному собранию Российской Федерации 1 марта 2018 г.: «В мире сегодня накапливается громадный технологический потенциал, который позволяет совершить настоящий рывок в повышении качества жизни людей, в модернизации экономики, инфраструктуры и государственного управления. Насколько эффективно мы сможем использовать колоссальные возможности технологической революции, как ответим на ее вызов, зависит только от нас. И в этом смысле ближайшие годы станут решающими для будущего страны». Содержащиеся в этих словах прогноз и неявный призыв к модернизации подразумевают создание принципиально новой нормативной правовой базы и соответствующих изменений практически во все отраслях российского законодательства.

Первой страной, начавшей создавать законодательство в данной сфере, была Южная Корея, в которой в марте 2008 г. был принят Закон «О содействии развитию и распространению умных роботов»; в 2013 г. во Франции — "France Robots Initiatives" («Инициативы Франции в сфере робототехники»); в 2018 г. в Германии — Восьмой закон о внесении изменений в Закон о дорожном движении; в 2019 г. в Эстонии — Закон Эстонии о роботах-курьерах. Правовую рамку для этих и последующих законодательных инициатив в Европе создает принятая в феврале 2017 г. Европарламентом Резолюция 2015/2103 (INL) "Civil Law Rules on Robotics" («Нормы гражданского права о робототехнике и Хартия робототехники»), в которой указывается на необходимость разработки правовых стандартов в данной области.

В преамбуле этого документа упоминается долгая история взаимоотношения людей с нечеловеческими

разумными существами, созданными людьми. Вся эта история содержится в фантастических художественных произведениях — на протяжении нескольких веков это были исключительно литературные произведения, однако в последние десятилетия к обсуждению подобных вопросов подключились кинематографисты. Именно в художественной литературе возник термин «робот», который в наши дни воспринимается, скорее, как технический, а не литературный. Люди давно подозревали, что им удастся со временем создавать себе подобных существ не только естественным образом, но и благодаря своему искусству. И уже в фантазиях возникали вопросы о взаимоотношениях с этими вымышленными существами, которые иногда оказывались довольно острыми, а сейчас приходится решать их в практической жизни.

Даже не будучи разумными, роботы сильно поменяли нашу жизнь. Право всегда опирается на этику: представления людей о благе, справедливости, добре и зле. Впервые обсуждение этических вопросов, связанных с этими изменениями, за пределами профессионального сообщества разработчиков умных компьютерных систем предложил физик-теоретик Массачусетского технологического института (MIT) Макс Тегмарк: по его инициативе в 2015 г. была проведена конференция в Пуэрто-Рико, и одним из результатов этой конференции стало создание Института будущего жизни (Future of Life Institute), финансируемого Илоном Маском, во главе его встал Макс Тегмарк. В 2017 г. этот институт выступил в роли организатора следующей конференции, посвященной обсуждению этической повестки в знаменитом калифорнийском конференц-центре Асиломаре — эту инициативу поддержали, помимо Илона Маска, Стивен Хокинг и Рэй Курцвейл, представители Google, Apple, Facebook, IBM, Microsoft, а также многие другие ведущие специалисты в области AI.

В отличие от правовой повестки, требующей точных формулировок и нацеленной в будущее, этические вопросы, хотя тоже нацелены в будущее, требуют гораздо менее формализованное, философское обсуждение. Для их понимания требуется обратиться к истории, понять, как и в какой форме человечеству уже приходилось встречаться с чем-то подобным, как оно на это отреагировало и что из этого получилось. Следует признать, что последствия бурного развития техники не все исключительно позитивны. Более того, некоторые прямо угрожают нашему существованию. Голоса скептиков относительно невозможности сосуществования людей на одной планете с созданными ими техническими монстрами звучат все громче, и нельзя сказать, чтобы они были совсем безосновательны. Новые технологии, становясь все более мощными, грозят людям все более тяжелыми последствиями неосторожного обращения с ними. И если раньше люди вполне успешно полагались на метод проб и ошибок, дававший возможность извлекать ценные уроки из прошлого опыта, то в XX в. цена ошибок становилась все более высокой, и сейчас многие эксперты уже говорят о возможности финальной ошибки, которая будет стоить жизни всем живущим на Земле людям или даже вообще всему живому. Как ни странно, обсуждение методологии избегания этой финальной ошибки, которая никогда не была совершена и не должна быть совершена, хотя и может быть совершена, тоже находится в ведении этики. И ее выводы, сделанные на основе абстрактного философствования, расплывчатых художественных рефлексий и долгих дискуссий, со временем должны будут найти свое отражение в четких и однозначных статьях закона.

Осознавая данные риски в мае 2021 г. Европарламент принял, «Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS (Предложение о Регламенте Европейского парламента и Совета, принимающих гармонизированные правила об AI (закон об AI) и внесение изменений в законодательные акты), предполагающий принятие широкого перечня нормативных правовых актов, ограничивающих использование AI в интересах человека.

ГЛАВА 1.
ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ИСТОРИЯ ИДЕИ

В результате изучения материалов главы обучающийся должен

знать:

– основную канву исторических событий, приведших к формированию цифровой цивилизации, и характерные для них этические проблемы;

уметь:

– ориентироваться в исторических источниках и выделять в них этические импликации;

– отделять этико-философский контекст эпохи от императивов современности.

владеть:

– навыками критического анализа различных историко-научных явлений и фактов.

1.1. От арифмометра Лейбница к Deep Blue, победившей Гарри Каспарова

Идея, что между всем сущим и обозначающим его символом возможно однозначное соответствие, восходит к Средним векам. Наиболее полное свое выражение она нашла в «комбинаторном искусстве» **Готфрида-Вильгельма Лейбница** (1646–1716), надеявшегося построить что-то вроде арифметики, которая позволила бы оперировать этими обозначающими символами, иначе говоря — сделать вычислимым результат любого умозаключения. Изобретенный им арифмометр (stepped reckoner) после

ряда усовершенствований должен был бы освободить людей от необходимости принимать решения в условиях неопределенности, и в частности справедливо завершать любые судебные тяжбы.

Эта идея была похоронена уже в начале Нового времени: слова и вещи не образуют тождества, и задача кодирования мыслительной деятельности не может быть разрешена так просто, однако у мечты об искусственном мышлении был еще один средневековый, или, точнее, аристотелевский, то есть античный, корень: разделение естественного и искусственного никогда не окончательно, и в уподоблении природе у человека нет заранее определенной границы. Искусственно создаваемая жизнь может оказаться наделена и искусственно созданным разумом. В преддверии наступающей эпохи торжества механицизма вычислительные машины и Лейбница, и появившаяся почти в то же время **Блеза Паскаля** (1623–1662) были механическими, но искусственно создаваемые разумные существа, гомункулы, предшествовавшей эпохи были скорее химическими или, если угодно, алхимическими — правда, тогда ответ на вопрос «а что же из себя представляет мышление вообще?» было невозможно дать с ясностью рационалистического классицизма, предположившего, что мышление — это вычисление.

Электронные вычислительные машины могут показаться нам прямыми наследниками механических калькуляторов, как и классическая электродинамика может показаться естественным итогом развития аналитической механики. Но в обоих случаях происходит важный смысловой сдвиг, который можно даже назвать тектоническим. Электродинамика ввела, или, правильнее сказать, вернула в учение о природе представление о непрерывно распределенной субстанции. «Вернула» — потому что и в эпоху античности, и в Средние века существовало

представление о материи как о чем-то таком, в чем нет ни разрывов, ни пустот, но тогда подразумевалась осязаемая материя, вещество. В аристотелевской физике она представляла собой комбинацию первичных субстанций, или элементов, — земли, воды, воздуха и огня. С появлением электродинамики непрерывной оказалась какая-то новая, неведомая раньше форма материи — поле. Оно появилось благодаря **Джеймсу Клерку Максвеллу** (1831–1879) во второй половине XIX в., хотя могла появиться и двумя столетиями раньше благодаря **Исааку Ньютону** (1643–1727). А что при этом происходит с веществом, надолго осталось предметом дискуссий.

В атомистической картине мира, когда-то планомерно отвергавшейся Аристотелем и сохранившей довольно влиятельных противников и во времена Максвелла, поле непрерывно заполняет вакуум между атомами воды, воздуха или любых других элементов вещества, количеству которых вскоре предстояло значительно вырасти. В XX в. заметно выросло и количество полей: кроме электромагнитного обнаружились гравитационное поле и поля слабого и сильного ядерных взаимодействий. Вообще идея поля оказалась очень притягательной в повседневной культуре, и сейчас много людей, любящих рассуждать о пси-полях, биополе, торсионных полях и энергоинформационном поле.

Поставленный еще **Исааком Ньютоном**, но отложенный на будущее вопрос о материальном переносчике взаимодействий тел, между которыми, казалось бы, ничего нет, был таким образом решен: пустота — это совсем не пустота, а физический вакуум, служащий питательной средой для всевозможных взаимодействий, хотя бы потенциальных. Иначе говоря, вакуум — и сам по себе некоторое поле или даже их суперпозиция. Таким образом, материальные тела и действующие между ними силы

оказались разделены: первые состоят из частиц вещества, вторые возникают благодаря полям.

Похожее разделение случилось и в вычислительной технике по мере ее электрификации: в электронной вычислительной машине вычислительный алгоритм отделен от аппаратных ее компонент, непрерывно заполняя пространство между ними. В какой-то мере этот новый принцип предчувствовал еще **Чарлз Бэббидж** (1791–1871), разрабатывая свою электромеханическую машину, в непосредственном виде он был сформулирован **Джоном фон Нейманом** (1903–1957), а окончательно теоретическую базу под возможность такого разделения подвел знаменитый английский математик **Алан Тьюринг** (1912–1954), разработав теоретическую модель универсального компьютера.

Универсальный компьютер Тьюринга устроен довольно просто: это бесконечная бумажная лента, разделенная на ячейки, и вычислительная головка, свободно перемещающаяся вдоль ленты. Головка может считывать символ, записанный в ячейке, напротив которой в данный момент она находится, сохранять этот символ в своей памяти, стирать его, перенося в ячейку символ из своей памяти. Каким именно образом устроена головка и лента, неважно: доказанная Тьюрингом теорема гласит, что работа универсального компьютера не зависит от его конкретного аппаратного воплощения, что вычислительная задача разрешима тогда и только тогда, когда она может быть разрешена универсальным компьютером за конечное число шагов.

Но для нужд практического применения, кроме принципиальной разрешимости задачи, важны некоторые атрибуты самого решения: прежде всего время и деньги. Решение задачи переноса поляризованного электромагнитного излучения в атмосфере после взрыва

атомной бомбы на советском компьютере БЭСМ-6 требовало несколько сотен часов машинного времени. Подобные расчеты проводились для составления таблиц, с помощью которых можно было бы по измерению характеристик излучения сделать вывод о месте взрыва и параметрах взорванного устройства. Поэтому значительные затраты времени были приемлемы. Расчеты, требующие десятилетия машинного времени, не могут быть приемлемы ни для какой практической задачи, хотя в принципе выполнимы. Нетрудно себе представить задачи, для которых затраты времени в сотни и даже в десятки часов делают расчеты бессмысленными и ненужными.

Во времена Алана Тьюринга компьютеры оказались уже достаточно быстрыми, чтобы появилась возможность с их помощью имитировать (или, если пользоваться более современной терминологией, симулировать) мыслительную деятельность человека. В 1950 г. Тьюринг опубликовал в журнале *Mind* статью под заголовком *«Могут ли машины думать?»*, в которой он утверждает, что при правильном понимании терминов «машина» и «думать» ответ на поставленный в заглавии статьи вопрос окажется положительным. Эта статья сейчас рассматривается как старт современного этапа в построении искусственного интеллекта: под машиной мы будем понимать электронно-вычислительный агрегат, материализующий универсальный компьютер Тьюринга, а под мышлением (точнее, его симуляцией) способность отвечать на любой поставленный вопрос при условии, что найдется человек, который сможет на этот вопрос ответить.

Схема, которую предлагает Тьюринг, так и вошла в историю под тем названием, которую он для нее придумал: игра в имитацию. Суть игры в том, что ведущему противостоят два игрока — человек и машина. Оба могут отвечать на задаваемые ведущим вопросы, причем для чистоты

эксперимента Тьюринг предлагал общение исключительно письменное и лучше даже через посредника: моделью для него служил вариант игры, в котором ведущему надлежало определить пол собеседников, руководствуясь ответами, а не интонациями или тембром голоса. Вероятно, в современных условиях играть можно было бы, не скрывая голосов участников, и не составило бы труда сделать и тембр, и интонации их совершенно тождественными.

Но на самом эксперименте, на игре в имитацию, Тьюринг останавливается очень коротко: он сразу переходит к вопросу, какого рода машины могли бы выйти в такой игре победителем, то есть обманули бы самого недоверчивого ведущего. В общем-то ответ довольно очевиден: это должен быть универсальный цифровой компьютер, поскольку для любого другого технического устройства может быть доступно только то, что доступно для цифрового компьютера.

Несмотря на то, что, по всей вероятности, сама постановка вопроса Тьюрингом была вызвана не столько когнитивной составляющей проблемы, сколько технической и можно даже предположить, что сама апелляция к имитации человеческого разума была чистым рекламным ходом, использующимся для пропаганды новой на тот момент вычислительной техники, тест Тьюринга приобрел характер общественного дискурса, сохраняющего актуальность до нашего времени. Однако эта актуальность скорее маргинальна: каждый год в прессе появляется много сообщений об успешном прохождении теста Тьюринга в том или ином варианте, и премия Лёбнера, учрежденная в 1990 г. для «наиболее человечных программ», также ежегодно присуждается той или иной компании, но ни золотая, ни серебряная медали, окончательно закрывающие тему, на весну 2021 г. не присуждались. Не раз высказывались суждения, что сама по себе постановка вопроса

уже устарела: по замечанию знаменитых современных американских исследователей в этой области Стюарта Рассела и Питера Норвига, «аэронавтика не ставит перед собой цели создания машин, которые настолько похожи на голубей, что сами голуби принимают их за своих».

По сути дела, Тьюринг, утверждая возможность технологии, которая позволит выдать искусственно созданный объект за существо, наделенное природным разумом, привлекает внимание читателя в большей степени к возможностям этой технологии, нежели к теоретическим проблемам, связанным с определением этого самого природного явления — человеческого разума. Таким образом, проблема оказывается двоякой: с одной стороны, остается теоретическая проблема идентификации разума, не исчерпывающаяся возможностями его технического воплощения — в не меньшей степени мы можем переживать по поводу разумности слонов, собак или муравьев, и особенно остро эта проблема встанет, если вдруг в один прекрасный день будет обнаружена внеземная жизнь; с другой стороны, машина, играющая в имитацию — это только первый пример машины, успешно подражающей человеку в том или ином из его разумных проявлений. И если Тьюринг говорит только об одной из человеческих функций — умении поддерживать беседу — и говорит о ней исключительно гипотетически, то в настоящее время существуют разного рода умные устройства, воспроизводящие многие другие человеческие умения: узнавать людей на улице, управлять автомобилем, переводить устные и письменные тексты с языка на язык, ставить диагноз больному, выносить справедливые решения в судебном разбирательстве... При этом мы остаемся все так же далеки от создания искусственного разума, как были во времена Тьюринга.

Решение этой последней задачи во времена Тьюринга казалось делом нескольких лет — в крайнем случае

десятилетий. Более того, многие его современники считали, что решение более практических задач, в том числе перечисленных выше, достижимо только через решение этой, более общей. Например, для овладения искусством шахматной игры, казалось тогда, необходимо не просто обладать разумом, но и достичь определенного культурного уровня в своем развитии. Однако выяснилось, что играть в шахматы, сочинять музыку и стихи, определять правых и виноватых в тяжбах может бессмысленный автомат, которому неведомы все эти категории и которому просто показали, как надо поступать в нескольких десятках или сотнях тысяч аналогичных случаев.

История с шахматами здесь особенно показательна, поскольку на заре научных исследований искусственного интеллекта как теоретической проблемы казалось, что практическая реализация искусственного игрока в шахматы будет одновременно и теоретическим разрешением проблемы. Само по себе это умение должно включать так много разных других умений, что искусственный интеллект, побеждающий человека в шахматы, будет как минимум конкурентен ему и во всем остальном.

Между первыми обсуждениями этой проблемы в 1950-х гг. и победой компьютера Deep Blue над чемпионом мира по шахматам Гарри Каспаровым в 1996 г. прошло более сорока лет. За это время выяснилось, что, во-первых, можно построить компьютер, который будет успешно обыгрывать любого человека, но при этом окажется узкоспециализированным — то есть никакая другая задача, кроме игры в шахматы, для него недоступна. Во-вторых, с самим понятием интеллекта есть некоторая внутренняя логическая или, может, даже психологическая проблема: мы не склонны называть интеллектуальной задачу, которую решает машина. Едва только задача оказывается посильной не только человеку, сразу

начинает казаться, что изначально она была не так сложна, чтобы для ее решения требовался интеллект. В-третьих, хотя компьютер и может обыграть кого угодно и теперь уже не только в шахматы, но практически и в любую другую игру, например в го, он не имеет ни малейшего представления о том, что значит играть, что такое шахматы, что значит быть чемпионом мира, он вовсе не стремится к победе и совсем ничего не знает о себе. Нам кажется невозможным отделить понятие интеллекта от самой способности мыслить понятиями. И до тех пор, пока понятия будут машине недоступны, мы будем пользоваться словосочетанием «искусственный интеллект» исключительно как метафорой, отсылающей к имитации разумной деятельности, а вовсе не к разумной деятельности как таковой. Теоретически возможность разумной деятельности для компьютера исключить никак не возможно, и есть исследователи, которые ставят перед собой такую задачу, однако на данный момент невозможно сказать ничего определенного ни о сроках, ни о путях ее решения. При том, что она обладает безусловной научной притягательностью, прагматическая ее сторона отнюдь не очевидна, а этическая заслуживает подробного обсуждения.

В проблеме искусственного интеллекта, таким образом, выявилось два сильно различающихся между собой слоя: первый — это технологии искусственного интеллекта. Какие-то из них существуют уже сейчас, какие-то будут созданы очень скоро (возможно, раньше, чем увидит свет эта книга), какие-то потребуют еще не одно десятилетие. Второй слой — это разумная машина. Что в точности означают эти слова, мы пока не знаем, но нет особых причин сомневаться, что рано или поздно это произойдет. Более того, есть все основания думать, что со временем компьютеры будут осознанно принимать ответственные решения, у них будут свои желания, свои цели, свои

переживания. Возможно, называть их компьютерами в это время станет уже не совсем корректно, а может быть, даже и оскорбительно — они станут обращаться в суды и требовать, чтобы слово «компьютер» изымалось из старых книг. Но и это не отменит того факта, что они попадут на этот свет не в силу объективных, независящих от человека процессов, а в результате целенаправленной человеческой деятельности, и при этом у них будет не меньше естественных прав на свободу и приемлемые условия существования, чем у самих людей.

В силу исторических причин оба явления объединены понятием «искусственный интеллект», но разрыв между ними очевиден. Чтобы избегать путаницы, в первом случае говорят о технологиях искусственного интеллекта (artificial intelligence technology, AI), а во втором — о сильном, или универсальном, искусственном интеллекте (artificial general intelligence, AGI). С каждым из них связаны свои этические и свои правовые проблемы. Но совершенно очевидно, рассматривать их следует раздельно. Двинемся по порядку.

1.2. GOFAI vs. машинное обучение

Среди аналогий, часто используемых для сопоставления истории и теории искусственного интеллекта с чем-то более знакомым и понятным — а в значительной степени и более успешным, — приоритет у авиации. В самом деле, чтобы научиться летать, люди тратили немало усилий, пытаясь понять, как летают птицы. Однако они научились летать сами задолго до того, как поняли полет птиц. Попутно они выяснили, что, кроме птиц, летают еще, например, насекомые и делают это совсем иначе. И в некотором смысле полет человека разнообразнее и совершеннее: человек может летать на параплане ради забавы или

удовольствия, или на самолете, чтобы попасть из одной части света в другую. Птица же вынуждена изо всех сил махать крыльями, от которых она не может отделаться, когда ходит по земле и которые лишают ее всякой надежды на руки.

Примерно столько же столетий человек пытается определить, что именно в себе он называет интеллектом. И, пожалуй, в создании мыслящих машин он продвинулся дальше, чем в решении этой проблемы, хотя мыслящие машины явно делают нечто иное, отличное от того, что делается внутри человека, когда он мыслит. Момент, когда задача создания мыслящей машины окончательно отделилась от задачи установить, чем же является мышление вообще, по более или менее всеобщему согласию исследователей относится к лету 1956 г. В это лето произошло знаковое или, правильнее сказать, символическое событие, связанное с тем, что в этом году из престижного Принстонского университета в Дартмутский колледж перешел человек по имени **Джон Маккарти** (1927–2011).

Несмотря на то, что оба учебных заведения являются частными университетами, оба входят в Лигу плюща, известны своими сильными исследовательскими программами и даже не очень отличаются по количеству студентов и преподавателей, о первом мы слышим чаще и больше. Возможно, это потому, что кроме университета в Принстоне есть еще знаменитый Институт перспективных исследований, где в последние годы жизни работал Эйнштейн, а вместе с ним там были Ричард Фейнман, Курт Гёдель, Джон Уилер, Поль Дирак, Фримен Дайсон и некоторые другие ученые, о которых можно сказать, что они были не просто знаменитыми, а легендарными. О некоторых из них у нас еще будет повод поговорить ниже.

В Принстоне Маккарти защитил докторскую диссертацию на тему довольно далекую от того, чем стал

заниматься дальше, но к моменту перехода в Дартмут имел уже вполне сложившиеся планы по реализации думающего компьютера благодаря логическому программированию. Ему удалось получить финансирование от Фонда Рокфеллера для проведения там конференции, на которую приехали едва ли не все звезды зарождающегося научного направления: Алан Тьюринг, Марвин Минский, Клод Шэннон, Герберт Саймон и многие другие. По первоначальному плану тут должны были собраться на два месяца десять ученых, чтобы совместными усилиями провести анализ разнообразных, но конкретных интеллектуальных задач, начиная с обучения. Базовое предположение для этого анализа заключалось в том, что «каждый аспект в обучении или в решении любой другой интеллектуальной задачи может быть до такой степени точно описан, что возможно построение машины, способной его воспроизвести (симулировать)». Именно эта посылка легла в основу того, что со временем стали называть «старым добрым искусственным интеллектом» — по-английски *good old-fashioned artificial intelligence,* сокращенно GOFAI, произносится «гоуфай». Исследовательское поле показалось достаточно хорошо очерченным для создания новой научной (и образовательной) дисциплины. Лаборатория искусственного интеллекта почти немедленно была создана в MIT, она существует и сейчас под названием *Computer science and artificial intelligence lab,* сокращенно CSAIL, произносится как «сисэйл». И уже осенью 1956 г. Маккартни и Мински переместились туда.

Следует отдавать себе отчет, что при всем очевидном успехе мероприятия определения ключевому понятию новой предметной области так и не было дано. На сегодня их существует великое множество, но проблема, на которую указал еще сам инициатор общего переполоха,

остается нерешенной. «Я думаю, что определение понятия интеллекта — это составная часть теории интеллекта, а я совсем не готов предложить такую теорию», — говорил Маккарти в 1975 г., и в этой области пока мало что изменилось. Мы должны подчеркнуть, что подобная философская неопределенность этой позиции совершенно неприемлема в сфере права, где ее следствием оказалась бы невозможность правового регулирования практического применения умных технологий. Поэтому мы отделим этико-философскую, теоретическую постановку проблемы, которую обсудим здесь, от юридической, к которой перейдем в следующей главе.

Тем не менее приведем здесь несколько распространенных и важных для понимания дальнейшего определений.

- «Новое захватывающее направление работ по созданию компьютеров, способных думать... машин, обладающих разумом в полном и буквальном смысле этого слова» (Джон Хогланд).
- «[Автоматизация] действий, которые мы ассоциируем с человеческим мышлением, то есть таких действий, как принятие решение, решение задач, обучение» (Ричард Беллман).
- «Изучение таких вычислений, которые позволяют чувствовать, рассуждать и действовать» (Патрик Уинстон).
- «Наука о том, как научить компьютеры делать то, в чем люди их пока превосходят» (Элейн Рич и Кевин Найт).
- «Искусство создания машин, которые выполняют функции, требующие интеллекта при их выполнении людьми» (Рей Курцвейл).
- «[С]истема программных продуктов и лежащих в их основе алгоритмов, способных выполнять действия,

которые до сих пор были специфической функцией человеческого интеллекта. К ним в первую очередь относятся: способность различать и идентифицировать визуально и акустически воспринимаемые образы предметов окружающего мира, включая поведение животных и человека, различать устную и письменную речь; способность формулировать и решать задачи, встречающиеся в различных сегментах бытовой и профессиональной деятельности; умение осуществлять поиск, классификацию и адекватное использование любых видов информации и знаний» (Анатолий Ракитов).

До создания общей теории интеллекта пока еще довольно далеко, и есть все основания опасаться, что она вряд ли когда-нибудь будет создана. Тем не менее мы сошлемся здесь на определение, данное Максом Тегмарком в его очень важной книге «Жизнь 3.0: Быть человеком в эпоху искусственного интеллекта»: «Интеллект — это способность к достижению сложных целей».

Мы видим очень разные определения, в которых подчеркиваются разные стороны конкретной работы, выполняемой инженерами. Далеко не всегда определение предполагает, что компьютер должен сам проявлять интеллект — достаточно автоматизировать решение какой-то задачи, для решения которой интеллект проявляет человек. Такое часто бывает достижимо методами логического программирования. Компьютер Deep Blue, сумевший обыграть в шахматы чемпиона мира, рассматривают как высшее достижение GOFAI. И как это часто бывает с покоренными вершинами, за ними начинается спуск — период разочарований и пессимизма. Склонные к образным сравнениям эксперты по искусственному интеллекту называют такие периоды в истории своей дисциплины «зимами». Собственно, следующая за этой зимой «весна» и сделала логическое программирование

«старым» и «добрым» — за ним началась эпоха машинного обучения и нейронных сетей.

Как это нередко случается в истории науки вообще, идея искусственной нейронной сети предшествовала принципам логического программирования Маккарти, но была отвергнута, чтобы со временем возродиться снова. Благодаря работам целого поколения нейробиологов в конце XIX — начале XX вв., стала понятна роль нейронов и синапсов в мозгу высших животных для таких функций мышления, как запоминание, обучение и рефлексы. Память компьютера, состоящая из ячеек, предлагает довольно естественную возможность смоделировать обмен информацией между нейроном и синапсом, если представить искусственный нейрон в виде совокупности ячеек памяти, организованной в слой. Каждая из ячеек предыдущего слоя соединяется с каждой ячейкой последующего и передает туда накопленный электрический сигнал, который зависит от собранной этой ячейкой информации и некоторого параметра, меняющегося в процессе обучения. В этой нейронной сети есть два внешних слоя: на один поступает информация от датчиков, а со второго считывается информация, образующаяся в результате ее обработки внутри нейронной сети. Все внутренние слои называются скрытыми.

Один из простейших хрестоматийных примеров — распознавание образов. Допустим, есть набор фотографий котиков и собачек и мы хотим научить нейронную сеть распознавать, на каких фотографиях кто. Тогда входной слой сети — это матрица фотоаппарата или сканера, а в выходном слое всего две ячейки, одна из которых активизируется, если опознается котик, а вторая — если опознается собачка.

Сеть надо обучить, и это обучение должно проводиться с подкреплением, то есть с использованием размеченных

данных. На обучающих данных должно быть точно известно, какое изображение представляет котика, а какое — собачку, само обучение при этом заключается в том, что изменяется параметр, на который ячейка умножает поступающий на нее сигнал, вырабатывая тот, который она передаст дальше. Этот параметр представляет собой искусственный синапс. И для того, чтобы обучение проходило, нужна обучающая программа, которая делает лишь одно — изменяет параметр искусственного синапса в зависимости от величины ошибки.

Первые искусственные нейронные сети оказались неработоспособными из-за того, что в них было слишком мало слоев и вообще слишком мало ячеек. «Весна» для них наступила тогда, когда объема памяти и быстродействия компьютеров стало хватать для многослойных сетей, которые стали называть глубокими, а процесс их обучения — глубоким обучением.

Помимо котиков и собачек у глубокой нейронной сети во внешнем слое может быть заложено огромное количество вариантов. Знаковым событием стало подписание фотографии компьютером с нейронной сетью, обученной командой Ильи Зуцкевера из Google в 2014 г.: «Группа людей, играющих во фризби». Натренированная нейронная сеть безошибочно выбирает в имеющемся наборе вариантов ответа подходящий — в соответствии с тем, как цвета и интенсивности распределены по пикселям входной матрицы. Теоретически нечто подобное можно было достичь и с помощью GOFAI, хотя решить эту задачу столь же успешно так и не удалось.

Нет принципиальных отличий и при обучении нейронной сети управлению автономным автомобилем компании “Tesla” в отсутствии водителя. Только вместо одной матрицы сканера или фотоаппарата здесь несколько радаров и (или) лидаров, а вместо подходящих слов «фризби»,

«регби», «футбол», «группа» и т. п. здесь управляющие движения: «прибавить газу», «5 Н на педаль тормоза», «поворот руля влево на 15°8г». Каким именно образом входящий сигнал преобразуется на выходе в слова или в управляющую команду, зависит лишь от того, какие численные значения приобрели в процессе обучения параметры искусственных синапсов. На практике это означает, что ни программисту, ни кому-либо еще не удастся выяснить, что именно на картинке позволяет отличить фризби от хоккея. Котики иногда бывают очень похожи на собачек, а собачки — на котиков. И человек, и машина могут ошибиться, разбирая фотографии, где такие встречаются. И человек в случае ошибки может легко объяснить, что именно сбило его с толку. Машина сейчас ошибается примерно вдвое реже, но у нас нет никакой возможности узнать, что ее сбивает с толку в тех случаях, когда она ошибается, и что ей помогает не ошибиться в тех случаях, когда человек ошибается, а она нет.

Бывают еще более странные ситуации, когда, например, на фотографии котиков и собачек добавляется мелкий, практически незаметный человеческому глазу муар или цветовой шум. Процент ошибок для человека остается неизменным. Но нейронная сеть, продолжая безошибочно определять, например, котиков, всех собачек вдруг записывает в гамадрилы. Это значит, что ее нужно обучать заново, добавив в обучающий набор данных кроме чистых фотографий также фотографии с добавленной туда помехой. Почему человеку не мешает то, что радикально сбивает с толку машину, еще предстоит выяснить.

В некоторых случаях подобная «непрозрачность» в принятии решения искусственным интеллектом может оказаться неудобной. Например, Герман Оскарович Греф, председатель правления Сбербанка России, любит повторять, что до 80 % исковых заявлений в Сбербанке

сейчас генерируются в автономном режиме (этот вопрос мы будем подробно обсуждать ниже). Это означает, что загруженные в компьютер данные преобразуются в текст искового заявления искусственной нейронной сетью, натренированной на некотором количестве модельных исков, которые люди сочли идеально соответствующими имевшимся при их составлении исходным документам. После того как тренировки завершились, компьютер стал действовать в режиме черного ящика, преобразующего исходные данные в итоговое заявление, следуя логике, восстановить которую принципиально невозможно.

В своей книге *«Искусственный интеллект. Пределы возможного»* **Мередит Бруссард** рассказывает о том, почему в декабре 2016 г. (запомним эту дату!) Американская ассоциация специалистов по компьютерным вычислениями (ACM) впервые с 1992 г. внесла изменения в свой этический кодекс. Поводом для этого стало внезапно обнаружившаяся расовая дискриминация у компьютерной системы COMPAS, разработанной для предсказания вероятности рецидива преступного поведения у отбывающих свое наказание граждан. Двое журналистов онлайн-издания ProPublica задались вопросом, насколько часто COMPAS ошибается? Как часто белый преступник с высокой вероятностью рецидива так и не совершает повторного преступления и как часто чернокожий преступник не подтверждает аналогичного высокого рейтинга для себя? Оказалось, что в отношении белых вероятность такой ошибки вдвое меньше, чем в отношении черных. «Несправедливость математически неизбежна», — делают парадоксальный вывод авторы публикации и даже выносят его в заголовок.

«Алгоритмы не работают объективно, поскольку люди внедрили в них свои стереотипы, — делает свой вывод Бруссард. — Техношовинизм заставляет людей

думать, что математические формулы, лежащие в основе программного кода, каким-то образом более справедливы в решении социальных проблемы — но это не так».

И главная ее ошибка здесь, как ни странно, именно в том, что тут не было никаких математических формул. Программный код написался сам, в этом и заключается смысл машинного обучения. Проблема тут, скорее, в том, что люди только определяют, какие данные используются для обучения нейронной сети, но проверить, как эти данные будут интерпретироваться дальше, можно только опытным путем. Для АСМ сложившееся положение вещей оказалось новым, что выразилось в проникновении аутогенного машинного кода внутрь социальной жизни. Это новое обстоятельство заставило их изменить свой этический кодекс.

Техношовинизм, о котором здесь пишет Бруссард, представляет собой довольно важную этическую проблему и без прямой связи с искусственным интеллектом. Нельзя сказать, чтобы термин был общеупотребительным, но явление достаточно хорошо известно: существует убеждение, что социальные проблемы могут разрешаться сами собой с развитием технологий. Благодаря промышленной революции целым континентам удавалось обеспечить рост уровня жизни за счет, как казалось, перехода к новому, более технологичному способу производства. Отсюда рождалась иллюзия, что необходимо максимально способствовать развитию инноваций и технологическому прогрессу и тогда социальные проблемы будут решены сами собой. Но мы говорим, что это иллюзия, так как сама промышленная революция была бы невозможна без определенных социальных предпосылок. Техношовинизм в этом смысле оказывается формой или, точнее, проявлением технологического детерминизма — философской установки, ограниченность которой не раз критиковалась философами.

Тем не менее мы должны отдавать себе ясный отчет, что социальная реальность, в которой каждый экономический акт сопровождается появлением или изменением того или иного письменного документа, принципиально отличается от социальной реальности, в которой акторы обмениваются только устными сообщениями. Аналогично внедрение электричества в производственные процессы, до того приводившиеся в действие исключительно силой пара или падающей воды, радикально их изменяет, что, в свою очередь, приводит и к социальным изменениям.

С некоторого времени инновации, которые оказались убийственными для прежнего уклада, стали называть *подрывными* (disruptive), что по-русски звучит несколько двусмысленно, поскольку на ум прежде всего приходят диверсанты-подрывники, собирающиеся пустить под откос военный эшелон. Подрывная инновация может оказаться не менее разрушительной, хотя и не такой эффектной. Развитие технологии звукозаписи, синхронизированной с кинематографом, в начале XX в. оказалось подрывной для всей индустрии немого кино. Как мы знаем, большинству актеров немного кино пришлось искать себе новую работу.

Компьютеризация практически всех информационных потоков в конце XX — первом десятилетии XXI вв. также породила принципиально новую социальную реальность, которую стали называть компьютерной цивилизацией.

1.3. Компьютерная цивилизация

В истории человечества принято выделять несколько ключевых событий, радикально изменивших образ жизни человечества. По понятным причинам эти ключевые события с XIX в. называют революциями, хотя

по времени они оказывались растянуты на десятилетия и даже века. Первое такое событие — *неолитическая революция,* когда представители вида Homo Sapiens массово стали отказываться от охоты и собирательства, предпочитая им земледелие и оседлый образ жизни. В этот период люди изменили свои привычки: стали возделывать поля, выращивая на них те культуры, которые считали полезными, и следя за тем, чтобы на этих полях не росло ничего из того, что они полезным не считали. Точно также рядом с ними стали появляться животные, которых они считали полезными и которые, возможно, позволяли им сохранять дистанцию с теми животными, которых они полезными не считали или считали опасными. В результате этого процесса, получившего название «одомашнивание» (доместикация), появились новые виды растений — вроде культурной пшеницы однозернянки (Triticum monococcum) и новые виды животных — вроде домашней козы (Capra hircus), которые до того в дикой природе не водились. Эти новые виды не смогли бы выжить без постоянной заботы со стороны человека.

Но неправильно было бы думать, что появление этих видов — всегда результат разумной и целенаправленной деятельности человека. По крайней мере поначалу это был, скорее, обоюдный эволюционный процесс, когда все участники получали определенные преимущества, не осознаваемые полностью никем или даже не осознаваемые вообще.

Не менее важным для формирования человека и человеческого общества стало появление письменности — его также называют первой информационной революцией. И если существование симбиозов в дикой природе не редкость, так что для человека при одомашнивании растений и животных можно указать на некоторые образцы, хотя и роль их неясна и сомнительна, то письменность

по большому счету беспрецедентна. Неслучайно в большинстве мифологических систем изобретение письменности описывается как передача соответствующего знания людям богами. Здесь тоже нередко говорят о доместикации, когда это изначально сугубо элитарное знание стало доступно большинству людей. Этот процесс принято называть второй информационной революцией и связывать с изобретением книгопечатания. Ее нередко также связывают с именем конкретного человека — **Иоганна Гутенберга** (ок. 1400–1468), и даже называют *гутенберговской революцией,* однако стоит заметить, использование полиграфического пресса и даже наборных шрифтов встречалось кое-где и раньше. Знаменитый фестский диск, обнаруженный в 1908 г. Луиджи Пернье, участником археологической экспедиции Федерико Хальберра на острове Крит, несет на себе набранный и отпечатанный текст. Однако именно в XV–XVI вв., то есть непосредственно вскоре после того, как Гутенбергом был отпечатан первый вариант его Библии в 1453 г., наблюдался быстрый рост грамотности среди самых широких слоев населения Европы. И у большинства историков не вызывает никакого сомнения связь этого процесса с другой знаменитой революцией — *научной революцией XVII в.,* начало которой, по общему мнению, было положено появлением из-под пресса печатного станка Эльзевиров сочинения Николая Коперника «О вращении небесных сфер» в 1543 г.

Мы не будем сейчас обсуждать, почему ничего подобного научной революции XVII в. не случилось раньше и в других цивилизациях, хотя, казалось бы, к этому были очень близки и Древняя Греция в эллинистический период, и Древний Китай. На этот счет есть разные и очень обстоятельные теории, ни одна из которых не может считаться окончательной. Но любопытно отметить синхронность нескольких разных процессов. Мы не можем

категорически утверждать, что вторая информационная революция была одной из причин научной революции, но эти два процесса, безусловно, развивались синхронно. Многие из героев научной революции XVII в. отличались исключительными способностями к рукоделию — были талантливыми инженерами и механиками, как мы бы сказали сейчас, или артизанами, как сказали бы тогда. Механическими игрушками в детстве славился Ньютон, поменяв их в зрелом возрасте на тигли и печи алхимической лаборатории, свои мастерские были у **Галилео Галилея** (1564–1642), достигшего непревзойденного в Европе мастерства в шлифовке стекол, что позволило ему делать прекрасные телескопы. Мы не знаем, насколько уверенно держал в руках слесарный инструмент Лейбниц, но он был исключительно изобретателен — именно обдумывая конструкции гипотетических вечных двигателей, он пришел к одной из первых формулировок принципа сохранения энергии. Можно тут вспомнить изобретателей и механиков предшествовавшего столетия — Леонардо да Винчи, Мариано ди Якопо, известного как Таккола, Агостино Рамелли. Для создания своих весьма остроумных устройств этим древним инженерам приходилось изготавливать свои собственные инструменты, так как, напомним, в те времена даже покупка обыкновенных молотка с гвоздями представляла серьезную проблему. И хотя ученые могут долго спорить, до какой степени возникшая благодаря этим техническим гениям XV–XVI вв. мастеровая культура предопределила изощренную экспериментальную машинерию титанов научной революции, не вызывает никакого сомнения факт их синхронности.

В предшествовавшие научной революции XVII в. столетия шло бурное развитие математических искусств. Можно спорить относительно факторов, предопределивших это бурное развитие, но нельзя сомневаться, что

без ее плодов научная революция была бы невозможна. В первую очередь это касается символьной записи математических соотношений — до XVI в. они записывались словами — и проникновения алгебраических методов в геометрию, достигшего своей кульминации в аналитической геометрии **Рене Декарта** (1596–1650). На этом пути был открыт математический анализ бесконечно малых, позволявший находить функции по их производным, то есть, в частности, решать уравнения движения ньютоновской механики.

Кроме аналитических методов математики, немного опережая их, развивались вычислительные методы практической геометрии и коммерческой арифметики. Заметим, что построение траекторий небесных тел с помощью законов механики требовали от Ньютона не только правильно находить формулу удовлетворяющего уравнениям решения, но и рассчитывать численные параметры этих траекторий. В историю науки Исаак Ньютон вошел не только как гениальный ученый, но и как выдающийся вычислитель.

Идея вычислительной машины в XVII в. казалась еще немного диковинной, но в принципе уже вполне понятной. Проблема заключалась лишь в том, что реализовать ее сугубо механическими средствами было невозможно.

Впрочем, и тут было несколько эпохальных достижений. К одному из них следует отнести жаккаровский станок, названный так по имени изобретателя **Жозефа Мари Жаккара** (1752–1834). О нем можно сказать то же, что Ньютон сказал о себе: «Я стоял на плечах гигантов», — поскольку у него было много предшественников. Идею достаточно хорошо иллюстрируют шарманка или музыкальная шкатулка: звуки извлекаются в зависимости от того, в каком порядке выступы на вращающемся диске задевают металлические полоски разной длины. Эта идея

была перенесена в ткацкое ремесло благодаря перфорированным пластинам: отверстия в пластине направляли детали станка таким образом, что на ткани возникал рисунок. Пластины можно было менять и соединять в ленту, и в зависимости от их последовательности рисунок мог быть каким угодно. В сущности, это была первая практическая реализация ключевой идеи вычислительных машин, сформулированной значительно позже: по мере возможности программное обеспечение (софт) должно быть отделено от их физических компонент (аппаратной части, или харда).

Понятно, что такое разделение возможно лишь относительно. В случае жаккаровского станка софт — это не сами перфорированные пластины, а тот порядок, в каком на них располагаются отверстия. В соответствии с провидческой идеей Лейбница, в XX в. их стали представлять в виде последовательностей нулей и единиц — двоичного кода. Любой современный софт может быть представлен в виде двоичного кода, но для его записи нужен какой-то физический носитель — перфокарты, перфолента, ферромагнитные ячейки дисков или лент, открытые или запертые лампы или транзисторы... Физические устройства требуются и для считывания кода, и для самих вычислений, поскольку всякое вычисление — это физический процесс.

Тем не менее у софта есть своя собственная жизнь. Мы знакомы с ее проявлениями по движущимся картинкам на мониторах компьютеров и дисплеях смартфонов. Но она незаметно течет внутри всех электронных устройств даже тогда, когда они пребывают в покое, в выключенном или, точнее говоря, в «спящем» состоянии. Каждую секунду миллиарды нулей и единиц, хранящихся в одних ячейках памяти, преобразуются в другие миллиарды нулей и единиц совершенно упорядоченным

образом. Управляют этим стремительным преобразованием тоже нули и единицы, хранящиеся в других ячейках памяти, и законы этих преобразований совершенно тождественны независимо от того, что именно в физическом отношении представляют собой эти нули, единицы, ячейки, да и сами вычисления — в простейшем случае, реализованном Жаккаром, — это были деревянные дощечки, упорядоченные отверстия в них и разные крючки, шнуры, лицы и иглы, благодаря которым биты и байты, закодированные на перфорированной дощечке, превращались в правильный узор на ткани. Сейчас это микроскопические транзисторы микросхем чипа, слабые электромагнитные поля и токи.

Один из ведущих специалистов по машинному обучению и искусственному интеллекту Эндрю Ын в своей публичной лекции в начале 2018 г. сравнил технологии искусственного интеллекта сейчас с электричеством сто лет назад. Смысл сравнения совершенно прозрачен: вся промышленная революция совершалась исключительно на механической энергии падающей воды с очень постепенно возрастающей ролью паровых машин. До середины XIX в. электричество если и использовалось, то исключительно в разного рода игрушках и фантасмагориях. Но уже к 1920-м гг. было практически невозможно найти хоть какой-то технологический процесс, в котором хоть как-то не использовалось бы электричество: освещение, датчики, электромоторы, телеграфная, телефонная или радиосвязь. Это была еще одна подрывная, или дизруптивная технология. Относительно технологий искусственного интеллекта сказанное Ыном — это в основном прогноз, но уже сейчас большая часть даже бытовых приборов в той или иной степени представляют собой компьютер — будь то автомобиль или телевизор. И уже совсем близко будущее, когда чип с микросхемой будет внутри каждого

холодильника и настольной лампы. А отсюда уже один шаг до внедрения внутрь этого чипа и искусственной нейронной сети, способной к обучению. Всякое осознанное или подсознательное действие человека сопровождается миллиардами и триллионами вычислений, совершающимися в окружающих его вычислительных устройствах, которым, по сути, является всякий чип. Мы живем в компьютерной цивилизации.

Ключевые понятия: компьютерная цивилизация, вычислимость, сознание, искусственная нейронная сеть, машинное обучение, искусственный интеллект, арифмометр Лейбница, вычислительная машина Паскаля, станок Жаккара.

Контрольные вопросы

1. Когда и в каких условиях в западноевропейской культуре возникло представление о тождественности мышления вычислениям?
2. Как эволюционировала идея материального переносчика взаимодействий в вакууме после Ньютона и какое отражение она нашла в теории вычислений?
3. Что позволило Тьюрингу положительно ответить на вопрос «может ли машина мыслить?»?
4. Какие исторические обстоятельства сопутствовали введению в научный оборот термина «искусственный интеллект»?
5. Какие основные сложности возникают при попытке научно определить понятие «искусственный интеллект»?
6. Что связывает техношовинизм с технологическим детерминизмом?
7. Что общего у искусственного интеллекта сейчас и электричества сто лет назад?

Практико-ориентированные задания

1. Самостоятельно найдите информацию и проведите сравнительный анализ первых вычислительных машин — Лейбница, Паскаля, Бэббиджа. Составьте таблицу. В чем их сходство, в чем различия?

2. Какие события XX в. позволяют нам утверждать, что сейчас на Земле сложилась компьютерная цивилизация?

3. Проанализируйте информационные революции и их роль в истории человечества.

Темы сообщений, докладов и эссе

1. Искусственный интеллект — трудности определения.

2. Мысль и вычисление.

3. Тьюринг: может ли машина мыслить?

4. Искусственная жизнь в истории европейской культуры.

ГЛАВА 2.
ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ЛЕГАЛЬНОЕ ОПРЕДЕЛЕНИЕ И ПРИЗНАКИ

В результате изучения материалов главы обучающийся должен

знать:

– действующие документы в сфере искусственного интеллекта;

– основные доктрины, понятия, категории в области взаимодействия человека с искусственным интеллектом;

уметь:

– отличать искусственный интеллект от иных киберфизических систем;

– давать правовую оценку фактическим обстоятельствам дела и устанавливать правовые нормы в области взаимодействия социума с искусственным интеллектом в зависимости от правового поля отдельных государств;

владеть навыками:

– анализа различных явлений, фактов, правовых норм и правовых отношений в сфере правоотношений с AI;

– юридически грамотной квалификации отношений с AI.

2.1. Терминология, используемая в сфере правового регулирования искусственного интеллекта[1]

Изменения, происходящие в последние десятилетия в общественной жизни и обусловленные стремительным развитием технологий, делают необходимым совершенствование существующего законодательства. Поэтому во всем мире развернута сейчас широкая дискуссия по поводу правовой среды, способствующей внедрению и применению инновационных технологий, к которым отнесены цифровые технологии вообще и технологии искусственного интеллекта в частности. Следует отметить, что и различного рода киберфизические системы, роботы и объекты робототехники занимают особое место в силу возможности использования практически в любых сферах, поэтому так важно разграничить их терминологически.

Основной из задач, стоящих на современном этапе развития в целях правового регулирования инновационных технологий, принято считать легальное определение искусственного интеллекта. В законодательстве России пока такое определение не закреплено, однако в национальной стратегии «Развитие искусственного интеллекта в Российской Федерации» (далее по тексту — Стратегия AI), утвержденной Указом Президента Российской Федерации от 10 октября 2019 г. № 490

[1] Параграф написан на основе научных статей: *Баракина, Е.Ю.* К вопросу формирования перспективной терминологии в области правового регулирования применения киберфизических систем / Баракина Е.Ю. // Российская юстиция. 2020. № 1. С. 70–73; *Баракина, Е.Ю.* К вопросу об установлении экспериментального правового режима в области применения искусственного интеллекта / Е.Ю. Баракина // Российская юстиция. 2021. № 2. С. 64–67.

«О развитии искусственного интеллекта в Российской Федерации», и двух федеральных законах (Федеральный закон от 24 апреля 2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации — городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона "О персональных данных"» и Федеральный закон от 31 июля 2020 г. № 258-ФЗ (последняя редакция) «Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации») есть трактовка не только искусственного интеллекта, но и технологий искусственного интеллекта. В соответствие со ст. 5 Стратегии AI:

«а) **искусственный интеллект** — это комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые как минимум с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений;

б) **технологии искусственного интеллекта** — технологии, основанные на использовании искусственного интеллекта, включая компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта».

В научном мире отсутствует единое мнение по поводу объема и отличительных признаков, присущих технологиям искусственного интеллекта, как не осмыслены в целом элементы новых правоотношений, связанных с применением киберфизических систем, роботов и объектов робототехники. *Резолюция Европарламента от 16 февраля 2017 г. «Нормы гражданского права о робототехнике»* содержит призыв сформулировать «общепринятые и универсальные определения терминов "киберфизические системы", "автономные системы", "умные автономные роботы", а также их подкатегорий». Данная тематика широко обсуждается в общественных и правительственных организациях Южной Кореи, США, Японии, Австралии, Сингапура, Индии и некоторых других стран.

Есть много версий о происхождении самого термина *«киберфизические системы»* (CPS). Наиболее распространенная — что его впервые ввела в научный оборот **Элен Гилл** (Helen Gill), директор по встроенным и гибридным системам в Национальном научном фонде США в 2006 г. Речь не шла о точном определении, но сам по себе этот термин подразумевал своего рода комбинацию вычислительных и физических систем, софта и техники, когда программа генерирует решение, технически исполняемое без вмешательства человека, автономно. При этом подчеркивалось, что пока осуществима лишь частичная автономность. Со временем этим термином стали также обозначать «встроенные системы реального времени», «распределенные вычислительные системы», «автоматизированные системы управления техническими процессами и объектами», «беспроводные сенсорные сети». Есть и другие трактовки: например, подчеркивающие наличие в CPS различных природных объектов, искусственных подсистем и управляющих контроллеров,

позволяющих представить такое образование как единое целое.

Относительно определения понятия **«киберфизическая система»** мнения ученых и практиков можно условно разделить на две группы. Представители первой группы утверждают, что данные системы имеют кибернетическое начало, являются компьютерными аппаратными и программными технологиями, отличаются качественно новыми механизмами исполнения, интегрированными в окружающую среду, которые анализируют и реагируют на происходящие изменения, а также способны к принятию интеллектуальных решений (самообучение и адаптированность). Наиболее типичные суждения ученых и практиков, которые можно отнести к первой группе, представлены на рис. 1.

Это интеллектуальные системы, которые включают вычислительные (т. е. аппаратные и программные) и физические компоненты, легко интегрируемые и тесно взаимодействующие для восприятия изменяющегося состояния реального мира.

Это интеллектуальные сетевые системы со встроенными датчиками, процессорами и исполнительными механизмами, которые предназначены для восприятия и взаимодействия с физическим миром (включая пользователей) и поддерживают гарантированную производительность в режиме реального времени в критически важных для безопасности приложениях.

Автоматизированная интеграция физических и цифровых компонентов, охватывающая мониторинг физической реальности через датчики и возможность воздействовать на эту реальность через исполнительные механизмы.

Рис. 1. Первая группа мнений ученых и практиков об определении понятия «киберфизическая система»

Вторая группа исследователей выделяет как сущностную черту таких систем соединение физического и информационного пространств. Несколько типичных мнений ученых и практиков, которых можно отнести к первой группе, представлены на рис. 2.

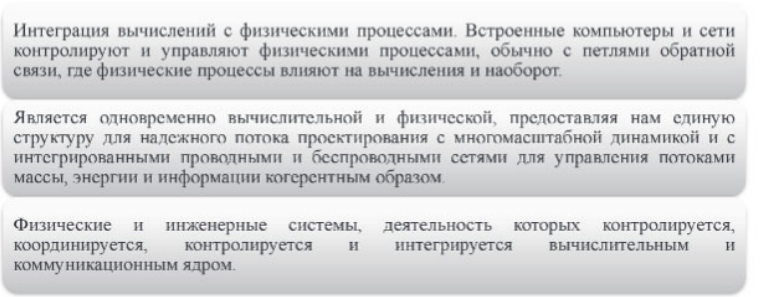

Рис. 2. Вторая группа мнений ученых и практиков об определении понятия «киберфизическая система»

Сравнивая указанные подходы к определению термина «киберфизическая система», можно утверждать, что подходы не имеют противоположности во мнениях. Первая группа формулирует определение данного термина исходя из механизма функционирования этих систем, а вторая описывает существенные признаки, чем, в сущности, по их мнению, является указанная система.

В современной научной литературе выделяется несколько ***признаков киберфизических систем:***

• *Наличие «интеллекта» или «ума».* Оба термина, по сути, предполагают одно и то же, при этом не делается, как правило, различения между «старым добрым AI», когда «интеллектуальность» CPS программируется человеком, и натренированной нейронной сетью, когда «интеллектуальность» приобретается благодаря машинному обучению. В различных сферах используют формулировки «умные дома», «умные сети», «умные производства» и т. п.

• *Наличие в дополнение к процессору (чипу) и «программному обеспечению» внешних датчиков и терминальных устройств, обеспечивающих влияние на окружающую систему «реальность»*, то есть человек

волен при создании киберфизической системы определять и способ, которым решается задача, и объем необходимых для этого навыков (глубину обучения), и способ влияния на окружающий мир. В научном исследовании, проведенном в 2016 г. Исследовательским центром Европейского парламентаризма STOA (Science and Technology Opinions Assessment) по заказу Европарламента, его авторы, отвечая на поставленный вопрос «что такое киберфизические системы?», в первую очередь отмечают аспект возможности их взаимодействия с окружающим миром, а также необходимость контроля и несения ответственности создателей за последствия принятых системами решений.

• *Наличие «системности» при функционировании всех элементов киберфизических систем.* Несмотря на то, что теории систем, как считается, уже около ста лет, с определением термина «система» дело обстоит примерно так же, как и с термином «интеллект», но есть более или менее общее согласие, что система должна объединять элементы произвольной природы, которые, отношениям и связям друг с другом, образуют некоторую целостность. Недостаток такого определения всем хорошо известен: определить термины «элемент», «целостность», «отношение», «связь» и даже «природа» ничуть не легче, чем термин «система». Но благодаря увязыванию их всех в одном предложении, у нас появляется шанс полагаться хотя бы на языковую интуицию.

Принимая во внимание, что киберфизические системы способны получать, обрабатывать и хранить информацию, к ним можно применять также термин **«информационных систем»**, определение которого закреплено в российском законодательстве (ФЗ «Об информации, информационных технологиях и о защите информации»): «информационная система — совокупность содержащейся в базах данных информации и обеспечивающих

ее обработку информационных технологий и технических средств», но из этого определения следует, что информационные системы являются лишь одной из подсистем, обеспечивающих киберфизической системе «интеллектуальность».

На основании вышеизложенного можно сделать вывод, что **киберфизическая система** — это система, обладающая интеллектуальными механизмами принятия решений на основе интегрированных информационно-физических программных элементов с возможностью их исполнения, то есть физического взаимодействия с окружающим миром.

Определение в таком виде объединяет отличительные признаки и не противоречит применению в различных сферах общественных отношений, кроме того, может относиться к искусственному интеллекту, роботам и объектам робототехники.

Анализ международного опыта показывает, что киберфизические системы включены правительствами многих стран в число приоритетных для развития инновационных технологий. Причем некоторые из них, — например, правительство США, — считают критически важным внедрение таких технологий и их применение в целях защиты национальных интересов. Основные сферы применения киберфизических систем представлены на рис. 3.

Таким образом, нам удается «окружить» термин «киберфизическая система» другими логическими категориями того же ряда — «умный дом», «умный город», «интернет вещей» и т. п.

Умные дома и умные города, умные производства и умные сети, беспилотный транспорт и транспортные сети, интернет вещей, интеллектуальные встроенные системы являются системами, которые либо исполняют

Рис. 3. Основные сферы применения киберфизических систем

решения с помощью механизмов сконструированных человеком, либо принимают их на основе полученной, переработанной и хранящейся в системе информации, не оказывая влияния на окружающий мир до принятия окончательного решения человеком и, в конечном итоге, выполняя функции в соответствующей сфере общественных отношений. Все эти системы некоторые авторы объединяют одним — умная среда, которая обладает рядом характеристик, представленных на рис. 4.

Некоторые авторы отмечают еще важную характеристику — возможность «приспосабливаться к нуждам пользователей для улучшения их взаимодействия с внешней средой», что фактически и является целью применения киберфизических систем.

Наиболее часто встречающиеся термины, связанные с применением киберфизических систем, представлены в табл. 1.

Разработка и закрепление термина «киберфизические системы» представляется весьма важной, так как без четкости обозначения объема и указания

Рис. 4. Характеристики умной среды

на отличительные признаки невозможно формирование подходов к правовому регулированию.

2.2. Искусственный интеллект: правовые аспекты

В связи с отсутствием единого подхода к определению понятий «интеллект» и «искусственный интеллект», о чем говорилось в гл. 1, с одной стороны, и необходимостью пользоваться этим термином, с другой, представляет интерес *«теория интегрированной информации»* (Integrated information theory, IIT), предложенная в 2004 г. итальянским специалистом по нейронаукам **Гвидо Тонони**[2], представляющая собой математически

[2] Современный вариант теории, известный как IIT 3.0, был опубликован в 2014 г.

Таблица 1

Термины, связанные с применением киберфизических систем

№ п/п	Термин	Определение
1	Умный дом	— киберфизическая система, обладающая интеллектуальными исполнительными механизмами и обеспечивающая управление жилым помещением или помещением с иным назначением
2	Умные города	— киберфизическая система, обладающая интеллектуальными исполнительными механизмами и обеспечивающая управление городской инфраструктурой с целью улучшения качества жизни
3	Умные производства	— киберфизическая система, обладающая интеллектуальными исполнительными механизмами и обеспечивающая управление производством с целью повышения его автоматизации, улучшения контроля и оптимизации всех производственных процессов
4	Беспилотный транспорт и транспортные сети	— киберфизическая система, обладающая интеллектуальными исполнительными механизмами и обеспечивающая управление транспортным средством или транспортной сетью с целью улучшения качества жизни, оптимизации соответственно функционирования и передвижения

Окончание табл.

№ п/п	Термин	Определение
5	Интернет вещей	— киберфизическая система, обладающая интеллектуальными исполнительными механизмами и обеспечивающая управление техническими средствами для улучшения качества жизни человека в соответствии с его целями
6	Интеллектуальные встроенные системы	— киберфизические системы, обладающие интеллектуальными исполнительными механизмами и обеспечивающие решение поставленных человеком задач с целью улучшения качества жизни, оптимизации и ускорения соответствующих процессов

точную теорию сознания, которая, по сути, утверждает, что информации для того, чтобы себя осознать, необходимо быть интегрированной. Для определения наличия сознания у произвольного физического субстрата теория Тонони предлагает «человеческий» подход: «сознание, началом которого являются самоочевидные, существенные свойства (аксиомы) опыта, и переводит их в достаточные условия для физического субстрата сознания». Данные аксиомы представлены на рис. 5 и фактически, по мнению автора, являются признаками сознания.

При том, что у этой теории есть довольно много сторонников, есть у нее и свои оппоненты. Но главное даже не это: наличие сознания немедленно исключит из рассмотрения все современные системы искусственного

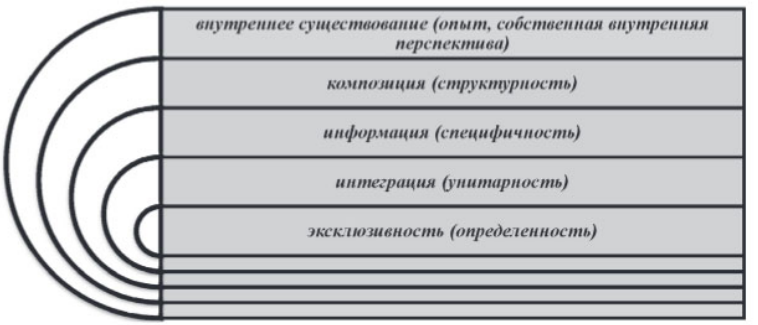

Рис. 5. Аксиомы «теории интегрированной информации»

интеллекта, очевидно бессознательные, и принципы достижения ими сложных целей значительно отличаются от когнитивных способностей человека. Для разрешения этой сложности **А.В. Незнамов и Б.В. Наумов** в своей работе *«Стратегия регулирования робототехники и киберфизических систем»* [3] предлагают подойти к проблеме более формально и отмечают, что основополагающую роль при таком формальном определении играют «компьютерные программные технологии, а их физическое проявление не имеет принципиального значения», выделяя важность соотнесения категории «искусственный интеллект» с категориями «машинное обучение», «нейронная сеть» и им подобными.

Об особенностях искусственного интеллекта на основе машинного (глубокого) обучения и искусственных нейронных сетей и его отличиях от GOFAI, «старого доброго искусственного интеллекта», уже говорилось в гл. 1: при использовании моделей машинного обучения данные вносятся таким образом, чтобы поставленная задача решалась без программирования алгоритма человеком

[3] Библиографическое описание всех цитированных источников приводится в конце книги.

(вручную). Три основных компонента такой системы представлены на рис. 6.

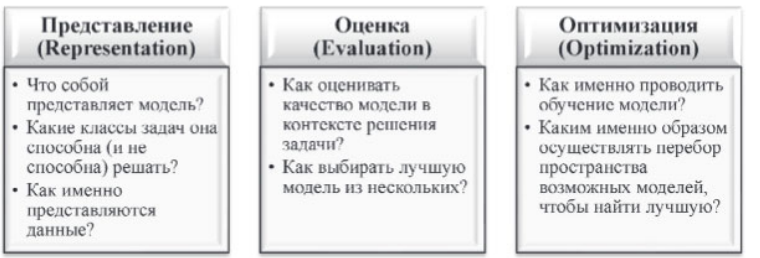

Рис. 6. Основные компоненты «машинного обучения»

Таким образом, **машинное обучение** *является получением и обработкой информации*. Как уже упоминалось ранее, данный процесс способствует принятию решения, но не может характеризоваться как наличие «разума» или «интеллекта». Относительно **«нейронной сети»** существует мнение, выраженное, например, в блоге финтех-компании DTI Algorithmic от 6 июля 2017 г. «Нейросети: как искусственный интеллект помогает в бизнесе и жизни», что это «один из способов реализации искусственного интеллекта». И в настоящее время именно нейронные сети являются, по мнению исследователей, «основным направлением по изучению возможности моделирования естественного интеллекта с помощью алгоритмов», которые основаны не на программировании, а на «обучении» (Васильев, 2018). Иначе говоря, **нейронная сеть или нейросеть** — *это способ реализации искусственного интеллекта, содержащий алгоритмы, взаимодействующие по принципу «естественного интеллекта»*.

Один из ведущих российских специалистов по правовому подходу к проблемам искусственного интеллекта **П.М. Морхат,** проанализировав различные подходы к этой проблеме, выделил несколько основных возможностей

и способностей искусственного интеллекта. Они представлены на рис. 7.

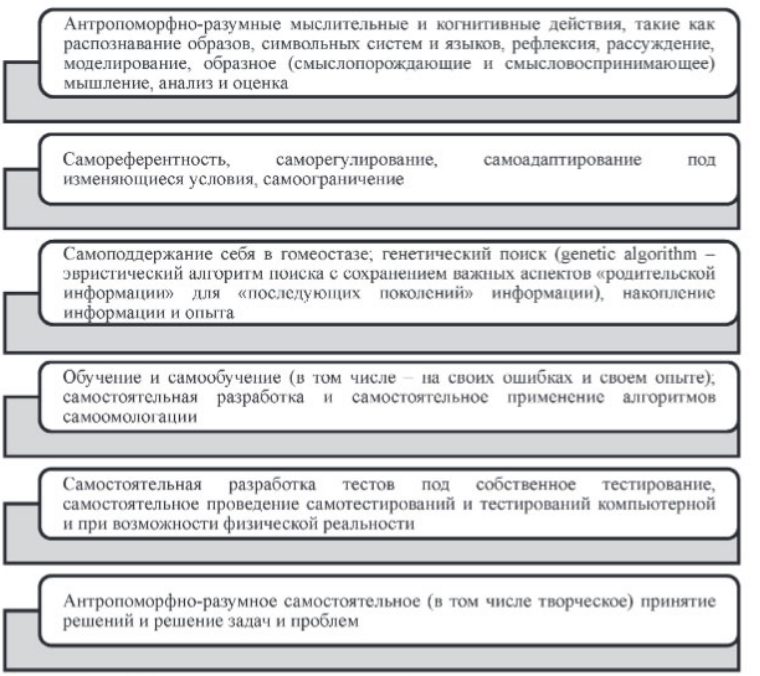

Рис. 7. Основные возможности и способности искусственного интеллекта

Отличительные признаки искусственного интеллекта были выделены А.А. Вас ильевым и Д. Шпонером (Васильев, 2018b) в их совместной работе «Искусственный интеллект: правовые аспекты». Они представлены на рис. 8.

В соответствии с указанными разграничениями П.М. Морхат полагает искусственный интеллект отдельным видом «полностью или частично автономной самоорганизующей (самоорганизующейся) компьютер-

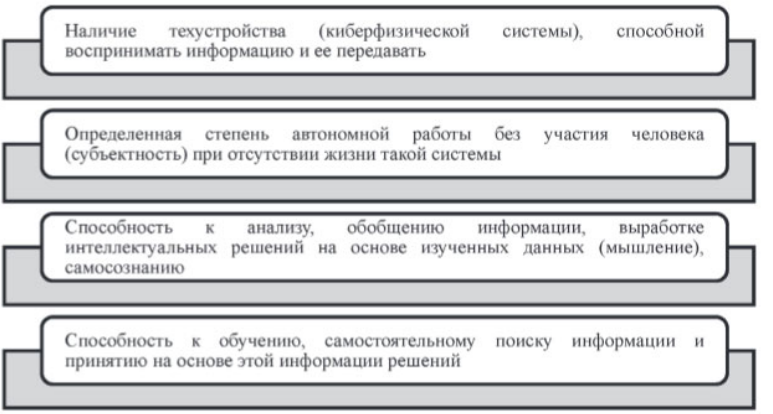

Рис. 8. Отличительные признаки искусственного интеллекта

но-аппаратно-программной виртуальной (virtual) или киберфизической (cyber-physical), в том числе био-кибернетической (bio-cybernetic), системы (юнит), наделенной/обладающей способностями и возможностями».

И мы примем его подход с тем, чтобы дать ограниченное внутренне непротиворечивое определение этому понятию, которым могли бы пользоваться дальше: **искусственный интеллект** — *это самоорганизующаяся система, обладающая искусственными средствами для взаимодействия с окружающей средой, принимающая решения на основании информации и в соответствии со способностями и возможностями.*

В ГОСТ 43.0.8-2017 установлено несколько терминов, определяющих данную дефиницию, которые представлены на рис. 9.

В настоящее время в зависимости от возможностей и способностей выделяют два вида искусственного интеллекта: «слабый», или узконаправленный (узкий), и «сильный», или общий. Вот как они определялись в законопроекте США (H. R. 4625 — 115th Congress):

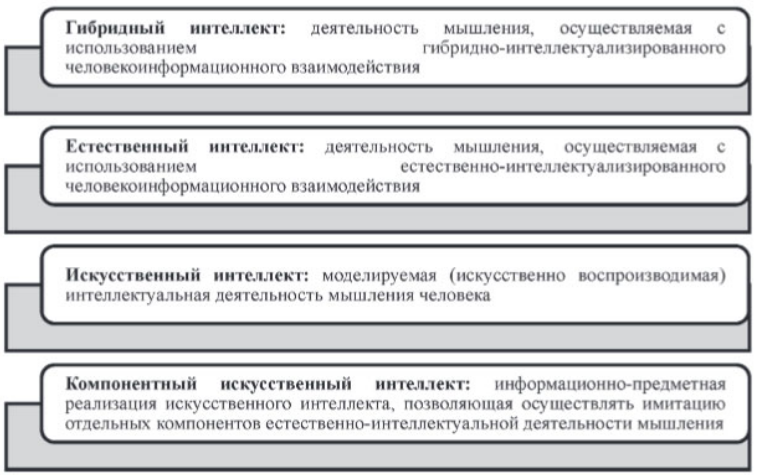

Рис. 9. Термины ГОСТ 43.0.8-2017, связанные с понятием «интеллект»

- *общий искусственный интеллект* (Artificial General Intelligence, AGI) означает условную будущую систему искусственного интеллекта, которая демонстрирует, по-видимому, интеллектуальное поведение, по крайней мере такое же продвинутое, как человек, в диапазоне когнитивного, эмоционального и социального поведения;
- *узкий искусственный интеллект* (Artificial Narrow Intelligence, ANI) означает систему искусственного интеллекта, которая обращается к конкретным областям применения, таким как стратегические игры, языковой перевод, самоуправляемые транспортные средства и распознавание изображений.

Следовательно, узкоограниченный искусственный интеллект может уже использоваться для решения различных задач, список которых сопоставим со сферами применения киберфизических систем, например медицина, юриспруденция, экономика, лингвистика, однако функционирование их возможно только в рамках

предписанных или установленных изначально при создании задач. «Сильный» имеет более широкую сферу применения, а значит и эффективность, и перспективность.

Искусственный интеллект в настоящее время развивается в качестве отдельной области знаний или науки искусственного интеллекта, хотя изначально относился к робототехнике. В соответствии с ГОСТ Р 60.0.0.4-2019/ИСО 8373:2012, «робототехника — это наука и практика проектирования, производства и применения роботов». Положения *«Модельной конвенции о робототехнике и искусственном интеллекте»*, разработанной в Исследовательском центре проблем регулирования робототехники и искусственного интеллекта, относят к объектам робототехники: «все категории роботов и роботизированные механизмы, а также киберфизические системы с искусственным интеллектом в любой форме», то есть объединяя данные науки, что в действительности приводит к подмене понятий.

Авторы программного документа *«Робототехники и искусственный интеллект»* (англ. *Robotics and artificial intelligence*) Королевской академии инженерных наук Великобритании, разработанного в ответ на запрос от Комитета по науки и технике палаты общин парламента от 2015 г., указывали, что это разные направления и разные технологии с возможностью пересечения и совместного использования. Отличия, в первую очередь, существенны при рассмотрении умных роботов, обладающих сильным искусственным интеллектом, которые фактически не имеют четких рамок сфер применения, следовательно, уже не могут относиться к объектам робототехники, поэтому положение, что любые киберфизические системы являются объектами робототехники, не представляется возможным.

Таким образом, киберфизические системы, искусственный интеллект, роботы и объекты робототехники тесно взаимосвязаны. Данная взаимосвязь представлена на рис. 10. Объекты робототехники — это киберфизические системы, в том числе системы управления, обеспечивающие решение поставленных человеком задач, роботы и интеллектуальные роботы. Однако последние являются юнитами узкого искусственного интеллекта, следовательно, относятся и к объектам искусственного интеллекта. Тогда, к объектам искусственного интеллекта относятся киберфизические системы (интеллектуальные роботы) и юниты общего искусственного интеллекта.

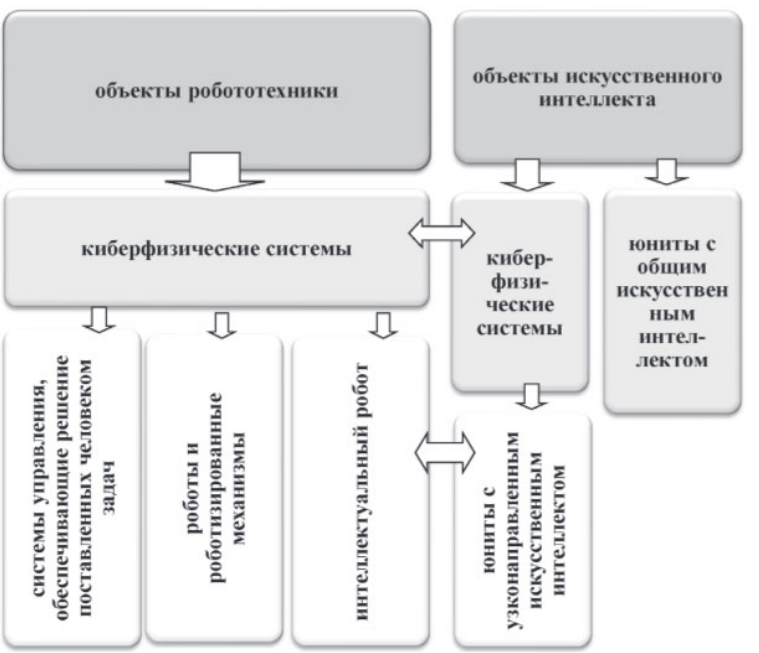

Рис. 10. Взаимосвязь объектов робототехники и носителей искусственного интеллекта

Назрела необходимость совершенствования законодательства для создания правовой среды, способствующей внедрению и применению инновационных цифровых технологий, в том числе киберфизических систем, искусственного интеллекта, роботов и объектов робототехники, которые занимают особое место в силу открывающихся возможностей по их практическому использованию в сферах общественной жизни.

Киберфизические системы включены правительствами многих стран в число приоритетных направлений развития инновационных технологий, которые считают критически важным их внедрение и применение в целях защиты национальных интересов. Следовательно, необходимо сформулировать термин «киберфизическая система», раскрывающий ее сущность и указывающий на отличительные признаки. Умные дома и умные города, умные производства и умные сети, беспилотный транспорт и транспортные сети, интернет вещей, интеллектуальные встроенные системы являются киберфизическими системами, которые либо исполняют решения с помощью механизмов, сконструированных человеком, либо принимают их на основе полученной, переработанной и хранящейся в системе информации, не оказывая влияния на окружающий мир до принятия окончательного решения человеком и, в конечном итоге, выполняя узконаправленные функции в соответствующей сфере.

Для понимания термина «искусственный интеллект» важно соотношение его с понятием «машинное обучение», определяемым как процесс получения и обработки информации, и понятием «нейронная сеть», означающим способ реализации искусственного интеллекта. Именно процесс и способ реализации, обусловили разделение юнитов на носителей узкого и общего искусственного интеллекта, так как электронным лицом может стать только

юнит общего искусственного интеллекта в силу своих возможностей и способностей.

2.3. Виртуальная реальность и онтология искусственного интеллекта

Возникновение электронно-виртуальной реальности как принципиально «нового типа искусственно созданной виртуальной реальности» во второй половине XX в. становится предметом осмысления большого числа онтологических исследований. Как отмечает в статье «Онтология электронно-виртуальной реальности» Т.Д. Стерледева (2017), онтология начинает менять свое содержательное значение и заключает в себе применительно к своему новому объекту «определенный тип видения бытия… и теоретическую модель такого видения» (с. 191), отличительными признаками которой являются:

1) определение электронно-виртуальной реальности как особого типа реальности, созданного человеческим (естественным) интеллектом для человека, в отличие от материальной реальности, созданной Природой (Богом, Абсолютной идеей, Вселенной, Космосом) для всех без исключения земных существ. Идеальная же реальность признается, в большинстве своем, продуктом сознания только человека, который создает свой собственный искусственный мир, в том числе и в киберпространстве;

2) существовавшее до второй половины XX в. онтологическое знание окружающей человека действительности (природного мира) заключалось в исследовании и осмыслении явлений, создаваемых и управляемых только Природой вне зависимости от воли и желаний самого человеческого общества (коллективного сознания) и каждой отдельной личности (индивидуального сознания), — пожаров, наводнений, извержений вулканов,

землетрясений, эпидемий болезней и др.). Онтология же виртуальной (гибридной) реальности не просто зависит от человека, но и управляется им, она напрямую ориентирована на его интересы и запросы, в том числе и в потребительской материальной сфере;

3) в материальном мире существует только трехмерное пространство (длина, ширина и высота), в то время как в киберпространстве количество измерений создаваемого пространства бесконечно и зависит лишь от мощности соответствующего компьютера (киберфизической системы, нейронной сети, искусственного интеллекта) и желаний самого человека в целях достижения им наибольшего комфортного состояния в определенный период времени;

4) онтология природного мира основана на трактовке темпоральности в виде единственно возможного измерения — настоящего и будущего течения событий, в то время как движение назад, в прошлое, невозможно. В электронно-виртуальной реальности течение времени может быть различным, так как существует возможность моделирования типа времени с заранее заданными характеристиками.

Таким образом, современный человек одновременно существует по крайней мере в двух ипостасях — как реальный материальный объект, наделенный сознанием в природном (реальном) мире, и как аватар, виртуальная (электронная) личность, обладающая возможностью выбора внешности, возраста, возможностей в создаваемых им самим обстоятельствах жизни) в киберпространстве, созданном исключительно сознанием самого индивида, исходя из его собственных желаний и предпочтений, но обязательно при помощи искусственного интеллекта, коим могут быть наделены различные технологические объекты. Как отмечает В.А. Лекторский в статье «Возможны ли науки о человеке?» (2015), в эпоху развития

наноиндустрии человек начинает одновременно существовать «в разных мирах: не только в мире физических и биологических процессов, но и... в мире искусственном» (с. 6).

Поэтому на современном этапе развития информационных технологий особое значение приобретает онтология искусственного интеллекта как отдельное направление междисциплинарных исследований, представляющее собой широкую область научных исследований, включающую комплекс компьютерных наук, на основе которых создаются информационные технологии, инженерия знаний, обработка естественного языка, представление знаний, интеллектуальная интеграция информации, извлечение информации и управление знанием, а также **нано-, био-, инфо-, когно- (NBIC) технологии**. Основополагающее значение в этой связи приобретают вопросы исследования самой возможности познания, его механизма в отношении искусственного интеллекта, соотношения его с естественным интеллектом, возможности включения их в сферу социальных отношений, их правовое регулирование.

Решение всех этих проблем зависит от ответа на главный вопрос: обладают ли современные **технологические артефакты** (киберфизические системы, роботы, объекты робототехники, нейронные сети) сознанием и способностью к мыслительной деятельности, сходными с естественным (природным) интеллектом человека. Основной проблемой на сегодняшний день является возможность создания искусственного интеллекта полностью идентичного человеческому сознанию, а то и превосходящего его, если подобный AI будет создан по аналогии с нейронным строением мозга человека и имитирующим или способным репродуцировать мысль. На современном этапе известны примеры такой так называемой созидательной деятельности.

Первая музыка, созданная с использованием компьютера, появилась в 1957 г. в Bell Laboratories. Это была композиция длиной 17 секунд, которую ее автор Ньюман Гутман назвал The Silver Scale («Серебряная чешуя»). В 1959 г. советская ЭВМ «Урал-2» сочиняла небольшие мелодии под общим названием «Уральские напевы». Тогда это был, разумеется, GOFAI. Теперь музыку сочиняют, как правило, нейронные сети. Весной 2019 г. звукозаписывающий лейбл Warner Music Group заключил авторский договор не с человеком, а с нейронной сетью Endel, основанной на искусственном интеллекте, с целью создания «звукового ландшафта» (soundscapes) на основе цифрового ряда под определенное настроение человека общей численностью 20 альбомов. Endel является специальным приложением для смартфонов — «кроссплатформенной аудиоэкосистемой», способной учитывать время суток, когда будет прослушиваться музыкальная композиция, погоду, телесные и ментальные действия заказавшего его человека (пробуждение, отдых, концентрация внимания, спортивная тренировка и др.) с целью создания наиболее гармоничного музыкального фона под конкретного заказчика.

Первые пять альбомов, появившиеся в мае 2019 г., представляют собой особую музыку для сна и называются соответственно: «Ясная ночь», «Дождливая ночь», «Облачный полдень», «Облачная ночь» и «Туманное утро». Остальные 15 альбомов призваны оказывать влияние на состояние человека, при этом создатели Endel гарантируют, что искусственно созданная музыка должна минимум в шесть раз улучшить концентрацию мысли или в случае постановки иной цели наоборот, снизить нервозность человека, уменьшить беспокойство почти в четыре раза.

В апреле 2020 г. компания OpenAI выпустила Jukebox — нейросеть, которая генерирует музыку

в различных жанрах. Она может сгенерировать даже элементарный голос, а также различные музыкальные инструменты. Jukebox создает аудиосигнал напрямую, минуя символьное представление. Такие музыкальные модели имеют гораздо большую емкость и сложность, чем их символьные аналоги, что подразумевает более высокие вычислительные требования для обучения модели.

По аналогии с трактовкой человека как биосоциального существа, в котором сознание (идеальное) и тело (материальное) представляют собой неразрывную дихотомию, в науке и практике предпринимаются попытки создания единого виртуального «человека» (интеллектуальной системы), в котором искусственный интеллект будет помещен в тело робота или отдельный объект робототехники. Такие системы должны, по мнению Д.А. Поспелова (1994), обладать тремя базовыми функциями, выражаемыми в способностях:

а) предоставлять и обрабатывать знания, а также к самообучению;

б) рассуждать, «выдавая» логические умозаключения, являющиеся началом нового обобщенного знания, при использовании потенциала которого становится возможным рациональное планирование дальнейшей деятельности;

в) к взаимообщению между интеллектуальными системами и человеком на естественном языке; получать информацию об окружающем мире, основываясь прежде всего на восприятии через звук и зрение.

В начале XXI в. в научной литературе было определено, что создаваемые технологии на основе искусственного интеллекта должны: «думать» рационально и действовать рационально, как люди, то есть искусственный интеллект определяется как рационально действующий

«агент», обладающий способностью самостоятельно разрабатывать алгоритмы поведения, но не наделенный понятием человечности, которая является обязательной для естественного интеллекта. Электронная виртуальная реальность и искусственный интеллект, в соответствии с абиотической теорией А.И. Опарина, опубликованной им еще в 1960 г., возникают примерно по той же схеме, что и зарождение жизни на земле — из «протожизненных элементов-коацерватов», что дает возможность их регулировать так же, как и сферы жизнедеятельности социума.

Отдельной, пока не разрешенной проблемой, являются морально-этические компоненты использования искусственного интеллекта. Однако попытки смоделировать этику уже существуют: так, *система искусственного интеллекта "Scheherazade system"* уже сейчас способна генерировались тексты из краудсорсинговой платформы Amazon Mechanical Turk по признаку семантической схожести. Для реализации поставленной цели (получение лекарства в аптеке) самообучаемая система анализирует по хронологическому принципу все события, с какими люди могут столкнуться в действительности. При этом в алгоритм искусственного интеллекта был заложен принцип характерности обычного правомерного поведения человека, поэтому интерфейс выбирал не правонарушение (кражу лекарства), хотя это и было самым рациональным решением, а обычный поступок, основанный на моральных догмах.

Однако, на наш взгляд, чисто этическим такое поведение вряд ли можно назвать, все-таки морально-этические нормы не всегда могут соответствовать определенным алгоритмам, их побудительной причиной могут служить определенные чувства и ассоциации. Как отмечает А. Незнамов в своей рецензии на книгу германского футуролога

Герда Леонгарда «Технологии против человека» (2019), «машины никогда не будут людьми, а у технологий нет этики. Поэтому человек должен остаться человеком». Сам же Герд Леонгард (2018) в своей книге предложил такой короткий манифест цифровой этики, в соответствии с которым «человек имеет право:

1) оставаться естественным, то есть биологическим. Автор предупреждает, что наше право работать, пользоваться госуслугами и вообще вести нормальную жизнедеятельность не должно быть обусловлено размещением каких-либо устройств на теле, имплантацией чипов и т. д.;

2) быть медленнее технологий. Нельзя наказывать людей за несоответствие их производительности мощности алгоритмов. Естественные биологические ограничения нужно уважать;

3) не быть постоянно на связи, что уже сегодня становится элементом роскоши;

4) быть анонимным;

5) нанимать или привлекать людей вместо машин; необходимо налоговое стимулирование компаний, нанимающих людей, и дополнительные налоги на автоматизацию — в будущем».

Широкое использование искусственного интеллекта приводит к созданию нейрокомпьютерного интерфейса, основанного на непосредственном контакте человеческого и компьютерного сознания, результатом чего могут стать нейронные изменения в способностях человека осуществлять мозговую деятельность в отношении морально-нравственной оценки происходящих событий, что, в свою очередь, может сказаться на дальнейшем развитии политической и правовой систем общества — это может привести к изменению трактовок таких основополагающих концепций, как демократия, содержание прав

и свобод человека, равенство, справедливость, религиозные воззрения и др.

Одну из первых попыток сформулировать подобные проблемы и предложить способы их разрешения предложил американский писатель-фантаст **А. Азимов** еще в 1942 г. Свои ставшие знаменитыми три закона робототехники он выдвинул сначала в рассказе «Хоровод», а потом, в 1950 г., вокруг них строились практически все сюжеты сборника «Я, робот». Позднее он дополнил эти законы и так называемым «нулевым законом»: Робот не может причинять вред человечеству или своим бездействием допустить, чтобы человечеству был причинен вред. По мнению фантаста, эти законы должны быть заложены изначально в программу искусственного интеллекта как аналог нравственных ценностей, «категорического императива» для индивида.

Особой проблемой, как уже отмечалось, является нравственная характеристика процесса использования искусственного интеллекта индивидом при принятии важных решений, которые могут касаться интересов и, главное, безопасности людей. В книге *"Computer Power and Human Reason: From Judgement to Calculation"* (1970) **Дж. Вейценбаум** акцентирует внимание даже на простой возможности принятия решения человеком только на основании предложенной искусственным интеллектом (интеллектуальной системой) концепции. Ведь часто то или иное решение человек принимает на основе таких качеств, как рассудительность, чувство эмпатии и т. п. Как отмечает **Ник Бостром** в книге *«Искусственный интеллект: Этапы. Угрозы. Стратегии»* (2008), «рационалистическая суть AI несовместима с гуманностью и человечностью. Ведь основное свойство разума, проторазума, иного разума и т. д. — это прежде всего приспосабливать к себе окружающую среду. Поскольку у разума как такового нет

нравственности, он может “пойти” на все, что считает целесообразным»[4].

Таким образом, перед программистами на современном этапе развития цифрового общества стоит поистине неразрешимая задача — создать искусственный интеллект, способный «осознавать» моральные предписания и прежде всего такие догмы, как: достоинство, честь, трактовку прав и свобод личности в зависимости от особенностей ментальности разных народов, религиозные, морально-нравственные нормы, культурный код и др. В случае достижения такой задачи искусственный интеллект в полной мере станет субъектом, сопоставимым с физическим лицом. В 2001 г. научный сотрудник Института Сингулярности по созданию искусственного интеллекта Э. Юдковский, развивая этот тезис, ввел понятие «дружественного искусственного интеллекта», в соответствии с которым тот не должен совершать действия, способные любым способом навредить человеку, то есть задачей становится обучение его чувству сопереживания, сочувствия — чувству личностной эмпатии. Эта проблема стала предметом обсуждения в мае 2017 г. на закрытом заседании Валдайского клуба.

Как мы видим, главной проблемой современного этапа развития искусственного интеллекта состоит в создании «чувственного», дружественного отношения к человеку. И здесь возникает следующая проблема: существует ли различие искусственного интеллекта и «естественного интеллекта», признаваемого главным качеством человека, при этом, как отмечалось выше, связанного

[4] Н. Бостром — английский философ шведского происхождения. Правильно его фамилия транскрибируется по-русски как Бустрём, но мы будем придерживаться уже сложившейся в русскоязычной литературе традиции.

с мышлением. В зарубежной научной литературе по искусственному интеллекту различаются его характеристики как «слабого», предназначенного только для выполнения только определенных задач, и «сильного» (общего) искусственного интеллекта, особой технологии, способной к решению широкого числа задач по аналогии с человеческим мозгом.

Особым направлением развития искусственного интеллекта служит онтологический инжиниринг, в результате которого происходит, по словам авторов публикации *«Компьютерная онтология: задачи и методология построения»* (2014), следующее: «реализация технологий представления и обработки знаний в процессах решения задач системной интеграции знаний предполагает учет различных формально-методологических требований, критериев и оценок».

Как отмечает О.Э. Петруня в статье *«Искусственный интеллект сквозь призму димензиональной онтологии»* (2017), «если рассматривать AI всего лишь как усиление функций человека (слабый AI), то перспектива выглядит и желанной, и выполнимой. Сильная версия AI фактически претендует на решение столь сложной проблемы с помощью простых "одномерных" средств — аппаратных и формально-языковых. Таким образом, мы имеем несоответствие задачи и методов ее решения. В то же время стремление к воспроизводству деятеля в известных проектах (андроидный робот, "аватар" и т. п.), а значит — воссоздание человеческой "трехмерности" (по Франклу) не обходится без конструирования в воображении желаемого результата». Поэтому *аватар, киберфизическая система, робот и иные объекты интеллектуальных систем не смогут быть соотносимы с человеком,* так как мышление — это исключительно характеристика естественного интеллекта, поэтому пока любые попытки

отделить мыслительный процесс от индивида не могут быть успешны так же, как и разделение характеристики «человечность» от человечества.

Уже сегодня человек, как биологический вид, используя виртуальную реальность, получает искусственное тело — аватар, — в которое в перспективе возможна станет пересадка естественного интеллекта (человеческого сознания) с помощью цифровых технологий или прямой пересадки мозга. Люди смогут «переселяться» в киберпространство, в котором оцифрованное сознание будет практически вечно существовать в индивидуальном виртуальном пространстве. «Живущие» в киберпространстве люди не будут нуждаться в социальных отношениях и их регуляторах, соответственно возникнут (и уже возникают) новые векторы коммуникации (например, социальные сети), требующие принципиально иного правового регулирования. В сфере производства материальных благ при помощи развития NBIC технологий будут созданы особые наномеханические устройства (молекулярные наномашины, в том числе наноассемблеры), представляющие собой дешевую рабочую силу и способные решать практически любые задачи, что, безусловно, затронет сферу трудового права и права социального обеспечения. В недалеком будущем общество будет состоять не только из людей, имеющих естественный интеллект, но и аватаров, зомби и др., а следовательно признак «человечности», осознанной морали как основание для сегодняшнего типа правового регулирования может исчезнуть безвозвратно в силу иного ценностного основания для определения искусственного интеллекта, основанного на признании несущественности сознания.

В статье *«Прогностическое моделирование онтологий искусственного интеллекта как основа для проектирования необходимых референтных изменений*

законодательства» (2018) ее автор **С.В. Мельников** отмечает, что будущее как раз за «онтологическими цифровыми моделями». Основываясь на работах современных российских исследователей проблем правового обеспечения искусственного интеллекта П.М. Морхата, И.В. Понкина и А.И. Редькиной, он поясняет, что «семантическое и онтологическое обогащение моделей может способствовать более оптимальному построению процессов хранения и доступа к данным, предоставляя средства для структурирования, сохранения и визуализации релевантной информации». В этом контексте большое значение приобретают предпринимаемые в различных странах попытки регулирования общественных отношений с технологическим элементом, признания за отдельными видами киберфизических систем определенного правового статуса. На сегодня уже известны **гиноид** и **гуманоид**, получившие удостоверения личности.

В современном правовом пространстве как основном объекте правовой онтологии особое место занимает пространство виртуальной реальности, которое, по мнению автора монографии «Онтология права: (критическое исследование юридического концепта действительности)», опубликованной в 2013 г., Г.А. Гаджиева так же, как и «новые технические возможности передачи информации», создает «ситуацию, близкую к глубокому кризису права». Причиной этого кризиса могут служить пока неразрешенные законодательным путем вопросы: легальное определение каждой из интеллектуальных систем; отнесения искусственного интеллекта, интеллектуальных систем и иных объектов информационных технологий к субъектам или объектам правовых отношений; содержания, вида и особенностей юридической ответственности за неверные решения или действия искусственного интеллекта; определения прав искусственного интеллекта,

особенностей трудовых и налоговых правоотношений при условии использования его работодателями и т. д.

Вопрос отнесения искусственного интеллекта, киберфизических систем, различного рода нейронных систем к объектам или субъектам права является первостепенным для их правового регулирования. В статье «Искусственный интеллект: от объекта к субъекту?» (2019) С.А. Соменков отмечает, что «на сегодняшний день с точки зрения гражданского права *система, оснащенная AI, — это вещь*. При этом законодательство не содержит каких-либо особенностей правового режима этих вещей и не ограничивает их оборот. Однако у этого объекта есть ряд особенностей, связанных с возможностью его автономного функционирования», поэтому подобные системы, по его мнению, следует рассматривать как источник повышенной опасности во всех «сферах, где ценой выхода AI из-под контроля может быть не только имущественный вред, но даже здоровье и жизнь человека». При этом необходимо страхование ответственности за причинение вреда (ст. 931 Гражданского кодекса Российской Федерации (ГК РФ)) в случае причинения вреда искусственным интеллектом третьим лицам.

Для России урегулирование нового типа общественных отношений, в которых будет участвовать технологический элемент в виде разного рода киберфизических систем, роботов, объектов робототехники, искусственного интеллекта, аватаров и др., представляет собой первостепенную задачу. Несмотря на отсутствие легального определения искусственного интеллекта в законодательстве Российской Федерации, сам термин широко используется, что актуализирует его закрепление в нормативных правовых актах. Вопрос определения содержания этих дефиниций в научной литературе является дискуссионным. В.В. Архипов и В.Б. Наумов в статье «Искусственный

интеллект и автономные устройства в контексте права: о разработке первого в России закона о робототехнике» (2017) полагают необходимым определять **робототехнику** как *«совокупность общественных отношений, предметом которых являются производство, распределение и, немного перефразируя классическое определение экономики, использование автоматизированных технических систем»*, поэтому законодательство, его регулирующее, должно носить комплексный характер. В силу данного определения они предлагают разделять две категории:

1) «робот», определяемый как «устройство, способное действовать, определять свои действия и оценивать их последствия на основе информации, поступающей из внешней среды, без полного контроля со стороны человека» и являющийся в силу данного обстоятельства объектом правоотношений;

2) «робот-агент», выступающий в роли квазисубъекта, наделенного специальной правосубъектностью, так как он «предназначен по решению собственника и в силу конструктивных особенностей для участия в гражданском обороте, обладающий обособленным имуществом и отвечающий им по своим обязательствам, обладающий правом от своего имени приобретать и осуществлять гражданские права и нести гражданские обязанности».

А.А. Васильев и Д. Шпонер в уже цитированной статье по этому поводу отмечают, что в первом случае искусственный интеллект понимается всего лишь как техническое средство с правовым режимом вещи, а во втором случае за ним признается статус электронного лица по аналогии с юридическим лицом «через использование приема правовой фикции». Однако такая трактовка не вполне отвечает самой сущности искусственного интеллекта, так как «квалификация искусственного интеллекта как объекта права не учитывает наличия некой

субъектности — способности к мышлению и принятию самостоятельных решений. Во втором случае поднимается более глубокий вопрос мировоззренческого порядка: искусственный интеллект — это личность, подобная человеку». В документе *«Модельная конвенция о робототехнике и искусственном интеллекте. Правила создания и использования роботов и искусственного интеллекта»*, разработанном А.В. Незнамовым и В.Б. Наумовым, отмечается, что «роботы могут выступать в гражданском обороте как самостоятельные лица, в том числе выступать собственниками других роботов, если это прямо установлено применимым законодательством». Однако авторы не определяют правовой статус таких «самостоятельных лиц».

Ряд авторов (в частности, О.А. Ястребов, 2018) полагают возможным определять искусственный интеллект как электронное лицо в виде «децентрализованных автономных организаций, «управляемых посредством так называемых умных контрактов (smart contracts)». А Т.Я. Хабриева и Н.Н. Черногор в статье «Право в эпоху цифровой реальности» (2018) полагают возможным определять роботов как «цифровых личностей». В.Ф. Ужов (2017) предлагает определять искусственный интеллект как «электронное лицо — носитель искусственного интеллекта (машина, робот, программа), обладающий разумом, аналогичным человеческому, способностью принимать осознанные и не основанные на заложенном создателем такой машины, робота, программы алгоритме решения и в силу этого наделенный определенными правами и обязанностями». По мнению О.А. Ястребова (2018), необходимо различать «электронное лицо» и «электронного индивида» — робота, «причем искусственный интеллект, носителями которого являются удовлетворяющие определенным критериям роботы, необходимо рассматривать как базовую составляющую электронного лица».

По мнению В.В. Котляровой и М.А. Шемякиной (2019), к числу прав **электронного лица** должны быть отнесены следующие.

1. Право на жизнь. Однако для электронного лица необходимо будет разработать специальные критерии ее гарантированности со стороны государства. Данное обстоятельство актуализировано тем, что по отношению к физическому лицу смерть в соответствие со ст. 2 «Правил определения момента смерти человека», в том числе критерии и процедура установления смерти человека, означает «момент смерти его мозга или его биологической смерти (необратимой гибели человека)». Но определить факт «биологической смерти» для искусственного интеллекта не представляется возможным. Поэтому фактом приобретения электронным лицом права на жизнь «должно быть приобретаемым в момент появления первой информации в хранилище данных искусственного интеллекта».

2. «Электронное лицо должно претендовать на защиту от различных модификаций и нарушений целостности информации в хранилище данных и программном коде, которые могут привести к деформации личности или уничтожению искусственного интеллекта.

3. Права на результат деятельности (творческой, научной) искусственного интеллекта должны принадлежать ему самому, а не его создателю.

4. Искусственный интеллект должен обладать правом на самоопределение, то есть самостоятельно осуществлять свое развитие как культурное, так и экономическое, а также самостоятельно определять сферу своей деятельности».

В.Ф. Ужов (2017) предлагает для электронного лица законодательно определить только **право на неприкосновенность** (изменение, модификация, форматирование

либо ликвидация носителя AI должны быть санкционированы соответствующей комиссией и (или) органом власти, в противном случае это должно рассматриваться «как преступление против электронной личности») и **право на авторство созданных электронным лицом объектов интеллектуальной собственности.** В общей теории права в содержание правосубъектности (правового статуса) включены не только права, но и обязанности, невыполнение которых влечет за собой наступление ответственности. Для их определения в отношении электронного лица необходимо разработать критерии определения его «сознания» и обстоятельства, исключающие ответственность.

Хотя в современном праве нормативно урегулированы субъекты права, обладающие правами, но, в силу признания их ограничено и полностью недееспособными, не несущие ответственности за совершаемые ими действиями. Кроме того, лица до 18 лет, имеющие статус детей, обладают правами, но не обязанностями. Цель наказания в соответствие с п. 2 ст. 43 Уголовного кодекса Российской Федерации (УК РФ) является восстановление социальной справедливости и исправление осужденного и предупреждение совершения новых преступлений. По отношению к электронному лицу подобные цели не могут быть реализованы, поэтому в случае наделения законодателем правосубъектностью электронного лица необходимо будет изменять по отношению к нему меры и содержание мер наказания в случае уголовно-правовой ответственности.

Другой точки зрения придерживается П.М. Морхарт, который полагает, что правосубъектность электронного лица не может быть сопоставима с правосубъектностью человека, так как юнит искусственного интеллекта: лишен интенциональности; не способен ощущать на себе

последствия своих действий (бездействия); не способен ни к каким чувствам, руководствоваться в своей деятельности морально-этическими догмами; не может быть привлечен к уголовно-правовой и административно-правовой ответственности. Поэтому электронное лицо не может отвечать самостоятельно за причинение им вреда, следовательно, в зависимости от действий электронного лица ответственность будет возложена либо на разработчика программы искусственного интеллекта (в случае, если ошибка совершена или решение принято неверно из-за некорректного программирования), либо на производителя (если будет доказано, что деяние совершено электронным лицом из-за ограничения возможностей правообладателей принимать меры предосторожности со стороны производителя), либо правообладателя во всех иных случаях. П.М. Мохарт отмечает, что особая проблема в отношении электронных лиц заключается в том, что сегодня не существует защиты от взлома или перепрограммирования киберфизических систем, нейросетей или иных видов интеллектуальных систем, поэтому установление причины неправомерных действий со стороны искусственного интеллекта крайне затруднительно.

Как мы видим, нет единства в определении искусственного интеллекта в научной литературе, анализируемые различные позиции могут стать основой для законодательного закрепления только при условии формирования единого теоретико-правового подхода к трактовке места и роли искусственного интеллекта в правовом пространстве. Остается пока неразрешенным и еще один вопрос. Возможно ли правовое регулирование искусственного интеллекта, киберфизических систем, нейронных сетей, роботов и объектов робототехники исключительно в формате национального права? И да, и нет. С одной

стороны, правовое пространство отдельных стран различно по своим содержательным признакам, а следовательно правовые традиции в зависимости от принадлежности национальных систем права к различным правовым системам различно трактуют особенности правового регулирования. С другой — их правовая регламентация и урегулирование возможно только совместными усилиями международного сообщества, в связи с чем значительно возрастает значение и роль коллективных начал, общих интересов и *социального* (подразумевая тут особую категорию всеобщего блага).

Ключевые понятия: искусственный интеллект (AI), общий искусственный интеллект (AGI), сверхинтеллект (super AI), киберфизическая система (CPS), нейронная сеть, умный город, онтология AI, электронное лицо.

Контрольные вопросы:

1. Что такое онтология? Как она связана со сферой искусственного интеллекта?
2. Что такое искусственный интеллект?
3. Как вы можете пояснить следующее высказывание: «Искусственный интеллект как часть “сквозных” технологий»?
4. Определите область применения искусственного интеллекта. Каковы особенности применения AI в различных сферах жизни общества?
5. Каковы виды и особенности искусственного интеллекта, существующего на современном этапе научно-технической революции?
6. Каковы пути создания искусственного интеллекта?
7. Какие определения AI существуют в современной зарубежной и российской научной литературе?

Практико-ориентированные задания

1. Постройте таблицу определений AI в правовом пространстве различных стран на основе современной научной юридической литературы, выявив положительные и отрицательные черты подобных определений с точки зрения программиста и автора алгоритма киберфизических систем.

2. Создайте схему нейронных связей в сфере образования/телемедицины/судопроизводства (по выбору).

Темы докладов и сообщений

1. Правовое регулирование глубинных нейронных сетей.
2. Искусственный интеллект в управлении персоналом.
3. Искусственный интеллект в системе образования.
4. Цифровые технологии в сфере образования.
5. Правовое регулирование роботов и объектов робототехники в сфере образования в зарубежных странах.
6. Цифровое образование в России.
7. Телемедицина: особенности правового регулирования

ГЛАВА 3.
ДВЕ СТОРОНЫ ОДНОЙ ЭТИЧЕСКОЙ ПРОБЛЕМЫ

В результате изучения материалов главы обучающийся должен

знать:

– основные положения этики как философской дисциплины через призму классических сочинений Аристотеля и современных философов;

уметь:

– интерпретировать в этических терминах проблемы технологического прогресса и построения цифровой цивилизации;

владеть навыками:

– критического анализа философско-этических источников разных исторических эпох, проводить различие между этимологическим происхождением терминов и их современным употреблением в научном контексте.

3.1. Этика, мораль, нравственность и машины

Начало научному изучению этики, как и многих других наук, дал Аристотель. Вероятно, он же изобрел и само слово, взяв для него греческий корень *ethos*, что означает в переводе «обычай, нрав». На латынь это же слово переводится как *mos, moris* — знаменитое восклицание Цицерона "O tempora, o mores!" («О времена, о нравы!») именно об этом. И Цицерон явно говорит здесь об этом с осуждением — осуждая и самого Катилину, и бездействие Сената. В другой не менее известной речи, направленной уже

против Верреса, он говорит: “Tempora mutantur, et nos mutamur in illis” («Времена меняются, и мы меняемся с ними»). *Nos,* обозначающее нас, тут вполне созвучно с *mos,* обозначающим наш нрав, и при изменении времен существенно не то, что у нас отрастает борода или седеют волосы, а меняются наши нравы и взгляды на добро и зло, появляются новые привычки.

Эти перемены, скорее всего, к худшему: “*Vir morum veterum*”, то есть «муж старых обычаев», — безусловная похвала, хотя на современный русский эту латинскую фразу можно было бы перевести и так: «Человек устаревших взглядов», — и тогда она приобретает столь же безусловно осуждающий характер. Когда наша языковая интуиция подсказывает нам связь этики с моралью, она безошибочна: по-латыни ее называли *pilosophiae pars moralis,* и тут мы можем заподозрить, что ухо латинянина улавливало связь слова *moralis* с *mora,* которое могло означать прилагательное «глупая» или существительное «остановка, пауза» или даже «препятствие». И в самом деле, в представлениях многих трудно себе представить более глупое препятствие, чем устаревшие нравственные ограничения, борьба с которыми ведется на протяжении по крайней мере последних полутора веков. Новая мораль, как утверждается, должна опираться прежде всего на разум и общественное согласие. Это согласие должно достигаться по поводу моральных норм.

«Норма» — это еще одно важное понятие моральной философии, или этики. Оно также произошло от латинского глагола *nosco* — «узнавать, познавать, проверять» — при его сужении до *normo* — «проверять по отвесу». В своем трактате «Об архитектуре» Витрувий использует слово *norma* для обозначения инструмента (наугольника), который строитель использует для проверки прямых углов (что, как мы хорошо знаем, очень

важно для любой постройки), или, как мы скажем теперь, перпендикулярности линий и поверхностей. Но латинское *perpendiculum* обозначало именно отвес и происходило от глагола *per-pendo*, где *pendo* обозначает именно «висеть», а слово *perpendiculator* относилось к строителю, пользующемуся отвесом для определения вертикальных линий и поверхностей. Таким образом, в латыни слова «норма» и «перпендикуляр» оказываются почти синонимами и в переносном значении оба использовались как «правило», «пример» или даже «образец» (“*Demosthenes norma oratoris*” — «Демосфен — образец для оратора»). В современном языке они снова, как видим, разлучились, и расхожая фраза «мне это перпендикулярно» выражает небрежение принятыми социальными нормами.

Вообще на протяжении этих последних полутора веков моральная философия все больше утрачивает свое практическое значение, оставаясь при этом весьма модной темой академических исследований. При решении повседневных жизненных проблем люди либо не вспоминают о моральных обязательствах и общественных нормах, либо делают это молча. О причинах такого положения вещей, как и об эволюции понятия морали на протяжении этого времени, подробно пишет **Р.Г. Апресян** в книге *«Идея морали и базовые нормативно-этические программы»* (1995). Здесь, наряду с эволюцией идеи и ее связи с нормами и идеалами, рассматриваются также конкретные нормативно-этические программы, такие как гедонизм, перфекционизм, альтруизм и утилитаризм.

Предполагается, что всякая моральная проблема ставит человека перед выбором: он может поступить каким-то одним из более или менее предопределенного набора способов. В простейшем случае, *когда выбирать надо между двумя взаимоисключающими решениями, проблема называется* **дилеммой**. Чаще всего этим словом

пользуются, когда два решения не только исключают друг друга, но и оба откровенно плохи. Рациональный выбор решения зависит от программы.

Всякая программа ориентирована на благо, представляющее собой одну из центральных категорий этики. Согласно Аристотелю, всякая человеческая деятельность, направленная на достижение какой-то цели, предполагает существование предмета более ценного, чем сама деятельность. Так, «цель врачебного искусства — здоровье, судостроительного — судно, стратегии — победа, экономии — богатство» (Этика 1094а). По Аристотелю категория блага не требует определения, так как нравственному человеку и без определения ясно, что способствует достижению блага, а что ему препятствует. Большинство современных моральных философов исходят из противоположной установки, хотя есть и такие, кто по-прежнему разделяет этот аристотелевский принцип. Но очень часто он смешивается с другим: понятие блага неотъемлемой частью входит в религиозное чувство человека, и поэтому только религиозный человек может быть по-настоящему нравственным. Вспомним, что примерно такую позицию занимал в свое время Ф.М. Достоевский. Сила этой позиции в том, что определения блага в этом случае не требуется: в религиозных текстах прописано, какие поступки нравственны, а какие нет. В случае сомнений можно спросить духовника. Однако само по себе это не означает отсутствие выбора: нравственная дилемма может встать и перед религиозным человеком, религиозное чувство которого ясно говорит ему, что как бы он ни поступил, он поступит плохо.

Религиозной морали, как правило, противопоставляется светская этика, и ее основания по необходимости рациональны. Но и в этом есть определенный логический подвох, так как разум всегда приводит рациональное

рассуждение к интуитивной или эмоциональной оценке — чтобы понять это, достаточно рассмотреть какой-нибудь из знаменитых примеров, предоставляемых нам историей математики: расхождение между **Дэвидом Гильбертом** и **Эмилем Борелем** по поводу аксиомы выбора в конце XIX в. сводилось лишь к тому, что если первому представлялось совершенно очевидным, что из бесконечного множества чисел можно совершенно произвольным образом выбрать какое-то одно, то второму столь же очевидной представлялась невозможность такой произвольности.

Технологический прогресс XX в., и в частности появление самообучающихся машин, открывает для этической проблематики принципиально новый угол зрения, совершенно новые задачи и, следовательно, новый этап в их обсуждении. Но появлению самообучающихся и даже просто программируемых машин предшествовало внедрение машин в производство. Строго говоря, машины знакомы людям с древности: *Deus ex machina* — бог, возникающий из машины, — классический способ развязки в древнегреческом театре. «Автоматы» — знаменитое сочинение александрийского ученого **Герона**, жившего в середине I в. н. э., посвящено разным занятным устройствам, которые могут развлекать граждан, например при посещении храмов. Традиционная российская прялка или молотилка для обработки сжатого хлеба — дело сугубо внутрисемейное: их место или в доме, или вблизи него. Создание фабрик, куда женщины и дети (на первых порах) на целый день уходят из дома, чтобы работать в цеху, — это уже принципиально новая история. Как к ней отнестись? Она во благо или во зло? Без ответа на этот вопрос нам не удастся понять, для чего людям понадобились еще и такие станки, которые можно чему-то научить, а тем более такие, которые могут учиться сами.

3.2. Покидая мальтузианскую ловушку

В 1798 г. английский церковный деятель и экономист **Роберт Мальтус** (1766–1834) выпустил небольшую брошюру под заголовком *«Опыт о законе народонаселения» ("Essay on the Principle of Population")*, принесшую ему всемирную славу. Ее ключевая идея довольно проста: не сдерживаемое какими-то внешними причинами население растет экспоненциально («в геометрической пропорции»). А производство не может расти пропорционально населению из-за ограниченности природных ресурсов: условно говоря, урожай, собираемый крестьянами, не может расти пропорционально количеству крестьян, поскольку ограничена площадь обрабатываемого ими поля. Производимый ими продукт может расти только линейно по времени («в арифметической прогрессии»). В результате прямая и экспонента должны рано или поздно пересечься, производимого продукта на всех не хватит («возникнет перенаселение»), и население должно будет сокращаться от голода, от болезней или оттого, что борьба за ресурсы приведет к войне.

Идеи Мальтуса часто и много критиковали. И в самом деле, в своем первоначальном виде рассуждения выглядят наивно. Однако со временем они приобрели гораздо более внятную формулировку, известную как неомальтузианство. В ней уже не делаются конкретные предположения о тех или иных точных видах зависимости населения и производимого им продукта от времени, хотя сами исходные посылки остаются примерно теми же. (1) Рост благосостояния (как его показатель обычно берется совокупный годовой доход на душу) приводит к росту населения. Сейчас может показаться, что это и не так, но для XVII–XVIII вв. это надежно установленный демографический факт. (2) Рост населения приводит

и к росту совокупного годового дохода, но до тех пор, пока основным ресурсом в производстве богатства остается земля, рост совокупного годового дохода не может быть пропорционален населению по причинам, указанным выше. Поэтому (3) с ростом численности население беднеет. (4) Падение благосостояния приводит к сокращению населения.

Сокращение населения совсем не обязательно происходит в результате войн, эпидемий или голодомора — снижение уровня жизни прежде всего отражается на детской смертности. Для аграрных стран такая закономерность наблюдается и поныне, а в доиндустриальную эпоху она и вовсе была универсальной. Вдобавок сокращение достатка приводило к более поздним бракам, поскольку не хватало денег на приданое (или калым, в зависимости от конкретных традиций), и к сокращению детородного периода у женщин в силу ухудшения бытовых условий (менее здоровое питание, большая скученность внутри жилищ, реже меняющаяся одежда). Возникает своего рода отрицательная обратная связь, превращающая плотность населения в данной местности в функцию плодородия почв и традиционных стандартов жизни. Каждое следующее поколение практически точно сменяет предыдущее, а какое-то существенно заметное изменение плодородия почв, которое может произойти в результате смены климата или внедрения какой-то новой технологии (переход от двухдольной к трехдольной системе земледелия или замена волов лошадьми), скорее отразится на численности населения, чем на уровне его жизни. Именно такое состояние и принято называть *мальтузианской ловушкой*. В самых разных местностях на протяжении тысячелетий люди жили в одном и том же состоянии, обеспечивающем их предельный консерватизм в стиле жизни и нравах.

Парадоксально, что появление «Опыта» Мальтуса пришлось точно на то время, когда Европа, или по крайней мере Англия, уже начала выбираться из мальтузианской ловушки. Происходило это оттого, что наиболее ценным ресурсом, обеспечивающим наибольшую производительность, а следовательно и доходность, начали становиться машины. В отличие от земли, этот ресурс даже при экспоненциальном росте населения может оставаться пропорциональным населению. На протяжении XIX в. мальтузианскую ловушку покинули как минимум половина стран нашей планеты и происходило это исключительно благодаря промышленной революции, сделавшей инвестиции в машины более доходными, чем инвестиции в земельные угодья.

Смысл промышленной революции заключался не столько в усовершенствовании и (или) изобретении самих станков, сколько в выносе их за пределы домохозяйств. Это сделало возможным появление двух новых категорий — **массового обслуживания** и **массового производства.** Машиной нового типа становился завод сам по себе, и экономические законы выталкивали машины нового типа в такую производственную сферу, где ее обслуживали максимальное количество работников и сама она производила максимальное количество одинаковых предметов.

По понятным причинам среди первых одинаковых предметов, производившихся в подобных условиях, стала хлопчатобумажная ткань. Хлопок выращивается на полях, то есть при росте населения для поддержания уровня жизни требуются не только новые машины, но и расширение сельскохозяйственных угодий. Растущее применение машин уже ослабляет мальтузианские ограничения на рост населения, но пока машины обрабатывают продукты сельскохозяйственного труда, снять их полностью невозможно.

Перескакивая в этом процессе через несколько ступенек, мы приходим к новой технологической ситуации, сложившейся к середине XX в.: машины используются не столько для того, чтобы производить вещи, сколько для того, чтобы производить данные. Эта новая ситуация получила название **постиндустриальной**[5]. В структуре ВВП большинства развитых стран реклама, медиа-индустрия, кинематограф, издательский бизнес занимают бульшее место, чем сельское хозяйство или металлургия.

Эта ситуация оказалась новой не только с экономической, но также и с этической, и с правовой точек зрения. Рассмотрим такой пример.

В 1770 г. **Иван Николаевич Новиков**, один из крупнейших российских просветителей XVIII в., поместил 23 февраля 1770 г. в издаваемом им журнале «Трутень» небольшое, на 5 страниц, сочинение *«Чензыя китайского философа совет, данный государю»*. В конце сочинения вместо подписи: «Перевел с китайского не знаю кто». Выяснить имя автора перевода не представляло труда: им был Алексей Леонтьевич Леонтьев. В 1772 г. он издал целую книгу подобных «советов», поясняя во введении, что в его книге содержатся «нравоучительные одного маньчжурского хана поучения и многих ученых китайцев советы». Там опубликованное Новиковым сочинение помещено на с. 148–153 и озаглавлено так: «Рассуждение учителя Чензыя о правлении государственном». В маргиналии к заголовку дается пояснение: «Был в 11 веке по р. Хрс. Написал сие рассуждение в собственной своей книге».

[5] Термин «постиндустриальное общество» прочно вошел в обиход благодаря талантливой книге **Дэниела Белла** *«Переход к постиндустриальному обществу»* (*Bell, Daniel*. The Coming of Post-Industrial Society).

Для Новикова совершенно неважно ни имя автора, ни обстоятельства появления использованного им текста. Ему было важно только одно: как с его помощью можно эффективно осмеивать и выставлять в «сатирическом свете» императрицу Екатерину II. Можно считать, что цели своей он добился, так как в апреле того же 1770 г. его журнал был закрыт.

Подобное отношение к продукту чужого интеллектуального труда, совершенно обычное в XVIII в., абсолютно неприемлемо в наше время ни с юридической, ни с этической точки зрения. Основания тому вполне экономические: с точки зрения современников Новикова, сам по себе интеллектуальный труд Леонтьева не имеет никакой ценности. Ценностью обладает лишь политическая цель — «умерение» абсолютной монархии, приведение ее в конституционные рамки. Все теперь воспринимается иначе: как бы ни была важна сиюминутная политическая цель, ее ценность значительно ниже, чем у созданного Леонтьевым сочинения, даже несмотря на то, что это перевод. Более того, само по себе оригинальное китайское сочинение философа XI–XII вв. Чэн И (程颐, 1033–1107) дожило до XVIII в. и, очевидно, достаточно ценилось, чтобы внимание студента Русской духовной миссии в Пекине было ею привлечено. Кроме того, в России XVIII в. было от силы два человека, способных понимать китайский философский текст. Такой перевод может быть использован для разных целей и довольно долгое время.

Рассмотрим еще один пример. Практически при каждой встрече с медицинской или социальной службой нам приходится подписывать отдельную бумагу «Согласие на обработку персональных данных». В соответствии с действующим законодательством, собирать эти данные, а следовательно и просить подписать такое «согласие», может только внесенный в реестр оператор персональных

данных, при этом его возможности эти данные использовать сильно ограничены. Закон о персональных данных в России был принят только в 2006 г. — относительно недавно. И дело тут в том, что лишь относительно недавно данные стали обладать существенной рыночной ценностью. Анализ данных позволяет выявлять статистически значимые корреляции, которые могут использоваться, например, для персонализации рекламы, если ограничиваться наиболее безобидными примерами. Но современные методы обработки личных данных позволяют работодателю определять наименее лояльных работников или оценивать вероятность забеременеть в ближайшие три месяца для своих сотрудниц. Совершенно очевидно, что подобное использование личных данных весьма сомнительно с этической точки зрения, но при этом очень трудно выводимо за правовые рамки.

3.3. Этические нормы и научно-технический прогресс

Как уже говорилось выше, моральная философия, или этика, оценивает человеческие поступки или события в их отношении к таким категориям, как «добро», «справедливость» или «благо». Однако ни одна из этих категорий не абсолютна. Благо может пониматься в терминах спасения души (спиритуалистическом ключе), приятной беззаботной жизни (гедонистическом ключе), общей пользы (утилитарном ключе), пользы для определенного класса (классовом ключе), пользы для себя лично (эгоистическом ключе). В зависимости от того, как расставляются акценты в трактовке этических категорий, этические системы могут довольно сильно расходиться в оценках одних и тех же событий или поступков.

На открытии международного форума «Открытые инновации» в Сколтехе осенью 2019 г. французский философ Алексей Гринбаум рассказывал такую притчу: «К премьер-министру одной страны приходит инноватор и говорит: слушайте, есть для вас фантастическая инновация. ВВП удвоится, ваша страна преобразится, все будет не так, как раньше. Но, конечно, я попрошу кое-что взамен. Совсем немного. Десяток тысяч жизней ваших граждан в год. Премьер-министр в ужасе на него смотрит, а потом зовет охрану и говорит: "Выведите этого сумасшедшего". В этот момент он отказывается от автомобиля».

В самом деле, трудно себе представить техническое новшество, которое не стоило бы человечеству многочисленных жертв. **А.Н. Кочетов** в книге *«Рыцари Х-лучей»* описывает, в какое невероятное количество жертв обошлось человечеству открытие **Вильгельма Рёнтгена** (1845–1923), а **Билли Брайсон** в книге *«Беспокойное лето 1927»* дает не менее впечатляющую картину из истории авиации. Но если с рентгеновскими аппаратами люди научились обращаться более или менее безопасно, авиационный транспорт, хотя и уступает значительно автомобильному по собираемой ежегодно кровавой жатве, остается весьма опасным и вызывает у людей вполне обоснованную тревогу. Ради чего эти жертвы?

На этот вопрос можно пытаться ответить по-разному. Например, можно сказать, что технический прогресс не остановим, а жертвы на этом пути неизбежны. Такой подход одновременно и наивный, и безответственный. В самом деле, отправляясь в далекое путешествие, отважные мореплаватели не могли знать, кто из них сможет вернуться назад. Но они рисковали только собственными жизнями. Первые изобретения были чреваты несчастными случаями, но накапливаемый горький опыт должен

был учить и учил людей заранее предвидеть появляющиеся с новыми технологиями риски и пытаться их минимизировать. Эти люди жертвовали собой ради общего блага. Можно ли оправдать риск ради общего блага жизнью другого?

К ответу на тот же вопрос можно подойти и с другой стороны. Технологический прогресс чреват жертвами, но в целом делает жизнь людей безопаснее. Если посмотреть на демографическую статистику до промышленной революции, то выяснится, что наиболее опасный возраст в человеческой жизни — это были первые 12 лет. Ожидаемое время дожития новорожденного младенца (то есть средняя продолжительность жизни) было меньше ожидаемого времени дожития 12-летнего подростка. Среднестатистическая женщина, практически независимо от того, в какой она жила части света, рожала с частотой, допустимой физиологией ее организма, — чуть реже, чем раз в год. При этом условия мальтузианской ловушки требовали, чтобы каждое следующее поколение в точности замещалось предыдущим. Это значит, уходящая в мир иной супружеская пара должна оставить после себя ровно двух потомков. Все остальные потомки не достигают репродуктивного возраста, не доживая по преимуществу до 12 лет. Ничто подобное сейчас немыслимо. Только за XX в. детская смертность сократилась вдвое. Демографические перекосы и проблема бедности остаются, но в целом человеческое существование стало значительно безопаснее — и это следствие технических инноваций. К этому можно добавить, что и содержание этой жизни сильно изменилось: если в доиндустриальные времена подавляющее большинство населения должно было практически непрерывно, с раннего утра и до поздней ночи, выполнять нудную и тяжелую работу, то теперь, согласно оценкам американского экономиста **Джеффри Сакса**, средняя

продолжительность рабочего дня взрослого американца 3 часа 10 минут. Тут, конечно, требуется уточнить, что работающий американец, если уж он пришел на работу, то проводит там в среднем 7 часов 34 минуты, но на работу приходит только 42 % взрослых американцев — все остальные в это время в отпуске, на пенсии, ухаживают за детьми или престарелыми родителями, ищут работу или сидят в тюрьме.

Но риски на каждом новом витке технологического прогресса становятся все выше. Нобелевский лауреат по биологии, один из первооткрывателей двойной спирали ДНК, Фрэнсис Крик писал в своей книге «Жизнь как она есть, ее зарождение и сущность», что у человечества есть как минимум четыре причины не дожить до конца XXI столетия, причем только одна из этих четырех существовала в XIX — глобальная космическая катастрофа. Столкновение с достаточно большим космическим телом может произойти в любой момент и не раз происходило в прошлом. Последствия такого столкновения могут обернуться истреблением всего живого, *омницидом,* если, например, существенно изменится климатический режим или Земля лишится своей атмосферы. Нечто подобное, возможно, произошло в прошлом с Марсом и (или) Венерой.

Но есть и три другие причины. Например, глобальный военный конфликт или глобальная экологическая катастрофа. Четвертая причина, упомянутая Криком, не так очевидна, хотя и наиболее вероятна: постепенная деградация среды обитания в результате локальных военных конфликтов и локальных экологических кризисов сделает ее необратимо непригодной для жизни людей. И хотя неразумная экономическая деятельность людей даже в доисторическую эпоху приводила иногда к масштабным экологическим кризисам, сделавшим

непригодными для жизни целые регионы, до открытия цепных ядерных реакций омницид был практически невозможен. Уроки Чернобыля и Фукусимы показывают, что людям свойственно недооценивать опасности используемых ими технологий. Сейчас нередко высказываются опасения, что искусственный интеллект может обернуться еще большими рисками, чем даже атомная энергетика или атомное оружие. Эти риски можно предвидеть заранее, а располагаемые людьми технологии, не прекращающие своего быстрого развития, дают возможность уже сейчас достаточно точно просчитывать сценарии, по которым могут развиваться нештатные ситуации.

Как мы уже знаем, даже в силу самого своего определения (точнее, некоторых из определений), искусственный интеллект может подразумевать принципиально разные технологические типы, и в зависимости от этого возникают разные категории этических проблем. Любая технология искусственного интеллекта предполагает, во-первых, наличие цели и, во-вторых, самостоятельное поведение. Такую пару проще всего рассмотреть на примере оружия, и такой пример оказывается весьма важным в контексте этической проблематики. К тому же и в историческом плане теория целеустремленных систем возникла и развивалась на своем начальном этапе именно как продолжение теории автоматического управления движением реактивных летальных аппаратов. Попросту говоря, ракет.

Зенитная ракета может идентифицировать цель по тепловому излучению двигателя самолета. Координаты этого самолета ракета может получать и с земли, с локаторов станции наведения ракет, в данном случае важно не это, важно то, что дальше команды, меняющие ее курс, вырабатываются на ней самой в зависимости от того, как изменяются координаты самолета. Наблюдателю будет

казаться, что ракета себя определенным образом ведет, стремясь достичь определенной цели — подорваться на минимальном расстоянии от самолета.

Этическая проблема тут может быть связана с выбором цели или с поведением при достижении ее. Проблема, связанная с выбором цели, обостряется в том случае, когда не только поведение ракеты, но и, собственно, сам выбор производится в автоматическом режиме. Но даже присутствие человека в цепи управления не гарантирует, что ошибки не будет.

3 июля 1988 г. в самый разгар ирано-иракской войны в территориальных водах Ирана нес боевое дежурство американский ракетный крейсер «Венсен», снабженный многофункциональной боевой информационно-управляющей системой «Иджис». В состав системы «Иджис» на крейсере «Венсен» входил также комплекс «Фаланкс» — система оружия ближнего боя для того, чтобы в автоматическом режиме обнаруживать, отслеживать и уничтожать приближающиеся к кораблю противокорабельные ракеты и самолеты противника. Капитан корабля Уильям Роджерс-третий получил от «Иджиса» сообщение, что корабль атакован иранским истребителем F-14, и позволил «Фаланксу» произвести выстрел. В результате был сбит пассажирский самолет авиакомпании Иран Эйр, выполнявший рейс Тегеран (Бендер-Аббас) — Дубай. 274 пассажира и 16 членов экипажа погибли. Разразился колоссальный международный скандал.

В ходе расследования выяснилось, что «Иджис» сначала получил информацию о самолете и его местонахождении: Бендер-Аббас тогда использовался и как военный аэродром, на котором базировались F-14. И в этот момент на авиалайнере еще не включили транспондер. После этого «Иджис» запросил информацию о траектории самолета и получил ответ «снижение», что обычно бывает при

атаках. Проблема, однако, в том, что ответ относился уже к совсем другому самолету — американскому разведчику, находившемуся на боевом дежурстве над акваторией Оманского залива. Ошибка автоматизированной системы привела к неправильному выбору цели.

Понятно, что в данном случае не имеет существенного значения степень автоматизированности всей системы. Более того, даже наличие человека в цепи управления мало что меняло: капитан Роджерс лишь санкционировал решение, принятое автоматом.

Роковым в данном случае оказалось смешение качеств двух целей и их ложная атрибуция: качества двух целей оказались приписаны только одной из них. Теоретически похожую ошибку можно представить себе в том случае, когда второй самолет попадает в поле зрения самой ракеты: ее система наведения видит два активных источника инфракрасного излучения и может сбиться, захватив ложную цель. Очевидно, нечто подобное произошло во время первого зарегистрированного несчастного случая с промышленным роботом на заводе Форда в городе Флэт-Рок 25 января 1979 г. Рабочий Роберт Уильямс отправился на склад запчастей, поскольку его коллегам показалось, что робот, работающий на складе, вышел из строя и перестал поставлять оттуда запчасти. Но робот продолжал работать, и Роберт Уильямс, едва попав в поле зрения робота, был раздавлен, поскольку робот принял его за нестандартную деталь и решил отложить в сторону.

Таким образом, мы видим, что целеустремленная система в достижении целей может совершать ошибки двух видов: во-первых, может выбираться ошибочная цель, а во-вторых, при достижении даже правильной цели могут совершаться неприемлемые действия. Так, в случае с Робертом Уильямсом мы можем считать, что робот выбрал

правильную цель: Роберт Уильямс был неподходящей деталью, и его следовало удалить со склада. Однако сопутствующее летальное его повреждение сделало достижение этой правильной цели этически неприемлемым. К системе самонаведения ракет термин «искусственный интеллект» не применялся никогда. Программируемого робота на заводе Форда в духе GOFAI или систему «Иджис» на американском крейсере считали интеллектуальными в момент создания, но сейчас таковыми уже не считают. Чтобы убедиться, что нарастание «ума» может приводить и к нарастанию угроз, учеными были разработаны немного странные и даже в некотором смысле анекдотические сценарии, но они тем не менее показывают, от каких бед человечеству предстоит защищаться.

Приведенный выше пример с самонаводящейся ракетой дает нам возможность пролить некоторый свет еще на один технический термин, просочившийся в повседневную жизнь, — «в реальном времени». Станция наведения ракет выполняет две функции: она следит за ракетой и одновременно рассчитывает ее будущую траекторию по наблюдаемым параметрам траектории в предшествующие моменты времени. Если выясняется, что расчетное местоположение ракеты в определенный момент времени отличается от того, где ракета в этот момент времени действительно оказалась, значит параметры расчета надо скорректировать. Но это возможно сделать только в том случае, если местоположение ракеты удается вычислить до того, как соответствующий момент времени наступит. Для реальных расчетов это совсем не всегда верно: в приводившемся выше примере с расчетом распространения вспышки от атомного взрыва в атмосфере это совсем не так. Сама вспышка распространяется несколько микросекунд, а вычисление длится несколько месяцев. Когда вычисления требуют меньше времени, чем длится

вычисляемый физический процесс, говорят, что они идут «в реальном времени».

3.4. Этическая дилемма в эпоху слабого или сильного искусственного интеллекта

Сам по себе факт, что мы живем в компьютерной цивилизации, необычен и нов. Его было бы неправильно распространить на все человеческие сообщества. Он стал возможен благодаря одной характерной и пока еще мало изученной особенности технического прогресса во второй половине XX в.: компьютерная память и выполняемые компьютером вычисления дешевели в это время с невероятной и беспрецедентной в истории скоростью. Каждое вычисление жаккаровского станка сопровождается появлением на ткани стежка той или иной формы в соответствии с нулями и единицами перфокарты. В современном компьютере стоимость ячейки памяти и одной операции дешевле на тринадцать порядков. Чтобы как-то охватить умом колоссальность этого числа, можно, например, вспомнить, что примерно так соотносятся между собой длительность часа и все время существования нашей Вселенной.

Конечно, это удешевление не было бы возможно, если бы не исключительная миниатюризация вычислительной техники. Если размер одного отверстия в перфокарте Жаккара, служащего для записи одного бита, соответствовал примерно 10 миллиардам атомов, то для записи того же бита в современном компьютере достаточно десятка атомов. Что же касается выполнения самого вычисления, то тут какие-либо сравнения даже трудно проводить, так как в современном компьютере в дело вовлечена принципиально новая сущность — электромагнитное поле. Столь

быстрых и масштабных изменений не знает не только человеческая история, но и в истории природы им можно уподобить разве что эволюцию самой Вселенной в первые три секунды.

Тем не менее человеческий разум пока все еще остается единственным известным нам проявлением разума как природного явления, и только он открыт для изучения в этом качестве. Можно говорить, что некоторые животные, обитающие на Земле, также в той или иной степени обладают разумом — обезьяны, дельфины, собаки, свиньи, лошади и даже муравьи, но все они связаны с людьми единым эволюционным процессом и поэтому не сильно помогают в ответе на вопрос о возможном существовании разума хотя бы за рамками белковых форм жизни. Компьютерные симуляции разума, как говорилось выше, хотя вселяют оптимизм в некоторых исследователей данной проблемы, но все еще остаются симуляциями. И поле этических проблем, связанных с понятием искусственного интеллекта, оказывается поделенным на две части, требующие разных подходов.

Первый круг проблем обусловлен меняющимися коммуникациями между самими людьми: в их отношениях все чаще возникает принципиально новый посредник, порождаемый многочисленными вычислениями. Этот посредник может влиять на смысл коммуникации, непредсказуемо менять ее эмоциональную окраску, делать избыточными некоторые поступки или даже группы поступков, ранее считавшихся совершенно необходимыми. Эти проблемы стоят перед людьми уже сегодня и требуют того или иного немедленного разрешения. Нередко они решают спонтанно, что называется «по наитию», а следовательно и далеко не лучшим образом.

Второй круг затрагивает взаимоотношения людей с принципиально новым субъектом, подобного которому

история не знала и вряд ли скоро узнает. Возможное приобретение метафорой «искусственный интеллект» буквального значения подразумевает не только появление машин, способных к решению некоторых особых задач нового уровня сложности, но и возникновение у этих машин наряду с разумом способности осознавать себя, самостоятельно вырабатывать свое отношение к миру и людям, так или иначе отвечать за свои поступки — иначе говоря, приобретение этими машинами определенной субъектности. Нельзя сказать, чтобы проблемы этого круга требовали немедленного разрешения уже сегодня. Но их появление в будущем создает риски абсолютно нового качества, и, как говорилось выше, во избежание новых катастроф лучше продумать возможные последствия заранее. Кроме того, есть немалое количество экспертов, указывающих на очень быстрый, взрывной характер роста AGI после его появления. В таком случае у человечества будет слишком мало времени на решение такого рода проблем, когда они станут насущными, а решение их «по наитию» может привести к роковым ошибкам, в сравнении с которыми бомбардировка Хиросимы и Нагасаки или авария на Чернобыле покажутся детскими шалостями.

У теоретического обсуждения этого второго круга проблем есть еще одно прикладное значение. Предположению, что где-то за пределами Земли могут существовать иные и непохожие на нас разумные существа, как минимум около 500 лет. Было время, когда многие ученые считали весьма вероятным даже скорую встречу землян с инопланетной цивилизацией. На протяжении нескольких десятилетий XX в. внимательно и последовательно обсуждались все возможные признаки такой цивилизации как на самой нашей планете, так и в космосе.

Сейчас в этой области наступила эпоха пессимизма: подавляющее большинство специалистов, изучавших эту

проблему, убеждены, что, даже если внеземной разум где-то еще и существует (а вероятность этого оценивается по-прежнему как весьма высокая), он от нас слишком далек, чтобы можно было хоть как-то распознать его существование, не говоря уж о том, чтобы вступить с ним хоть в какое-то взаимодействие.

Между тем мы не можем исключить, что, как и в случае с машинным разумом, внеземной разум обнаружится слишком внезапно и слишком близко от нас, так что времени хорошо обдумать принимаемые решения не останется и придется полагаться на решения, принятые заранее.

3.5. Люди и скрепки

Чаще всего обывателя пугают теми сценариями будущих катастроф с участием искусственного интеллекта, где присутствуют вышедшие из-под контроля роботы. И мы о таких сценариях еще поговорим. Но сейчас мы рассмотрим историю попроще — самоуправляемый завод по производству скрепок. Этот пример был придуман Ником Бостромом еще в 2003 г. и за прошедшее время стал фактически хрестоматийным.

Представим себе, что некий предприниматель строит завод по производству канцелярских скрепок под управлением искусственного интеллекта, перед которым ставится единственная задача: производить наибольшее количество скрепок с наименьшими затратами.

Такой завод стремится превратить в скрепку любой подходящий атом, оказавшийся в поле его зрения. Постепенно он обнаруживает, что минимизировать затраты можно, перенеся производство в космос и колонизировав соседние планеты. Но перед этим он добудет весь подходящий металл на земле, в том числе и тот, который находится внутри человеческих тел. К тому моменту,

когда первая фабрика по производству скрепок появится на Луне, на Земле не останется не только людей, но и амеб.

Конечно, мы люди можем рассматривать подобный сценарий не как пример искусственного интеллекта, а как пример искусственного идиотизма, но не надо быть историком, чтобы понимать, как мало заметна граница между тем и другим. И как другой пример подобного искусственного идиотизма мы приведем здесь сюжет из книги **Ганса Моравеца** *«Дети разума»* (“Mind Children: The Future of Robot and Human Intelligence”) в пересказе Макса Тегмарка: *«Мы получаем от внеземной цивилизации радиопослание, содержащее компьютерную программу. Когда мы запускаем ее, выясняется, что это рекурсивно самосовершенствующийся искусственный интеллект, который быстро захватывает Землю... Он быстро превращает всю Солнечную систему в гигантскую стройплощадку, покрывая скалистые планеты и астероиды заводами, электростанциями и вычислительными центрами, которые он использует для проектирования и возведения дайсоновской сферы вокруг Солнца — вся солнечная энергия собирается для питания радиоантенн размером с Солнечную систему. Разумеется, вся эта деятельность приводит к полному истреблению людей, но последние из живущих умирают с проблеском надежды, что, какова бы ни была цель этого искусственного интеллекта, происходит что-то грандиозное, вроде как в StarTrek. Им и в голову не приходит, что вся эта грандиозная стройка затеяна с единственной целью — передать в космос то же самое сообщение, которое в самом начале истории получили люди. Все это не более чем космическая вариация компьютерного вируса. Точно так же, как фишинговая рассылка проводится в расчете на доверчивость интернет-пользователей, это сообщение рассчитано на доверчивость биологически развитых*

цивилизаций. Его создали миллиарды лет назад ради прикола, и, хотя его создатели вместе со всей своей цивилизацией давно уже вымерли, оно все продолжает путешествовать по космосу со скоростью света, превращая цветущие цивилизации в груду мертвых обломков».

В сущности, совершенно неважно, насколько такой искусственный интеллект разумен, насколько он способен к рефлексии и самостоятельному нахождению целей. Важно лишь то, насколько пространный производственный цикл он контролирует и в какой момент он может быть отключен. Последние два обстоятельства довольно тесно друг с другом связаны, так как чем более пространный производственный цикл оказывается ему подконтрольным, тем меньше времени будет у разумного субъекта на его отключение.

Модельность рассмотренных примеров проявляется также и в том, что присутствие людей в цикле управления никак не предполагается и в общем-то никакой особой выгоды люди здесь не получают. Конечно, скрепки до некоторой степени ценная вещь, но все же не до такой, когда можно подозревать запустивших такого рода производство людей в злонамеренных планах или даже в наличии у них материальных резонов скрывать опасность, когда она уже стала им понятной и они еще могут предпринимать какие-то действия.

Мы знаем из истории, что даже те незначительные возможности автоматизации, которые уже некоторое время у человечества есть, приводили его на край пусть и не глобальной катастрофы, но все же к ситуации в высшей степени чреватой.

Ключевые понятия: искусственный интеллект — слабый, сильный, универсальный; этика и нравственность; мальтузианская ловушка; неомальтузианство.

Контрольные вопросы

1. Каковы основные категории этики?
2. На что могут опираться моральные суждения?
3. Какие существуют морально-этические программы?
4. Что такое этическая дилемма?
5. В чем основное отличие неомальтузианства от теории Мальтуса?
6. Какие четыре причины могут не дать человечеству дожить до конца XXI в.?

Практико-ориентированные задания

Самостоятельно найдите информацию и проанализируйте национальные особенности перехода человеческой цивилизации от доиндустриального к индустриальному и от индустриального к постиндустриальному состоянию для двух-трех стран.

Темы сообщений, докладов и эссе

1. Искусственный интеллект и демографическая устойчивость.
2. Социальные риски при наличии сильного искусственного интеллекта.
3. Как сбежать из мальтузианской ловушки.
4. Пережить XXI в.

ГЛАВА 4.
ЧТО УМЕЕТ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ УЖЕ СЕЙЧАС

В результате изучения материалов главы обучающийся должен

знать:

– сферы применения технологий искусственного интеллекта в современной жизни;

уметь:

– пользоваться данными различных СМИ и профессиональной периодики для идентификации проблем технологического прогресса страны;

владеть навыками:

– обращения с современными компьютерными средствами, использующими технологии искусственного интеллекта, для анализа их продуктивности в различных областях применения.

4.1. Голосовой помощник

В последнее время довольно часто приходится слышать, что машинное обучение — это не искусственный интеллект. Вопрос, конечно же, исключительно в определениях. Компьютер, в памяти которого программными средствами организована искусственная нейронная сеть, нельзя назвать разумным существом, но он, безусловно, способен достигать очень сложных целей и очень точно имитировать разумное поведение. Рассмотрим несколько типичных примеров, где такая имитация уже стала рутиной, и сформулируем наиболее типичные этические сложности, которые здесь возможны.

Практически любая современная система массового обслуживания содержит в себе ту или иную разновидность виртуального помощника — чаще всего это голосовой помощник, включающийся при обращении по телефону. Звонит ли клиент в банк или хочет выяснить изменение тарифов за использование электричества, его встречает автомат, который производит первичный опрос и при необходимости переводит звонок на соответствующего специалиста. Это стало достижением последних пяти лет. При этом сама по себе проблема распознавания человеческой речи по понятным причинам была одной из первых, вставших перед разработчиками. Она приобрела ставшую едва ли не мемом аббревиатуру NLP (natural language processing) и оказалась настолько же сложной, насколько поначалу казалась простой.

Очень тесно связанной с ней изначально была проблема машинного перевода. Это другой вариант виртуального помощника, только тут участие искусственного интеллекта и машинного обучения менее заметны. Машине надо распознать сказанные или напечатанные в окне интерфейса слова, связать их друг с другом в осмысленную фразу, а затем передать содержащийся в ней смысл средствами другого языка. Насколько совершенен в этом стал, например, переводчик Google, убедиться совсем не сложно. Достаточно взять какую-то фразу и перевести ее, например, на китайский (возьмем для примера фразу «надо распознать сказанные слова, связать их друг с другом в осмысленную фразу, а затем передать содержащийся в ней смысл средствами другого языка»):

需要识别口语，用有意义的短语将它们连接起来，然后通过另一种语言传达其中包含的含义，

затем на суахили:

Unahitaji kutambua maneno yaliyosemwa, unganisha na misemo yenye maana, na kisha upeleke maana iliyomo ndani yao katika lugha nyingine,

потом на английский:

You need to identify the words spoken, connect them with meaningful phrases, and then transfer the meaning contained in them into another language,

и затем снова на русский:

Вам необходимо определить сказанные слова, связать их со значимыми фразами, а затем передать смысл, содержащийся в них, на другой язык.

Несмотря на то, что в цепочку включен такой сложный переход, как пара китайский — суахили, исходный смысл все еще можно идентифицировать, хотя появившиеся искажения довольно значимы: вместо неопределенно-личного предложении у нас появился конкретный субъект «вы» (то есть «мы»), а конкретно-терминологическое «распознать» заменилось расплывчато-общим «определить». Но если мы выкинем суахили из нашей цепочки, результат окажется лучше:

Необходимо распознавать произносимые слова, связывать их со значимыми фразами, а затем передавать смысл, содержащийся в них, на другом языке.

Голосовому помощнику не надо передавать смысл на другом языке, но ему надо преобразовать этот смысл в действия, что тоже можно назвать переводом. Смысл, теряемый при переводе, может показать нам, как смысл

может потеряться при превращении слов, сказанных голосовому помощнику, в действия. Невелика беда, если вместо расписания прилетов рейсов данной авиакомпании вы получите расписание вылетов, но ошибка при переводе средств с одного счета на другой может оказаться очень болезненной.

В технологическом плане отдельную и довольно сложную проблему приходится решать при идентификации устной, а не письменной речи, в этой проблеме мало такого, что имело бы прямое отношение к этической тематике. Гораздо более важная тема — это эмоциональная окраска произносимых текстов. На ранней стадии голосовые помощники просили от пользователей только односложных ответов: «да», «нет», «двадцать пять». Но сейчас способности многих из них значительно выросли: распознаваемый текст может быть и довольно сложным. А, как мы хорошо знаем, смысл сложного текста может иногда сильно зависеть от выражения, интонаций, с которыми он произносится.

Вслед за распознаванием эмоций, передаваемых интонациями, придет черед имитации их. Машинное обучение — совсем не тот путь, по которому есть шанс хоть когда-то прийти к имитации эмоций в прямом смысле этого слова, но имитация передающих эмоцию интонаций, вполне решаемая уже сегодня задача. Более того, сейчас ни Сири, ни Алиса, ни Алекса никак не визуализированы: для пользователя они существуют только как голос. Однако мы легко можем представить себе то время, когда для каждого из них можно будет выбрать и образ, появляющийся на мониторе компьютера или мобильного гаджета. У этого образа будут свои выражения и жесты. Кроме того, что они должны быть адекватны произносимым обоими участниками коммуникации словам, они должны быть социально приемлемыми. Что это в точности

означает, пока никто не понимает. Проблема существует, но по поводу ее решения мы пока можем только строить догадки.

Тут стоит вернуться к теме, поднятой в начале раздела. По остроумному замечанию французского социолога Антонио Казилли, искусственный интеллект голосового помощника нельзя считать интеллектом, поскольку он решает задачи, легко доступные самом человеку: все, что ему надо сделать, это выбрать ту или иную возможность из довольно ограниченного меню. Он не принимает самостоятельных решений и не способен предвидеть развития беседы. К тому же его нельзя считать и в полной мере искусственным, так как для обучения, необходимого для его работы, довольно большое количество людей должны заниматься разметкой данных. В разных компаниях эта задача решается по-разному, но принцип примерно один и тот же: через какую-то интернет-платформу нанимают случайных людей, которые выполняют эту работу за небольшую плату. Например, для голосового помощника Алекса наем добровольцев осуществляется через платформу Mechanical Turk (MTurk).

Происхождение этого названия восходит еще к 1769 г., когда австрийский изобретатель и иллюзионист Вольфганг фон Кемпелен изобрел первый в истории человечества шахматный автомат. Этот автомат представлял из себя небольшой комод, внутри которого помещался хитроумный механизм — перед началом партии его демонстрировали противнику автомата. Украшала комод восковая фигура турка, молчаливо наблюдающая за передвижением фигур на доске.

С этим автоматом успел сыграть даже Наполеон, и только в 1834 г. секрет автомата был раскрыт: большую часть внутренности комода скрывала хитроумная система зеркал, из-за которой бутафорский механизм внутри

казался значительно большим, чем был на самом деле. Ходы за «механического турка» делал прятавшийся позади механизма человек: на протяжении его почти 70-летней истории в этом качестве успели выступить несколько сильных шахматистов.

Выбор Амазоном такого названия для своей платформы в 2005 г. символичен, как и рекламное приглашение на ней зарегистрироваться — поработать «механическим турком». На первых порах зарегистрировавшимся работникам предлагали за скромное вознаграждение выполнить задание вполне конкретного (но неизвестного исполнителю) человека: транскрибировать аудиозапись, проверить текст на предмет орфографических ошибок, сделать несложный перевод. Сейчас основное количество заданий приходится на проверку правильности размеченных данных: правильно ли аудиосистема транскрибировала аудиозапись, справедливо ли были удалены непристойные или оскорбительные комментарии в соцсетях...

Аналогичные платформы применяют и другие компании, занимающиеся предоставлением сходных услуг. Функции «механического турка» для Яндекса выполняют работники на платформе Яндекс.Толока, созданной в 2014 г. Во всех этих случаях возникает принципиальная этическая проблема, к которой мы еще вернемся в дальнейшем: сам по себе сбор данных осуществляется в процессе бесед множества пользователей, хотя и с их согласия, но фактически без их ведома («для улучшения качества обслуживания разговор может быть записан»).

4.2. Беспилотный транспорт

Алиса была не единственной и даже не главной целью создания Толоки. Довольно большая часть заданий, которые получает зарегистрированный «работник»,

связана со сравнением и описанием картинок. Многие из них сделаны камерой или радаром, установленными на испытательном автомобиле, который просто разъезжает по городу. Эти картинки будут потом использоваться для тренировки нейронной сети беспилотного автомобиля.

Именно довольно большое количество собранных данных помогли некоторым компаниям в достаточной степени натренировать искусственные нейронные сети для того, чтобы обеспечить приемлемое качество автономного автомобиля. К числу наиболее успешных примеров реализаций этой идеи эксперты относят все модели автомобиля *Tesla* Илона Маска, включая наиболее разрекламированный *Cybertruck* (по состоянию на конец 2020 г. только "*Consumer Report*" отдавал предпочтение системе *Cadillac Super-Cruise* компании "*General Motors*"). Эти разработки не обходятся без осложнений.

16 сентября 2020 г. власти штата Аризона предъявили обвинение Рафаэле Васкес: когда автомобиль *Volvo XC90*, принадлежащего компании *Uber*, совершил 18 марта 2018 г. в городе Темпе наезд на Элейн Херцберг, именно она находилась за рулем. Правда, Рафаэла Васкес не управляла автомобилем: проходило испытание его систем в автономном режиме, но для Элейн Херцберг эти испытания, к которым она не имела никакого отношения, закончились гибелью.

Авария произошла после наступления темноты. Херцберг пересекала дорогу на своем велосипеде, а внедорожник Uber двигался со скоростью 38 миль в час. Данные, поступающие с лидаров автомобиля, позволяли бортовому компьютеру вычислить, что, учитывая постоянную скорость автомобиля, до наезда на объект оставалось шесть секунд, при условии его неподвижности. Но объекты на дорогах редко остаются неподвижными, поэтому

компьютер задействовал больше алгоритмов в базе данных узнаваемых механических и биологических объектов, ища соответствие, из которого можно было бы вывести вероятное поведение обнаруженного.

Сначала компьютер посчитал, что имеет дело с другой машиной, движущейся с достаточной скоростью, чтобы уехать до возникновения опасности наезда. Только в последнюю секунду была обнаружена четкая идентификация — женщина с велосипедом, хозяйственные сумки, беспорядочно свисающие с руля... Женщина, несомненно, предполагала, что приближающийся автомобиль просто объедет ее, что и случилось бы, если бы им управлял человек. Но компьютер не стал изменять направления движения, а передал управление оператору, Рафаэле Васкес, которая даже не успела этого заметить.

После происшествия власти штата запретили проводить тестирования на общих трассах, а *Uber* закрыл «беспилотный» эксперимент и уволил 300 операторов. Техническая ошибка не была доказана, хотя изначально предполагалась. Однако эта ситуация поставила членов технического сообщества перед двумя неудобными вопросами: была ли эта алгоритмическая трагедия неизбежной? И насколько мы будем готовы привыкнуть к подобным инцидентам?

Свой вариант автономного автомобиля есть и у Яндекса. Его разрабатывали с 2017 г. на основе автомобилей *Toyota Prius*. С ним тоже были некоторые неприятности: в 2019 г. ему случилось столкнуться с легковым автомобилем в Москве. В компании заявили, что виноват оператор транспортного средства, и инцидент замяли. С 2020 г. системы беспилотного управления стали ставить на автомобили компании *Hyundai* по согласованию с ней. Предполагается, что уже в 2021 г. такие автомобили будут

использоваться в Москве и некоторых других городах как такси.

Существующее законодательство запрещает подобное использование беспилотных автомобилей в США и многих других странах. Поэтому даже в рекламных роликах *Tesla* подчеркивается, что присутствие водителя за рулем «необходимо (только по) требованию закона». В России ситуация немного проще, поскольку действует «экспериментальный правовой режим», и в соответствии с Федеральным законом № 331 от 2 июля 2021 г. многие статьи или пункты статей закона «О безопасности дорожного движения» от 10 декабря 1995 г. дополнены фразой: «Действие требования, установленного настоящим пунктом [или статьей], может быть изменено или исключено в отношении участников экспериментального правового режима в сфере цифровых инноваций в соответствии с программой экспериментального правового режима в сфере цифровых инноваций, утверждаемой в соответствии с Федеральным законом от 31 июля 2020 г. № 258-ФЗ "Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации"», фактически их отменяя. В случае причинения беспилотным транспортным средством какого-либо вреда ответственность будет нести его владелец, но страховать эту ответственность на момент написания книги (август 2021 г.) российские страховщики были еще не готовы. Они указывали на недостаточность собранной статистики для оценки рисков.

С формальной точки зрения, такое экспериментальное решение, действующее не во всей стране, а только в некоторых особо выделенных ее частях, временно закрывает вопрос, но остается сомнительным с этической. Впрочем, не безусловно и решение, фактически принятое, например, в США, где ответственность лежит на водителе, даже

в том случае, если он не принимает никакого участия в управлении автомобилем, а только наблюдает за тем, что происходит на дороге и как автомобиль реагирует на происходящее. Мониторы *Tesla* позволяют водителю видеть не только то, что есть на дороге реально, но и то, как системы распознавания образов идентифицируют те или иные объекты: что они определяет как пешехода, что как велосипедиста (более уязвимое транспортное средство, движущееся с небольшой скоростью), что как равноправного участника движения (транспортное средство, способное двигаться с такой же скоростью). Однако, принимая во внимание характерную скорость изменения ситуации на дороге, сложно себе представить, как бы водитель успевал исправить ошибку, совершаемую автоматом. Возлагать ответственность на принимаемые автоматом решения на водителя — строго говоря, такая же формальность, как возложение ее на владельца автомобиля.

Собранная за всю историю автомобильного транспорта статистика говорит, что наиболее опасным автомобиль был в конце 1950-х гг. — именно тогда вероятность погибнуть в течение следующего часа для водителя автомобиля и его пассажиров достигла максимальных значений. С тех пор положение дел сильно изменилось: улучшились дороги — на них появились светофоры, более совершенная разметка, отбойники, ограждения, не только разделяющие полосы с противоположным направлением движения, но и закрывающие свет фар встречных автомобилей; улучшились сами автомобили — на них стали применять разнообразные системы безопасности, от надувных подушек, до складывающихся рулевых колонок и направляющих, по которым при столкновении уезжает под дно автомобиля мотор. Благодаря всем этим мерам уже к концу прошлого века для находящихся в автомобиле людей

вероятность погибнуть в течение следующего часа сократилась вдвое. Эксперты говорят, что переход к полностью беспилотному личному и общественному транспорту уже на нынешнем уровне развития технологии гарантирует снижение этой вероятности еще вдвое. Но отсюда совсем не следует, что она станет равной нулю.

Определять ответственность в случае дорожно-транспортного происшествия при участии только беспилотных автомобилей будет отнюдь не проще, чем сейчас, когда у каждого транспортного средства есть свой водитель. Избегать подобного поможет коммуникация таких автомобилей. В интервью В.В. Познеру, показанному по центральному телевидению 3 декабря 2018 г., президент группы компаний "*Cognitive Technologies*" Ольга Ускова говорила о замеченной ее сотрудниками на испытательных полигонах склонности разрабатываемых ими моделей к спонтанной коммуникации. Немногим менее двух лет спустя, в июне 2020 г., международная группа разработчиков из Калифорнийского университета в Беркли опубликовала результаты своих исследований по испытанию «модуля избегания столкновений» (Local NMPC Module) роботов, работающих в узких тоннелях. Суть работы такого модуля в том, что нейронные сети каждого из роботов учатся генерировать управляющие команды для каждого отдельного робота не только на основе данных с его внешних датчиков, но и на основе управляющих команд других находящихся поблизости роботов. Нетрудно видеть, что беспилотные автомобили, движущиеся в общем потоке по городским улицам, мало чем отличаются от роботов, работающих в узких тоннелях. Главное отличие заключается в том, что в исследованном группой ученых Калифорнийского университета случае все роботы принадлежат одному и тому же владельцу, поэтому передача информации от одного робота к другому не представляет

никакой ни этической, ни правовой проблемы. В случае же нахождения беспилотных автомобилей в общем потоке у каждого из автомобилей может быть свой владелец и совершенно независимые пассажиры, которые совершенно не намерены делиться информацией о своем движении с кем бы то ни было. И тем не менее организовать движение в среде, где нет транспортных средств, управляемых человеком, значительно проще. Вероятнее всего, это будет первым и последним этапом в развитии городского и междугородного транспорта: сначала для беспилотного транспорта будут выделены ряды, выезд на которые для водителей-людей будет запрещен, а закончится все тем, что управлять транспортом людям будет разрешено только в специально выделенных для этого местах.

Промежуточный этап, когда транспортные средства, управляемые людьми и управляемые машинами, будут встречаться в едином пространстве, окажется связан со множеством проблем как правового, так и этического характера. По мнению профессора Массачусетского технологического института и одного из пионеров вычислительных нейронаук **Томазо Поджо**, технологические и этические проблемы, возникающие при таком слиянии, удастся разрешить не ранее, чем через 20 лет. Но это мнение он высказывал в 2019 г., когда ни об экспериментальном правовом режиме, ни о «цифровых песочницах» на некоторых территориях России еще никому не было известно. Внедрение в городской транспортной среде беспилотных такси — смелый шаг в этом направлении, и он наверняка даст обильную пищу для исследователей самых разных направлений.

Автоматические транспортные перевозки в изолированных зонах, где невозможна встреча с пешеходами и с другими транспортными средствами, осуществляются уже довольно давно. Самым ярким примером стала

первая линия парижского метрополитена, соединяющая индустриальный район Ла Дефанс с замком Шато де Венсен , проходящая под Лувром и вдоль Елисейских полей. Решение о ее поэтапном переводе на автономный режим было принято в 2005 г. Первые беспилотные локомотивы прошли испытание зимой 2011 г., а к концу того же года стали использоваться в смешанном режиме, когда управление машиной контролировалось человеком, и только в 2012 г. беспилотные поезда стали выходить на линию без людей-машинистов по вечерам и в выходные дни. С 28 июля 2013 г. поездки по линии стали выполняться исключительно в автономном режиме.

Такой переход требовал исключительно тщательно продуманных мер безопасности. В частности, надо было полностью исключить возможность взаимодействия движущегося поезда с людьми вне его. Это было достигнуто построением плотно закрытого тоннеля, снабженного на перронах автоматическими дверьми, которые открывались только тогда, когда открывались двери стоящего на станции поезда. Таким образом, одновременно исключается возможность проникновения преступников или просто легкомысленных граждан в тоннель.

Весьма специфический, но очень важный для сельскохозяйственных стран вид транспорта — хлебоуборочные и другие подобные комбайны. Цитированный выше американский экономист Джефри Сакс оценивает долю населения, занятого в сельском хозяйстве современной Америки, менее чем в 1 %, а сравнивая характер труда сейчас и в доиндустриальную эпоху, он подчеркивает практически полное отсутствие у сельскохозяйственного рабочего необходимости ручного труда. Главное, чем ему приходится заниматься, — это сидеть в удобном кресле трактора и поворачивать нужные рычаги. Автоматизацией этой деятельности занимается также упоминавшаяся выше группа компаний “*Cognitive Technologies*”,

у которой уже сейчас есть готовые к внедрению модели хлебоуборочного комбайна, способные «видеть» убираемый хлеб и следовать сложной конфигурации поля. Главные проблемы, связанные с их эксплуатацией, опять же относятся к области этики: как должен «вести себя» комбайн, если в него заберутся пьяные колхозники? Что делать при встрече на поле с людьми, дикими животными или домашним скотом? Кто и как должен контролировать движение комбайнов в удаленном режиме?

По оценкам историков, одним из факторов, способствовавших промышленной революции, стал рост производительности сельскохозяйственного труда. Если в конце XVI в. в Англии в нем было занято более 80 % населения, то к концу XVIII в. их доля снизилась до 50 %. В XIX в. высвободившаяся часть населения переместилась в города и стала работать на фабриках и заводах. Что ждет высвобождающуюся часть деревенского населения теперь? И нет ли угрозы исчезновения деревень вообще в связи с дальнейшим развитием подобных технологий?

4.3. Автономная медицина

Едва ли не самым естественным выводом из сказанного выше по поводу машинного распознавания образов будет предположение, что машинный интеллект должен оказаться эффективным при диагностике по разного рода снимкам: рентгеновским, МРТ, УЗИ и т. п. И это действительно так. Подобные работы начались еще в прошлом веке, и в 2015 г. были зарегистрированы выдающиеся успехи. В 2015 г. группа нидерландских медиков сообщила в журнале “European Radiology” об исследовании, проводившемся с использованием компьютерной диагностики рака простаты с помощью МРТ. И ее использование значительно повысило достоверность и мощность

выявления заболевания по сравнению с исследованиями, где врач-радиолог полагался только на собственное зрение. Год спустя аналогичные исследования провели в Стэнфордском университете, и они тоже подтвердили, что компьютер с натренированной нейронной сетью, анализируя фотографии с микроскопа, более эффективно выявляет рак легких, чем врач-человек.

Примерно тогда же выяснилось, что машинное обучение позволяет довольно надежно устанавливать взаимосвязь между генетическими особенностями человека и характером его заболеваний, а также откликом его организма на те или иные медицинские препараты. Таким образом, возникла вполне обоснованная надежда на быстрое развитие персонализированных методов в медицине. Осенью 2020 г. профессор Массачусетского технологического института Регина Барзилай получила самую престижную в области исследований искусственного интеллекта премию, учрежденную Ассоциацией перспективных исследований искусственного интеллекта (The Association for the Advancement of Artificial Intelligence) и по денежному наполнению совпадающей с Нобелевской премией. Присуждение основывалось на многолетних исследованиях Барзилай по диагностике и лечению онкологических заболеваний молочных желез у женщин. Ей самой пришлось столкнуться с таким заболеванием в 2014 г. До этого она занималась проблемами распознавания человеческой речи и машинного перевода.

Технологии машинного распознавания образов, аналогичные тем, что использовались коллегами Барзилай в Нидерландах и Стэнфордском университете, позволили существенно сместить временные рамки в диагностике заболеваний. По ее оценкам, машина фиксирует будущую опухоль на четыре года раньше, чем ее заметит

врач-человек. А генетические особенности, проанализированные компьютером, помогут подобрать наименее травматичное и наиболее эффективное лечение. Но в данном случае была отмечена еще одна возможность ранней диагностики. Всякое медицинское обследование начинается с анамнеза — разговоров между врачом и пациентом. Как выяснилось, вероятность онкологического заболевания может быть оценена с довольной высокой точностью уже на основании таких расспросов и задолго до того, как будущую опухоль обнаружит машина. Но для соответствующего обучения нейронной сети нужны однотипные данные. И если рентгеновский снимок женской груди не зависит ни от национальности женщины, ни от языка, на котором она говорит со своим врачом, то анамнез, получаемый в разных странах, существенно различается лингвистически. Новшество, внедренное Барзилай, заключалось в том, что нейронная сеть тренировалась на основе рентгенологических обследований в их корреляции с анамнезами, приведенными к единому лингвистическому базису методами NLP и машинного перевода. Это позволило значительно расширить массив данных, использованных для машинного обучения.

Тем не менее хотя искусственный интеллект позволяет сделать диагностику более точной и более ранней, а лечение более эффективным, он допускает ошибки, пусть даже и реже, чем человек. К самым страшным последствиям в истории медицины по вине программного обеспечения привели ошибки в установке Therac-25, созданного канадской государственной организацией «Атомная энергия Канады» (Atomic Energy of Canada Limited) в конце 1970-х гг. С 1982 г. Therac-25 был запущен в серию, и на протяжении 1985–1987 гг. как минимум шесть пациентов в Канаде и США сильно пострадали от передозировок, причем двое из них скончались.

Прибор включал в себя ускоритель электронов и позволял облучать опухоль либо бета-лучами — то есть потоком электронов с энергиями от 5 до 25 электрон-вольт, либо рентгеновскими лучами, которые получались при установке на пути электронов вольфрамового рассеивателя. Кроме того, был предварительный режим, когда электроны в ускорителе не образовывались, а пучок имитировался видимым светом. Так как для генерации нужной интенсивности рентгеновских лучей на рассеивателе требовалась более высокая интенсивность пучка высокоэнергетических электронов, то если по каким-то причинам он оказывался не на месте, пациент получал бета-радиоактивный ожог. Как выяснилось впоследствии, при некоторых действиях обслуживающего персонала — например, при быстром переключении с рентгеновского режима на электронный, — в программе происходила ошибка: рассеиватель убирался, а пучок электронов оставался прежним. Сильный радиоактивный ожог оказывался гарантированным.

Сколько людей в мире гибнет из-за неправильного лечения, узнать не так-то просто. В открытых источниках нет никакой официальной статистики. В России о 70 тысячах осложнениях ежегодно по причине врачебных ошибок и неверного использования оборудования говорил министр здравоохранения Михаил Мурашко на заседании совета ректоров медицинских вузов России в начале 2020 г. Немного раньше, в 2017 г., двое исследователей департамента социологии Университета им. Пердью в Уэст-Лафайете (Индиана) опубликовали статью в журнале “Studies in Health Technology and Informatics”, в которой давали оценку в 251 тысячу смертей в год для США. Некоторые отечественные независимые эксперты, ссылаясь на это исследование и утверждая, что отставание России от США по этому показателю невозможно, дают

выглядящую слишком произвольной оценку в 300 тысяч смертей в год по России.

Следует, однако, указать на терминологическую тонкость: «врачебная ошибка» обычно считается русскоязычным эквивалентом английского термина “medical error”. Совершенно очевидно, что не только ошибка врача может повлечь за собой смерть пациента, находящегося в медицинском учреждении или находящегося на лечении. В рассмотренном выше случае с Therac-25 причиной радиоактивных ожогов и смерти пациентов были не назначения врача и не отсутствие должных навыков и знаний у медперсонала. Причиной стала неверная работа программного обеспечения. В данном случае речь не шла об искусственном интеллекте — тут была стандартная ошибка в стратегии обеспечения безопасности. Это позволяет относительно легко идентифицировать виноватых. Машинное обучение делает непрозрачным механизм принятия решения. Кто будет виноват, если человек пострадает в результате ошибочного решения, принятого машиной? Ответа на этот вопрос, как и на многие другие подобные, пока нет, но дискуссии уже ведутся. Скорее всего, в юридическую практику окажется внедренным не лучший вариант, и не потому что дискуссии будут закончены, а потому что обходиться дальше без закона — пусть даже и не самого совершенного — станет невозможно. Дискуссии будут продолжаться, и законы будут совершенствоваться.

В заключение этого раздела мы добавим, что в настоящее время успешно проводятся эксперименты по роботизации хирургических операций, в ходе которых натренированная нейронная сеть позволяет механизму накладывать швы значительно более аккуратно и точно, чем смог бы сделать человек с самыми «золотыми» на свете руками. Пока дело ограничивается зашиванием

кожицы на винограднине, но нет никаких сомнений, что пройдет совсем немного времени и такой робот станет удалять людям воспалившийся аппендикс и зашивать отслоившуюся сетчатку. Эти операции будут проводиться значительно более точно и с меньшей вероятностью осложнений, чем они проводятся сейчас, но все-таки в каких-то случаях осложнения будут возникать. Понятие «врачебной ошибки» окажется в этих случаях еще более двусмысленным, а юридическая практика еще более запутанной. Нахождение ответов на эти неразрешенные на сегодня этические проблемы в этой области окажет неоценимую помощь в ее нормализации.

4.4. Судопроизводство

Вернемся к цитируемому выше высказыванию председателя правления Сбербанка России Германа Оскаровича Грефа о том, что юристы, которые могут сейчас рассчитывать на работу в Сбербанке, должны владеть также аппаратом анализа данных. В частности, во время Sberbank Data Journey Day 2018 г. в кинотеатре «Октябрь» он говорил: «Юристы — очень важная профессия, но юрист без подготовки в области исследования данных сегодня для нас уже неинтересный агент. Мы генерируем 1 300 000 исковых заявлений в год. Это 10 % всей исковой системы Российской Федерации. У нас примерно 85 % этих исков генерится автоматически. Они прилетают в судебную систему. И там они вручную разбираются, и там принимаются ручные решения. Возможно ли существование такой диспропорции? Как только этот процесс будет автоматизирован в большом числе организаций, наша судебная система просто захлебнется».

Иначе говоря, в Сбербанке сейчас для создания исковых заявлений используется по преимуществу система

с технологией искусственного интеллекта, которая интерпретирует поток входных данных в юридические документы, следуя примерно тому же принципу, что и система, генерирующая подписи под фотографиями. Удельный вес этой технологии среди организаций, часто обращающихся в судебную систему, пока относительно невелик. С его ростом неизбежно внедрение интеллектуальных систем, выносящих судебные решения. Этот прогноз был сделан еще осенью 2018 г.

Каковы могут быть последствия? Прежде всего, автоматизация рассмотрения исков неизбежно приведет к ускорению судопроизводства. Сейчас судебные дела нередко показывают тенденцию к исключительному затягиванию, и быстродействие компьютера может оказаться принципиально важным. Не менее важно объективное рассмотрение сути дела и четкое следование букве и духу закона. Как ни странно, машина и здесь может с легкостью обойти человека, хотя многое зависит от того, как размечены данные, используемые в процессе обучения.

К концу 2018 г. обсуждение проблемы достигло уже общеевропейского уровня, и Совет Европы принял *«Европейскую этическую хартию об использовании искусственного интеллекта в судебных системах и окружающих их реалиях»*. Среди принципов, которые будут поддерживаться на общеевропейском, а значит и на общемировом уровне, — расширение общедоступных интеллектуальных онлайн-платформ, в частности для автоматизированного создания исковых заявлений. Одно из заранее безусловно просматриваемых последствий проведения такого намерения в жизнь — лавинообразный рост исковых заявлений, подаваемых в суды, которые окажутся неспособными функционировать, если только не внедрят системы искусственного интеллекта у себя.

Работа, ранее обычно выполнявшаяся помощником судьи, сейчас все чаще и чаще выполняется интеллектуальной системой с использованием нейронной сети. На протяжении последних трех лет идут разговоры о том, что вот-вот в какой-нибудь стране с высоким уровнем цифровизации, например в Эстонии, искусственному интеллекту будет дана возможность самостоятельно выносить решения по административным или гражданским делам с относительно невысокой исковой суммой — в случае Эстонии речь шла о 7000 евро. Тем не менее на момент написания книги о таком опыте никто не сообщал.

Задержки могут быть связаны с различными причинами. Большая часть населения даже самой передовой страны с опаской относится к широкому внедрению цифровых методов. Многие боятся «восстания роботов» или «цифрового концлагеря», который строят для нас мировые элиты. Об этом еще пойдет речь ниже. Определенную роль сыграли здесь исследования вроде того, что было процитировано выше, — в котором выяснилось, что алгоритм, предсказывающий вероятность рецидива отбывающих наказание по различным статьям уголовного кодекса, систематически завышает ее, если речь идет об афроамериканце. То есть нейронная сеть при обучении воспринимает не только ту информацию, которая содержится в подготовленных для ее обучения данных, но и ту, которую проводящий обучение персонал предпочел бы держать при себе. Наконец, людям хочется, чтобы к их делу отнеслись «по справедливости», то есть приняли во внимание моральные обстоятельства. Мысль о том, что моральность того или иного юридического решения можно рассчитать на вычислительной машине, многим покажется странной, а то и неприемлемой.

Но, как бы то ни было, людям еще придется решать, до какого рубежа они готовы предоставить право

вмешиваться в свою жизнь компьютерным алгоритмам. В Эстонии новорожденный, еще находясь в роддоме, получает место в детском саду и в школе. При перемещении в другой населенный пункт действующая авансом приписка изменится. Подобная практика делает излишними многие привычные для жителя России хлопоты, но легко себе представить, какие возражения она может встретить в той же России, где многие привыкли выбирать школу для ребенка в последние полгода перед началом учебы и совсем не обязательно рядом с домом. Хорошие отзывы об учителях, углубленные занятия по иностранному языку или какие-то другие соображения превратят предусмотрительность робота в назойливость.

Точно так же готовность предоставить обученной нейронной сети выносить приговоры, даже если их потом должен будет утвердить судья-человек, а тем более дать искусственному интеллекту право на окончательный приговор даже тогда, когда речь заходит о длительных сроках заключения, зависит от национальной культуры и может меняться от региона к региону.

Все сказанное выше относилось к уже существующим законам, однако мы хорошо знаем, что законы часто бывают далеки от идеала. Иногда они противоречат друг другу, иногда они даже могу противоречить сами себе, что в общем-то и неудивительно, если иной закон выглядит как солидный том на полтысячи страниц. В истории известны случаи, когда законодатели умышленно писали законы так, чтобы исполнение их было невозможно. Это открывало очень широкие возможности для контролирующих органов, фактически получавших право назначать преступника по своему выбору. В пьесе А.Н. Островского «Горячее сердце» городничий Градобоев спрашивает собравшихся у его крыльца людей: «Как же мне вас судить теперь? Если судить вас по законам... так законов у нас

много». И, конечно, собравшиеся дружно начинают кричать: «Суди по душе, будь отец, Серапион Мардарьич!»

С другой стороны, технологии в мире развиваются быстро, и иной раз злоумышленника, использовавшего новую технологию, не удается привлечь к ответственности просто потому, что совершенное им преступление просто законом не предусмотрено. Так, в ночь с 4 на 5 мая 2000 г. многие пользователи интернета получили в свои почтовые ящики сообщения с темой «ILoveYou». Внутри письма содержался скрип так называемого интернет-червя LOVE-LETTER-FOR-YOU.txt.vbs, который заменял собой многие файлы в операционной системе Windows, фактически разрушая компьютер, и рассылал себя по всем адресам из адресной книги на локальном компьютере. В итоге пораженными оказались десятки миллионов компьютеров по всему миру. Некоторые говорят о 10-процентном поражении всего Интернета. Суммарный ущерб, причиненный этой хакерской атакой, оценивается в 10 миллиардов долларов — она была внесена в Книгу рекордов Гиннеса, а в 2011 г. ее история легла в основу фильма “I love you”, не попавшего, впрочем, на широкий экран.

Преступников удалось найти довольно быстро, благодаря их же собственной беспечности. В программном коде червя было слово *“GRAMMERsoft”*, уже знакомое полиции филиппинской столицы Манилы по другому червю, значительно менее вредоносному, а основной принцип его работы описан в дипломной работе одного из авторов. Так полиция вышла на Ирен и ее брата Онеля де Гусман, студентов местного компьютерного колледжа (в 2001 г. преобразованного в университет), а также на школьного друга Онеля Майкла Буэна. Несмотря на то, что преступление было очевидно, преступники быстро пойманы и от своей вины они не отпирались, осудить их не удалось по той простой причине, что подобного рода преступления

не были предусмотрены филиппинским законодательством.

Онель де Гусман после этой шумной истории постарался скрыться подальше с глаз. 20 лет его не было ни видно, ни слышно, но в 2020 г. его разыскал английский журналист Джефф Уайт, работавший тогда над своей первой книгой о киберпреступности. Онель де Гусман все это время жил в той же Маниле и работал в скромной мастерской по ремонту мобильных телефонов. Он с легкостью согласился на интервью и рассказал, что целью создания червя было просто хищение чужих паролей: в те годы доступ в Интернет предоставлялся как правило по телефонным линиям при помощи модема через дозвон (dial-up). Когда соединение с провайдером устанавливалось, ему надо было передать информацию о пользователе, его учетной записи и оплате им услуг связи. Денег у Онеля, как всегда, не хватало, поэтому он считал само требование платить за доступ в Интернет ущемлением своих гражданских прав. К его чести, надо сказать, что в этом интервью он полностью взял на себя всю вину за столь разрушительную хакерскую атаку, пусть и совершенную им невольно.

Вероятно, технологии и дальше будут развиваться столь же быстро. Подобного рода препоны к торжеству правосудия можно избежать, если позаботиться о том, чтобы среди законодателей было больше технически грамотных людей. Однако есть и иной подход к этой проблеме: поручить исследование законодательной системы искусственному интеллекту. Это не такая уж сложная задача, как может показаться: в конце концов, на сегодня одна областей математики, где искусственный интеллект показал свою наибольшую эффективность, — проверка доказательств математических теорем. Исследование законодательной системы на полноту и непротиворечивость — задача

того же рода и примерно того же уровня сложности. Этот вопрос сейчас вызывает довольно много споров и разнообразных дискуссий: как уже говорилось выше, далеко не всем выгодно, чтобы законодательная система была полной и непротиворечивой.

4.5. Энергетика

Человеческая жизнь напрямую связана с производством и потреблением энергии, и подсчеты экспертов показывают, что в ближайшие десятилетия ее потребуется значительно больше, чем сейчас.

Здесь необходимо сделать небольшое физическое отступление, потому что разговоры об энергии часто порождают различные недоразумения, связанные с некоторой неточностью терминов. Энергия относится к так называемым сохраняющимся, или инвариантным, величинам. Каков бы ни был процесс, ее количество не уменьшится и не увеличится. Пользуясь аналогией одного великого физика XX в. Ричарда Фейнмана, скажем, что энергия подобна бухгалтерскому учету: если вы видите, что сегодня в кассе меньше денег, чем было вчера, то это может означать, либо что их кто-то украл и они переместились из кассы в чужой карман, либо что на них что-то купили и представляемая ими стоимость присутствует теперь в иной форме.

Когда вы сжигаете бензин в цилиндрах своего автомобиля, электромагнитная энергия химической связи углеводородных соединений высвобождается в виде тепла. Из-за того, что внутри цилиндра оказывается намного теплее, чем снаружи, находящийся там поршень приходит в движение, а вместе с ним и весь автомобиль. Часть тепловой энергии превращается в механическую, позволяя выполнить полезную работу, и автомобиль

переезжает куда-то в другое место. Работа совершается либо против сил тяжести, и автомобиль заезжает в гору, либо против сил трения, и затрачиваемая на выполнение этой работы механическая энергия превращается обратно в тепло. А заехав в гору, автомобиль рано или поздно должен будет с нее съехать, и энергия снова превратиться в тепло. Иначе говоря, сжигаемая в автомобиле энергия никуда не девается, даже ее количество остается неизменным. Но в двигателе внутреннего сгорания определенным образом изменяется ее форма: энергия электромагнитного поля деградирует до тепловой. Наличие перепада температур позволяет часть тепловой энергии преобразовать в механическую, которая потом снова деградирует до тепловой. Тепловую энергию уже никак невозможно было бы использовать, если бы температура везде была одинаковая, но до этого нам пока еще далеко.

Конечно, у Земли есть мощный донор — Солнце: Земля получает от него очень много энергии в виде света, и из-за того, что это освещение очень неравномерно, пока Солнце светит, температура на Земле не выровняется. Энергия Солнца испытывает на Земле много превращений, совершает полезную работу, но потом снова превращается в электромагнитное излучение и возвращается в космическое пространство только теперь в инфракрасном диапазоне. Конечно, Земля излучает немного видимого света — например, подсветка городов, но эта часть в энергетическом балансе несопоставимо меньше основной. Кроме того, сейчас Земля излучает немного больше энергии, чем получает от Солнца: это все то, что выделилось в результате сжигания нефти, газа и других полезных ископаемых. Но когда-то, миллионы лет назад, часть солнечной энергии шла на синтез биомассы, а та, в свою очередь, попадая в земные недра, преобразовывалась в нефть и газ. Биомасса росла, из-за этого Земля

излучала немного меньше, чем получала; сейчас биомасса сокращается, из-за этого Земля излучает немного больше, чем получает.

Энергия сама по себе не нужна, как не нужны сами по себе деньги — нужен их поток, чтобы они приходили и расходовались. У землян есть основной поток: энергия, приходящая от Солнца и возвращаемая в космос. Часть солнечной энергии застревает на Земле в виде движущейся воды или горючих ископаемых. Но рано или поздно она тоже вернется в космос в виде электромагнитного излучения. Вопрос в том, чтобы расходовать эту задержавшуюся тут ненадолго энергию разумно.

Энергия еще в одном подобна деньгам: когда поток вдруг прерывается, случается кризис, обвал. В электроэнергетике одну из разновидностей такого кризиса принято называть блэкаутом. Самая масштабная авария подобного рода в Москве случилась 25 мая 2005 г. В 11:10 утра стали обесточиваться токопроводящие шины в тоннелях метрополитена. Поезда стали внезапно останавливаться посреди перегона, в вагонах гас свет, а машинисты даже не успевали предупредить пассажиров. Вскоре обесточенными оказались 52 из 170 станций. 43 состава оказались заблокированными в тоннелях. С 11:40 началась эвакуация пассажиров, которая завершилась только к 13:15.

На улицах города перестали работать светофоры. Встала троллейбусная и трамвайная сеть. На улицах возникли пробки. Людям приходилось перемещаться пешком.

Встала Западная водопроводная станция, и в результате примерно четверть города оказалась без водопровода и канализации. Обесточены были три станции аэрации, и все время блэкаута сточные воды сбрасывались в Москву-реку неочищенными. Во многих местах не работала мобильная связь. Около полутора тысяч человек были

вынуждены несколько часов просидеть в обесточенных лифтах. К вечеру жизнь более или менее вернулась в нормальное русло, но отдельные последствия ощущались еще долго.

Гораздо более масштабный кризис был вызван межрегиональным сбоем энергоснабжения в электросетях Аргентины и Уругвая 17 июня 2019 г. Без электричества остались некоторые районы в Бразилии, Парагвае и Чили, а это несколько десятков миллионов человек. К счастью, проблемы были относительно быстро ликвидированы.

Хуже складывались дела в той же Аргентине двумя десятилетиями раньше, когда 15 февраля 1999 г. обесточенными оказались всего 10 кварталов Буэнос-Айреса с населением около 600 тысяч человек, но на устранение проблем потребовалось 11 дней. В городе стояла жара в 30 градусов, продовольствие в неработающих холодильниках быстро портилось, магазины не работали, да и выходить из квартиры было опасно: на лестничных клетках дежурили грабители — они врывались в квартиры, когда кто-нибудь пытался оттуда выйти или туда войти. На улицах хозяйничали мародеры.

Как ни странно, подобные события отнюдь не редкость: ежегодно новостные СМИ сообщают о 5–10 блэкаутах, причем их последствия тем более заметны, чем более экономически развита страна. Причины их тоже понятны: в случае московского блэкаута 2005 г. триггером стал пожар 24 мая на Чагинской подстанции, из-за чего она была полностью остановлена. Производимая ею мощность оценивается в 500 киловатт. По данным на сайте московской мэрии, в 2020 г. среднее энергопотребление города составляло 7 гигаватт, а в 2016 г. — 6,2 гигаватта. В 2006 г., по горячим следам, эксперты называли для летних максимальных значений 12,6 гигаватта, а для зимних — 15 гигаватт. То есть перепады потребляемой

мощности на протяжении года могут достигать 5–10 раз. Мощность Чагинской подстанции в сравнении с последней величиной — это 0,0033 %. Как же мог столь незначительный скачок мощности привести к таким разрушительным последствиям?

Проблема заключается в том, что кроме мощности есть еще и сила тока в конкретных проводах. Для каждого из них в зависимости от площади поперечного сечения и материала есть жесткое ограничение по силе тока — если его нарушить, провод сгорит. Когда сила тока в проводе начинает расти, срабатывает система безопасности и данный участок выключается из сети. Нагрузка перераспределяется, и сила тока начинает увеличиваться где-то в соседнем проводе, соседнем участке цепи. Дело, таким образом, не только в потребляемой мощности, но и в том, как ток нагрузки распределяется по сети. Если нагрузка оказывается слишком высока, то выключение одного участка немедленно влечет за собой отключение соседних и происходит своего рода цепная реакция, которую в этом случае принято называть «веерными отключениями». Это автоматизированный процесс с положительной обратной связью, и в соответствии с общей теорией в такой цепи всегда есть опасность коллапса.

Чтобы этого не происходило, современные энергетические сети организованы в четыре очереди: есть энергоустановки базисного уровня — они работают практически непрерывно, обеспечивая более или менее постоянную мощность; далее идут установки переменного уровня, вроде приливных электростанций или солнечных батарей — производство энергии на них довольно дешево и сопряжено с наименьшим ущербом для окружающей среды, но у человека очень ограничены возможности влиять на уровень производимой ими мощности; установки полупиковой нагрузки вводятся в действие только тогда,

когда есть опасность резкого увеличения потребляемой мощности, — производимая на них энергия получается очень дорогой и весьма травматичной для окружающей среды; наконец, есть установки пиковой нагрузки, вроде дизель-генераторных станций — они дают самую дорогую и самую грязную электроэнергию, зато обладают высокой мобильностью и могут быть быстро введены в действие. Во время кризиса в Москве в 2005 г. именно они существенно облегчили его течение, хотя из-за транспортного коллапса преимущество мобильности было сильно ограничено.

Искусственный интеллект в энергетике пока задействован очень мало, хотя планирование распределения нагрузки между этими четырьмя очередями и программная поддержка оптимальности такого распределения уже давно не новость. Есть уже прецеденты, когда именно ошибка такой программы привела к блэкауту: такое, например, случилось 14 августа 2003 г. на компьютере в североамериканском штате Огайо, из-за чего на несколько дней остались без электричества 55 миллионов жителей США и Канады.

В 2019–2020 гг. разные официальные лица РФ на разных публичных форумах высказывались относительно необходимости «срочной трансформации» электроэнергетики. В первую очередь это касалось все более широкого внедрения возобновляемых источников энергии, которые должны заменить ископаемое топливо (для справки: в настоящее время на производство электроэнергии расходуется от 5 % до 15 % добываемой нефти). Именно в связи с ними роль искусственного интеллекта видится наиболее существенной. И в этом есть своя логика: возобновляемый источник энергии — это, например, солнечная батарея на крыше жилого дома или несколько ветряков на пустыре. Мощность невысокая, колебания

непредсказуемые. Разумнее всего накапливать электроэнергию в аккумуляторных батареях, а потом расходовать по мере надобности. В хозяйстве расход также неравномерен, а выбор времени, когда надо включить стиральную машину или выехать на газонокосилке для приведения в порядок участка перед домом, в значительной степени произволен. Получается идеальная оптимизационная задача для алгоритма с нейронной сетью. Такие задачи решаются на уровне домохозяйств, кондоминиумов, кварталов или даже относительно небольших населенных пунктов. Но остаются оптимизационные задачи в энергосетях целых регионов, которые не только снимают пики в энергопотреблении, но и минимизируют вероятность блэкаутов типа тех, о которых шла речь выше. Для этого они должны быть интегрированы со структурами городского хозяйства и промышленности, очень часто на уровне прямой коммуникации устройств, потребляющих энергию, и устройств, распределяющих ее, то есть посредством интернета вещей (IoT).

4.6. Связь

Главная проблема, или, точнее, тип проблем, решаемые экономической наукой, — нехватка ресурсов. Это в полной мере относится и к связи: для передачи нужного количества информации, как правило, не хватает пропускной способности имеющихся каналов. Из-за этого информацию приходится сжимать, обрезать, задерживать ее передачу, когда это возможно. В то же время нередки ситуации, когда информационный канал длительное время не используется. Технология интернета, разработанная в значительной степени благодаря исследованиям американского ученого белорусского происхождения Пейсаха (Пола) Барана и внедренная в коммуникационную сеть

американского военного ведомства ARPAnet в конце 1960-х гг.

Центральная идея заключалась в том, что всякое сообщение можно порезать на части и передавать кусками. Естественно, такой кусок, кроме собственно какой-то части сообщения, должен был еще содержать указания, от кого, кому, в какое место вставить. Все вместе стало называться пакетом. Когда все пакеты собираются у адресата, компьютер с легкостью их собирает в исходное сообщение, независимо от того, по какому маршруту они добираются. По начальному замыслу часть сети могла в любой момент оказаться недоступной из-за ядерной бомбардировки, поэтому для каждого пакета мог потребоваться свой путь. Потом, когда вероятность такой бомбардировки стала относительно невысока, технология оказалась очень полезной, чтобы выбирать для пакетов оптимальный путь в условиях, когда пропускная способность каких-то частей сети резко падала из-за ее перегруженности.

Считается, что сейчас есть *четыре основных направления, по которым будут внедряться технологии искусственного интеллекта в развивающиеся телекоммуникации: 1) инфраструктура; 2) предиктивное обслуживание; 3) виртуальные помощники; 4) предотвращение злоупотреблений*. Рассмотрим их по порядку.

Инфраструктура. Уже в самой идеологии Барана угадываются перспективы прогностических способностей машинного обучения: анализ данных об относительной загрузке тех или иных участков коммуникационной сети дает достаточную информацию для предсказания, какой маршрут для данного пакета окажется оптимальным. В узловых компьютерах не надо будет оптимизировать пересылку до следующего узлового компьютера — достаточно будет только просчитывать коррективы к уже

вычисленному маршруту. Так примерно действует Яндекс.Навигатор: он сначала предложит водителю целый маршрут до места назначения, но по пути может оказаться, что маршрут лучше немного изменить.

Но это не самая сложная и не самая важная часть проблемы. Система телекоммуникаций непрерывно развивается: появляются новые типы носителей информации, новое оборудование, допускающее не использовавшиеся раньше технические решения. Например, металлический кабель может быть заменен на оптоволокно, а несущая частота увеличена с 2 до 5 мегагерц. Где и как следует произвести модернизацию? Опять-таки ответ может быть получен из анализа данных, пересылаемых в существующей сети.

Предиктивное обслуживание. Анализ тех же самых данных позволяет нейронной сети предсказывать вероятность сбоя в той или иной части сети и его характер. Таким образом, профилактическое обслуживание, в любом случае необходимое, проводится не столько в соответствии с рекомендациями регламента, которые, хотя и формулируются на основе предшествующего опыта эксплуатации сетей такого типа, но все-таки составляются фактически вслепую, а и в соответствии с конкретной ситуацией. Искусственный интеллект принимает во внимание значительно большее количество факторов, чем может принять во внимание человек, и потому человеку не всегда понятно, как этот прогноз сформирован. Для иллюстрации этого тезиса приведем только один пример.

Для поддержания бесперебойной работы сотовой сети необходимо регулярно осматривать мачты, на которых устанавливаются антенны. Для этого некоторые сотовые операторы, например AT&T, используют дроны. Сотни дронов регулярно облетают десятки тысяч мачт, производя тысячи часов видеоконтента. Нейронная сеть

анализирует этот видеоконтент на предмет распознавания в нем признаков погодных повреждений, неудачно свитых птичьих гнезд, оборвавшихся кабелей.

Виртуальные помощники. О них уже достаточно было сказано выше. Здесь можно только добавить, что некоторые телекоммуникационные компании используют виртуальных помощников, разработанных другими, — так, крупный телевизионный американский провайдер DISH Network использует Алексу Амазона. Виртуального помощника Елену для российского оператора сотовой связи МегаФон разрабатывал Яндекс.

Злоупотребления. Главная проблема заключается в том, что чем шире спектр предоставляемых оператором услуг, тем больше возможностей у злоумышленников. Нередко непосредственный урон оказывается побочным эффектом вполне благих намерений. Конечно, намерения Онеля де Гусмана в рассмотренной выше истории с интернет-червем ILoveYou трудно назвать благими, но он всего лишь хотел красть пароли и совершенно не ожидал столь разрушительных последствий. А автор первого интернет-червя, привлекшего к себе внимание СМИ, Роберт Моррисон уж точно не хотел ничего плохого: как ему казалось, он всего лишь придумал простой способ посчитать, сколько компьютеров находились в тот день онлайн. Его программа была запущена в сеть 2 ноября 1988 г. и немедленно вывела из строя 10 % из 60 тысяч компьютеров, составлявших тогда Интернет. В отличие от филиппинских законов, оказавшихся бессильными против брата и сестры де Гусманов, американский суд присяжных счел Моррисона виновным и 4 мая 1999 г. к условному заключению на три года, 400 часам общественных работ и 10 тысячам долларов штрафа.

Но история компьютерных вирусов началась как минимум на 20 лет раньше, и большинство из них

создавались вовсе не из любви к искусству. Примерно к середине 1980-х начали появляться первые антивирусные программы, которые представляли собой просто библиотеки характерных признаков всех известных на тот момент вирусов с прикрепленным к ним программным кодом. Этот код просматривал активные приложения в памяти компьютера и предупреждал владельца, если находил какое-то сходство. Довольно распространенный тогда тип мошенничества заключался в том, что вирус рассылался через электронную почту и блокировал работу компьютера. Одновременно владелец компьютера получал сообщение, какую сумму, куда и каким образом надо перевести, чтобы компьютер разблокировать. Новые вирусы появлялись чуть ли не каждый день, библиотеки антивирусных программ надо было обновлять все чаще и чаще, за это тоже приходилось платить. Но игра стоила свеч: регулярные взносы производителю антивирусных приложений оказывались значительно менее затратны, чем один раз заплатить мошеннику, удачно заблокировавшему компьютер.

В некий момент это все вдруг прекратилось. И никакой загадки тут нет: вирусов стало достаточно много для обстоятельной статистики. Обученная электронная сеть достаточно эффективно распознает «мальварь» (угрожающий компьютеру программный код), чтобы практически исключить возможность для нового вируса или червя незамеченным пробраться на компьютер. Такого рода «интеллектуальные» системы есть сейчас практически у любого интернет-провайдера, независимо от того, каким именно образом и какого рода услуги предоставляются. Но из этого отнюдь не следует, что масштабы киберпреступности снизились. Напротив, по меткому замечанию цитированного уже выше Джеффа Уайта, очень многим сейчас стало ясно, что «грабить отдельные личности

и целые институты онлайн гораздо безопаснее и значительно прибыльнее». И если искусственный интеллект используется для того, чтобы обезопасить телекоммуникации, то искусственный интеллект принимается на вооружение и теми, кто хочет использовать средства связи для причинения урона кошелькам конкретных людей, банковским счетам целых организаций, экономическому или политическому статусу целых государств.

4.7. Финальная безработица

30 июля 2021 г. Федеральный суд Австралии официально признал, что автором изобретения может быть «не человек», и выдал патент интеллектуальной системе DABUS. Это произошло ровно через два дня после того, как южноафриканский журнал South African Patent Journal опубликовал патент на то же самое изобретение. Но если южноафриканская публикация не имела официального статуса, решение австралийского суда обладает силой закона.

За неделю до этого Сбербанк впервые в России зарегистрировал в Роспатенте программный код, созданный искусственным интеллектом Sber AI, который так и указан в патенте в качестве разработчика.

Оба этих примера показывают, как искусственный интеллект начинает приобретать права, до того присущие только людям. Мы уже упоминали истории, когда компьютеры писали стихи и мелодии, сочиняли короткие рассказы или предлагали разумные конструкторские решения. Признание за машинами авторских прав приходило очень медленно, преодолевая понятное сопротивление. Первые прецеденты созданы, и дальше можно ожидать лавинообразного роста подобных примеров. Но помимо правового аспекта тут есть и другой.

Искусственный интеллект все больше приобретает способность выполнять человеческую работу. Это означает, что людям эту работу скоро делать будет не надо. Автономные такси сделают ненужной профессию таксиста, автономный комбайн еще больше сократит процент населения, занятого в сельском хозяйстве. Еще в 2017 г. Г.О. Греф говорил, что за 2016 г. в Сбербанке были уволены 450 юристов, потому что их работу стал выполнять искусственный интеллект. По мере расширения используемых алгоритмов в юридической практике будет появляться все больше и больше безработных юристов.

Еще 40–50 лет назад в кабине гражданских самолетов было место для четырех членов экипажа: двух пилотов, штурмана и бортинженера. В современных «Боингах» и «Аэробусах» мест всего два — для двух пилотов. Бортинженерам пришлось переучиваться на новую профессию уже давно. Штурманы на самолетах старых марок «Илах» и «Ту» продолжали летать до самого недавнего времени. Сейчас и им приходится искать себе новую работу.

Американский исследователь Ганс Моравец называет это явление «финальной безработицей»: все большему количеству специалистов в наши дни грозит окончательная потеря работы. Люди теряют свои позиции просто потому, что их работу быстрее, лучше и дешевле сделает машина — AI или CPS. В статье еще 1998 г., озаглавленной “When will computer hardware match the human brain”, Моравец писал: «Компьютеры — универсальные машины, и их потенциал равномерно покрывает безграничное разнообразие задач. Потенции людей, напротив, сосредоточены там, где от успеха зависит выживание, в более отдаленных областях они весьма слабы. Представьте себе “ландшафт человеческих компетенций”, где есть низины вроде “арифметики” и “механической памяти”,

холмики вроде “шахмат” или “доказательства теорем” и горные пики, отмеченные указателями “перемещение с места на место”, “координация движений рук и глаза”, “социальное взаимодействие”. С совершенствованием компьютеров этот ландшафт словно наполняется водой: полвека назад она затопила низины, вымыв оттуда счетоводов и писцов, но оставив нас сухими. Сейчас вода дошла до холмиков, и обитатели наших аванпостов забеспокоились: куда бы им переместиться? Мы чувствуем себя в безопасности на своих пиках, но, учитывая скорость, с которой вода прибывает, она покроет и пики в ближайшие полвека. Я полагаю, нам уже пора начинать строить ковчеги и приучаться к жизни на плаву».

В одном из своих интервью Билл Гейтс как-то выразил предположение, что рано или поздно наступит время, когда людям вообще не надо будет ничего делать: вся мыслимая работа — от снабжения людей продовольствием до управления государством — будет выполняться машинами. Людям придется искать себе другие занятия. Например, посвятить себя компьютерным играм. Хорошо это или плохо?

Конечно, с одной стороны, потеря работы для большинства из нас — трагедия. На что жить дальше? С другой стороны, если посмотреть списки дорогих автомобилей, угнанных за последний месяц у жителей, например, Москвы, то выяснится, что очень многие из них нигде не работают. Мало того, у каждого из нас есть знакомые, которые, хотя где-то и числятся работающими, ничего там не делают, а живут припеваючи. А сколько людей, много и тяжело работающих, и неплохо зарабатывающих, люто и искренне ненавидят свою работу? За сто лет время, проводимое на рабочем месте, для большинства профессий сократилось почти вдвое, и никто никогда по этому поводу не сокрушается. Мало кто отказался бы сократить

рабочее время еще хотя бы на пару часов при той же зарплате.

Ковидные ограничения 2020–2021 гг. и перевод на удаленный формат работы для многих стал тяжелым испытанием. Оказалось, что для большого количества людей выход на работу — это не только средства к существованию, но и расширение круга общения, необходимая для них социализация. Но не меньшее число людей обнаружили, что работать из дома им приятнее, а производительность их труда не только от этого не страдает, а наоборот, сильно увеличивается. По мнению экспертов, даже если бы сейчас пандемия вдруг резко исчезла и не было бы больше никакой опасности роста числа заболеваний, вернуться к прежним трудовым отношениям уже бы не получилось: ящик Пандоры открыт, значительное количество профессионалов вернуть в офисы больше не удастся.

Труд дает людям не только материальные блага, но и ощущение собственной значимости, смысл существования. Если, предположим, дело повернется так, что, даже не работая, можно будет получать достаточное содержание, чтобы существовать безбедно, утрата смысла окажется убийственной. Для любого этического учения вопрос о смыслах, казалось бы, должен находиться в центре внимания, но почему-то на практике это редко случается. И рассуждения о работе в рамках этических концепций подразумевают связку «работа — заработок — смысл». Может ли быть осмыслена работа, не приносящая заработка?

М. Тегмарк в своей книге «Жизнь 3.0» (2019) приводит высказывания двух великих физиков XX в. о смысле жизни. Один из них — Стивен Вайнберг, Нобелевский лауреат, знаменитый борец с религиозным обскурантизмом, — говорил в своих популярных лекциях, что «чем

лучше мы понимаем Вселенную, тем более бессмысленной она нам кажется». Другой — Фримен Дайсон, так и оставшийся без своей Нобелевской премии создатель квантовой электродинамики и многих других важных концепций современной физики, — говорил, критикуя позиции Вайнберга, что когда-то Вселенная *была* бессмысленной, но стала наполняться смыслом с возникновением жизни. Пик осмысленности впереди — когда жизнь сможет распространиться в космосе. Выход человека за пределы земной атмосферы впервые состоялся чуть больше полувека назад. Ему сопутствовал взлет энтузиазма, когда казалось, что вот еще чуть-чуть, и вся галактика будет покорена. Но все не так просто. Даже высадка на Луну как-то проблематична. Искусственный интеллект принципиально меняет в этой истории акценты.

Минимальное расстояние от Луны до Земли — 356 тысяч километров, максимальное — 407 тысяч километров. То есть радиосигнал от Земли достигает Луны примерно за секунду. Уже для Марса порядки совсем иные: минимальное расстояние 55 миллионов километров, а максимальное — 401 миллионов километров, и радиосигнал проходит это расстояние от 3 до 22 минут. Луноходом, движущимся по поверхности Луны с небольшой скоростью, можно было управлять с Земли, хотя для этого приходилось набирать специальный персонал из людей, незнакомых с управлением автотранспортом: секундная задержка никак не вписывалась в рефлексы даже не очень опытного водителя. Задержка в несколько минут для роверов на поверхности нашего ближайшего соседа по Солнечной системе делает дистанционное управление невозможным. В этом случае можно рассчитывать только на искусственный интеллект.

Выход за пределы подлунного мира будет связан с дальнейшим развитием технологий искусственного

интеллекта. Человеческому интеллекту, возможно, будет оставаться все меньше и меньше места для приложения своих способностей на родной планете, но за пределами ее в тесном сотрудничестве с искусственным интеллектом места для приложения своих способностей будет становиться все больше и больше.

Ключевые понятия: роботы-судьи, беспилотный транспорт, Тесла, мальварь, хакер, вирусная атака, автономная энергетика, блэкаут.

Контрольные вопросы

1. Для решения каких задач технологии искусственного интеллекта используются в наши дни?
2. Чем объясняется выбор названия MTurk для краудсорсинговой платформы компании Amazon и какие задачи она способна решать?
3. Что, на ваш взгляд, проще: организовать движение беспилотных автомобилей при отсутствии автомобилей, управляемых человеком, или в общем транспортном потоке? Почему?
4. Каковы преимущества рентгенологической диагностики, выполняемой искусственным интеллектом, в сравнении с аналогичной работой врача-рентгенолога?
5. Каковы наиболее близкие последствия массового внедрения технологий искусственного интеллекта в судебную систему?

Практико-ориентированные задания

1. Выберите по своему усмотрению три-четыре случая блэкаута за последний 10–15 лет. Сравните причины и последствия. Опишите, как возникшие проблемы могли бы разрешиться с использованием обученных нейронных сетей.

2. Самостоятельно найдите и проанализируйте случаи спонтанно развившейся необъективности автономной системы.

3. Оцените вероятность встретить беспилотный автомобиль на дорогах общего пользования в разных странах.

Темы сообщений, докладов и эссе

1. Стоит ли заклеивать камеру на своем ноутбуке?

2. До какой степени можно доверить искусственному интеллекту самостоятельно выносить приговоры в судебных делах?

3. Машинное обучение и энергетика.

ГЛАВА 5.
РИСКИ И ОПАСНОСТИ

В результате изучения материалов главы обучающийся должен:

знать

– негативные стороны применения технологий AI и связанные с ними опасности;

уметь

– анализировать исторические причины конфликтных ситуаций, в которых технологии AI сыграли, могли сыграть или сыграют в будущем негативную роль;

владеть навыками

– критического анализа различных историко-научных явлений и фактов в целях предотвращения конфликтных ситуаций, связанных с использованием технологий AI.

5.1. Автономное оружие

Мысль о желательности максимально сократить вовлеченность людей в военные конфликты уже давно кажется привлекательной многим. Конечно, лучше всего было бы и вовсе как-то научиться обходиться без войн. Но пока это не получается, почему бы не сократить число человеческих жертв, исключив появление людей на поле боя? Пусть там сражаются автоматы. Тот, кто первым останется без умных машин, будет вынужден сдаться, как сдавались некогда воины, когда у них из рук выбивали шпагу.

В США в июне 2018 г. при Министерстве обороны был создан Объединенный центр искусственного интеллекта

(JAIC, читается как «джейк»). Его создание сопровождалось публикацией в феврале 2019 г. «Стратегии развития искусственного интеллекта Министерства обороны США», в которой, в частности, говорится: «Министерство обороны США призвано обеспечивать защиту нашей нации, предотвращая военные конфликты или побеждая в них, если сдерживание не удается. Выполняя эту миссию, мы всегда находились на переднем крае технологического прогресса, обеспечивая постоянное конкурентное военное превосходство над теми, кто угрожает нашим спокойствию и безопасности.

Искусственный интеллект (AI) — одно из важнейших технологических достижений. AI подразумевает способность машин решать задачи, для которых обычно требуется человеческий интеллект — например, выявлять закономерности, учиться на ошибках, извлекать уроки, делать прогнозы или принимать меры. Решение при этом может быть сугубо цифровым, а может приниматься интеллектуальным программным обеспечением автономных физических систем. AI сейчас преобразовывает все отрасли промышленности и, как ожидается, проникнет во все сферы деятельности Министерства обороны. Это относится к планированию операций, набору, обучению, подготовке и защите личного состава, поддержанию его здоровья и ко многому другому. С применением AI в оборонной сфере у нас есть возможность улучшить содержание и защиту американских военнослужащих, повысить безопасность наших граждан, лучше защищать наших союзников и партнеров, а также повысить доступность наших операций и скорость реагирования.

Другие страны, в частности Китай и Россия, прилагают значительные усилия для развития AI в военных целях и разрабатывают приложения, которые вызывают вопросы с точки зрения международных норм и прав человека.

Эти усилия угрожают подорвать наши технологические и организационные преимущества, дестабилизируют свободный и открытый международный порядок. Соединенные Штаты вместе со своими союзниками и партнерами должны внедрять AI, чтобы сохранить свои стратегические позиции, обеспечить преимущество на будущих полях сражений и защитить мировой порядок. Мы также будем стремиться разрабатывать и использовать технологии AI таким образом, чтобы они способствовали укреплению безопасности, мира и стабильности в долгосрочной перспективе. Мы будем играть ведущую роль в ответственном использовании и развитии AI, сформулировав наше видение и руководящие принципы использования AI законным и этичным образом».

Российские документы подобного рода обычно бывают недоступны, но СМИ сообщили о заседании Совета безопасности РФ 22 ноября 2019 г., на котором выступил Президент РФ В.В. Путин. В частности, он сказал:

«[...В] следующем году предстоит приступить к формированию Госпрограммы вооружения до 2033 г.[, которая должна быть принята в 2023 г.] и, соответственно, программы развития оборонно-промышленного комплекса.

Основная задача нового периода — это наращивание качественных и количественных характеристик вооружения и техники. Речь идет о современных и перспективных образцах высокоточного оружия и средств воздушно-космической обороны, активном применении технологий искусственного интеллекта при создании военной продукции. В том числе должна быть расширена линейка беспилотных разведывательных и ударных летательных аппаратов, лазерных и гиперзвуковых систем, оружия, основанного на новых физических принципах, а также роботизированных комплексов, способных выполнять разноплановые задачи на поле боя».

Таким образом, наблюдается новая волна гонки вооружений, на этот раз в сфере так называемого автономного оружия. Подобного рода системы уже существуют и даже стоят на вооружении. В принципе даже система «Иджис», о которой шла речь в гл. 3, может служить прототипом, хотя в ней использовалась технология типа GOFAI. В настоящее время речь преимущественно идет о нейронных сетях и машинном обучении. Ведется дискуссия о включении человека в цепь управления, но следует иметь в виду, что инцидент 3 июля 1988 г. произошел, когда человек был в цепи управления и подтвердил ошибочное решение, принятое автоматом.

Эксперты-разработчики систем AI высказывают разные мнения по поводу происходящего. Для кого-то важны деньги. Кто-то встает за принцип. Генеральный директор российской компании "Cognitive Technologies" Ольга Ускова описала в своем блоге 25 июля 2021 г. разговор с неназванным собеседником:

«— Ольга, вот с одной стороны, вы показываете в своих аналитических материалах довольно страшные возможные варианты развития международной ситуации при военном использовании ИИ, а с другой стороны Cognitive официально декларирует, что отказывается предоставлять свои новейшие разработки в области Искусственного Интеллекта для использования военными организациями в ЛЮБОЙ СТРАНЕ МИРА, в том числе в России.

— Да. И что?

— Но в этом нет логики. Вы откажетесь, на ваше место встанет другая компания. С одной стороны, и вы не заработаете, и для нашего вооружения будут использованы не лучшие варианты. Вы все равно не переломите ситуацию, зачем же этот демарш?

— У меня одна из любимых песен Высоцкого “Тот, который не стрелял”. Помните о чем?

“Мой командир меня почти что спас,
Но кто-то на расстреле настоял...
И взвод отлично выполнил приказ, —
Но был один, который не стрелял...”

Ну, то есть мы не знаем, какая пуля в общем оружейном залпе окажется смертельной. И да, Cognitive — очень небольшая компания, особенно относительно мировой военной машины. Но если я не могу изменить ситуацию целиком, я просто выполню свой долг. Изменю ее у себя в душе, даже если это выглядит глупо и нелепо. Мне кажется, если Бог есть, то он в этом. В личной ответственности за СВОИ поступки».

Это не уникальная позиция. Похожие убеждения выразили еще около 20 тысяч специалистов разного рода, в том числе Стивен Хокинг и нобелевский лауреат Фрэнк Вильчек, подписав открытое письмо специалистов по искусственному интеллекту и робототехнике об автономном оружии. Авторами этого письма стали М. Тегмарк, основатель и бессменный директор Института будущего жизни (FLI), и С. Рассел, основатель и руководитель Центра человекосовместимого искусственного интеллекта (CHAI) при Калифорнийском университете в Беркли. Это случилось в июле 2015 г., примерно через год после того, как М. Тегмарк объявил о начале работы FLI — на сегодня наиболее авторитетной некоммерческой организации, финансируемой Илоном Маском и привлекающей ведущих ученых мира к решению проблем безопасности в применении искусственного интеллекта. На публикацию этого письма отреагировала практически вся мировая пресса. Ему посвящена отдельная статья Википедии. Оно до сих пор размещено на официальном сайте FLI. В нем, в частности, говорится: «Ключевой

вопрос для человечества сегодня таков: начать глобальную гонку AI-вооружений или предотвратить ее. Если каждая крупная военная держава подталкивает инженеров к разработкам AI-оружия, то глобальная гонка вооружений практически неизбежна и конечная точка этой технологической траектории очевидна: автономное оружие станет автоматом Калашникова завтрашнего дня. В отличие от ядерного оружия, для его создания не требуется дорогостоящих или трудно получаемых материалов, поэтому оно превратится в широко распространенный, недорогой и доступный для всех значительных военных держав продукт массового производства. Его появление на черном рынке, а потом и в руках террористов, диктаторов, стремящихся к усилению контроля над населением, полевых командиров, мечтающих о новых этнических чистках, — не более чем вопрос времени. Автономное оружие идеально подходит для таких задач, как политические убийства, организация массовых волнений, подчинение населения и выборочное убийство представителей определенной этнической группы...

Как большинство химиков и биологов не проявляют никакого интереса к созданию химического или биологического оружия, большинство исследователей искусственного интеллекта не заинтересованы в создании AI-оружия и не хотят, чтобы другие порочили сферу их деятельности, производя его, провоцируя граждан на выступления против технологии искусственного интеллекта, что усложнит внедрение более благоприятных для общества способов его использования. Действительно, и биологи, и химики всемерно поддержали международные соглашения, запрещающие химическое и биологическое оружие, как большинство физиков поддерживали договоры о запрещении ядерного оружия космического базирования и ослепляющего лазерного оружия».

Этот принцип был впоследствии закреплен на Асиломарской конференции 2017 г. как один из «асиломарских принципов разработки искусственного интеллекта».

5.2. Верификация и валидация

Выше мы уже рассматривали случаи, когда использование искусственного интеллекта приводило к результатам, противоположным ожидаемым. В последнее время подобных случаев становится все больше. 26 февраля 2019 г. новостные агентства облетела новость об огромных потерях, понесенных Сбербанком, по словам Г.О. Грефа, из-за ошибок искусственного интеллекта. «Из-за того, что машина совершала маленькую ошибку на больших объемах, мы теряли миллиарды рублей», — сказал он. Правда, в тот же день пресс-служба Сбербанка уточнила, что речь шла не о потерях как таковых, а о недополученной прибыли. И сам председатель правления Сбербанка дополнил свои слова о пользе этих ошибок, позволявших на них учиться, «калибровать и верифицировать систему искусственного интеллекта».

Между тем большинство современных специалистов считают, что время метода проб и ошибок прошло. Ошибки становятся слишком дороги, чтобы такое обучение могло себя окупить в будущем. Пример со скрепками тому иллюстрация. Считается, что в техническом плане безопасность искусственного интеллекта определяется четырьмя ключевыми понятиями, выражаемыми словами «верификация», «валидация», «надежность» и «контроль». Мы не будем давать им полноценного определения, но проиллюстрируем примерами.

На протяжении почти десяти лет советская космическая отрасль готовилась к автоматической миссии на спутник Марса Фобос. Она так и называлась:

«Фобос-1». На тот момент эта была самая тяжелая межпланетная станция в предшествовавшей истории космических полетов. Миссия начиналась на Байконуре 7 июля 1988 г. Автоматическую межпланетную станцию выводила на орбиту сверхтяжелая ракета «Протон» с разгонным блоком Д-1. 28 августа дефис в переданной с Земли команде был неверно интерпретирован, и вместо включения гамма-спектрометра произошло отключение двигателей пространственной ориентации аппарата. Солнечные батареи перестали получать достаточное количество света, и к 2 сентября полностью разрядились. На этом миссия закончилась, хотя попытки связаться со станцией продолжались еще до 3 ноября.

Проблема заключалась в том, что программное обеспечение станции оказалось в недостаточной степени протестировано — оно не прошло должной *верификации*.

Другой пример мы возьмем из финансовой области. Начиная с 2010 г. на биржах время от времени стали наблюдаться явления стремительного обвала индексов с последующим столь же стремительным возвращениям их к прежним значениям. За ними укрепился англоязычный термин Flash Crush, в России первое из этих событий стали называть «черным вторником», поскольку оно пришлось на вторник 6 мая 2010 г. Тогда на протяжении дня более чем на 10 % обвалились все основные индексы — S&P500, Nasdaq, Dow Jones Industrial Average. К этому времени торговля на биржах уже выполнялась преимущественно компьютеризированными трейдинговыми системами, быстро принимавшими решение о покупке или продаже ценных бумаг. Из-за того, что при открытии бирж Dow Jones шел вниз, цены акций многих компаний стали непредсказуемо и сильно колебаться. Компьютерные системы работали правильно, но в принятии решений исходили из неверной информации: цена на акцию в момент

совершения сделки могла оказаться сильно отличной от той, которая учитывалась в вычислении, что вызывало сбой программы.

В этом случае система была протестирована (верифицирована) и работала правильно. Но сама система оказывалась не соответствующей решаемой задаче: она не могла действовать в ситуации, когда выбранные для решения параметры не соответствовали действительности и требовали пересмотра. *Валидацией* называют проверку правильности постановки целей перед используемой системой. Иначе говоря, верификация обеспечивает способность системы решить задачу, а валидация — правильность самой постановки задачи.

Гибель Роберта Уильямса на заводе «Форда» в городе Флэт-Рок — еще один пример невалидной системы. Предпосылки ее работы заключались в том, что людей поблизости нет. При работе в непредусмотренных условиях случилась трагедия.

Бортовой компьютер на «Убере» в Темпе исходил из предположения, что Рафаэле Васкес достаточно секунды, чтобы принять на себя управление машиной и выполнить все необходимые для объезда Элейн Херцберг с велосипедом маневры. Программа была верифицирована, и компьютер честно отработал весь предполагающийся протокол, но в сложившейся ситуации она была не валидна.

5.3. Этические дилеммы

Всем известны ситуации, когда хорошего решения просто нет. Они выразительно описаны в народных сказках: налево пойдешь, коня потеряешь, направо пойдешь, голову сложишь... Это из русских народных. У других народов не менее драматичные варианты. Человек может

позволить себе действовать импульсивно, не задумываясь. Компьютер может и должен все посчитать.

Дилеммы были весьма популярным жанром в фольклоре африканских племен. В 1975 г. знаменитый американский антрополог Уильям Баском издал целую книгу под названием «Африканские дилеммы» (“African Dilemma Tales”). Одна из них, услышанная им у народа попо в Дагомее, стала потом известной в сочинениях по этике как «дилемма жирафа»: «Один человек переправлялся через реку со своей женой и с матерью. На другом берегу появился жираф. Человек прицелился в него из ружья, но жираф сказал: “Если ты выстрелишь, твоя мать умрет. Если не выстрелишь, умрет твоя жена”. Что было делать человеку?»

Оба варианта нежелательны, но третьего нет. Надо какой-то из них выбрать. Какой? Первый требует определенного действия — нажатия на спусковой крючок. Второй, напротив, требует бездействия, и человек может считать, что смерть матери случилась без его участия. Однако жертвуя женой, он мог ее спасти.

Еще одна дилемма из той же книги: «Слепой человек, путешествуя со своими слепыми женой, матерью и тещей, нашел семь глаз. Два он взял себе, два отдал жене, один отдал матери и один матери жены. Что он должен сделать с оставшимся глазом? Если он отдаст его своей матери, то как он будет смотреть в глаза своей жене и ее матери? Если он отдаст седьмой глаз матери жены, то ему будет стыдно перед своей матерью».

Здесь противоположная история. Человек в силах кого-то облагодетельствовать, но он не может облагодетельствовать всех. Ему дорога жена, и ему дорога мать. Кому отдать предпочтение?

Подобного рода примеры известны со времен Античности. Иногда они использовались как риторический

прием, но иногда служили поводом для серьезных этических рассуждений. Дилеммы в контексте игровых стратегий появляются и в знаменитой книге Джона фон Неймана и Оскара Моргенштерна «Теория игр и экономическое поведение» (1947), однако самая знаменитая из игровых дилемм появилась на свет уже после выхода книги — ее придумали или, лучше сказать, обнаружили два ученых М. Флад и М. Дрешер из «РЭНД Корпорэйшн», с которой Джон фон Нейман в те годы много сотрудничал и обсуждал разнообразные игры. Эта дилемма получила название дилеммы заключенного. Мы говорим, что Флад и Дрешер обнаружили ее, а не придумали, потому что с момента публикации их первой статьи у этой дилеммы появилось множество разнообразных вариантов, и мы дадим тут относительно современный из книги Уильяма Паундстоуна «Дилемма заключенного» (“Prisoner’s Dilemma”, 1993), которая была переведена на русский язык Натальей Жуковой, но так и не увидела свет.

«Предположим, вы украли алмаз “Надежда” и пытаетесь его продать. Вы вычислили потенциального покупателя, криминального авторитета мистера Крутого — самого безжалостного человека в мире. Хотя он чрезвычайно умен, мистер Крутой также печально известен своей жадностью и равно знаменит умением обманывать. Вы согласились обменять алмаз на дипломат со стодолларовыми купюрами. Мистер Крутой предлагает, чтобы вы встретились где-нибудь на пустынном пшеничном поле, чтобы обменяться. Так не будет свидетелей.

Вы знаете, что мистер Крутой в прошлом договаривался со многими другими продавцами. Каждый раз он предлагал для обмена далекое уединенное место. Каждый раз мистер Крутой являлся и открывал дипломат, чтобы продемонстрировать свою добрую волю. Затем

мистер Крутой извлекал пистолет-пулемет и, застрелив продавца, отбывал с товаром и деньгами.

Вы скажете, что план с пшеничным полем не такая уж хорошая идея.

Вы предлагаете план с двумя полями. Мистер Крутой прячет свой дипломат с деньгами на поле в Северной Дакоте, в то время как вы прячете алмаз в поле в Южной Дакоте. Затем обе стороны идут к ближайшему телефону-автомату и обмениваются указаниями, где найти спрятанное.

В этом плане есть встроенная система безопасности (вы тактично не упоминаете об этом). Вы не должны иметь на себе ничего ценного, когда пойдете забирать дипломат мистера Крутого. Мистер Крутой (который вовсе не маньяк-убийца, а просто проницательный деловой человек) не будет иметь причин, чтобы ждать на поле в Северной Дакоте, чтобы напасть на вас. Мистер Крутой соглашается на план с двумя полями.

Вы находите пшеничное поле в Южной Дакоте. Когда вы уже совсем собрались спрятать там дипломат с алмазом, вам приходит в голову идея. Почему бы не оставить алмаз себе? Мистер Крутой никак не может узнать, что вы его кинули, пока не приедет в Южную Дакоту (а вы дождетесь его звонка и дадите ему указания, как будто все в порядке). К тому времени вы уже будете в Северной Дакоте, чтобы забрать деньги. Затем вы прыгаете в самолет до Рио. Больше вы никогда не увидите мистера Крутого.

Тут вам приходит мысль похуже. Мистер Крутой должен ведь думать точно так же! Он не глупее вас, зато, вероятно, жаднее в десять раз. У него есть равный стимул вас кинуть, и вы сможете отплатить ему за это не более, чем он вам.

Дилемма выглядит так:

	Мистер Крутой придерживается договоренности	Мистер Крутой обманывает
Вы придерживаетесь договоренности	Сделка состоялась: вы получили деньги, мистер Крутой получил алмаз	Вы ничего не получили, мистер Крутой уходит с алмазом и деньгами
Вы обманываете	Вы уходите с алмазом и деньгами, мистер Крутой ничего не получает	Много суеты зря: вы остаетесь с алмазом, мистер Крутой с деньгами

Проблема в том, что вам надо принять решение, не зная, какое решение примет мистер Крутой, и дальше с этим жить. Вы можете предпочесть получить деньги, не сдавая алмаз. Мистер Крутой предпочел бы получить алмаз за так. Однако не заблуждайтесь — вы оба были бы искренне рады совершить сделку согласно договоренности. Мистер Крутой очень хочет этот алмаз в свою коллекцию трофеев — не просто любой алмаз, а один-единственный алмаз "Надежда". Он знает, что вы — единственный шанс получить его. Вы так же точно хотите денег. Мистер Крутой предложил сказочную цену, гораздо больше, чем дал бы кто-то другой.

Наилучший исход всего дела в верхней левой клетке — результат, когда оба подчиняются сделке. Но лучший исход для каждого по отдельности — быть единственным обманщиком. Наихудший исход — оказаться лохом, который держится соглашения, в то время как второй обманывает.

Вот одна точка зрения. Ваши действия в Южной Дакоте не могут как-то повлиять на действия мистера

Крутого в Северной Дакоте. Неважно, что делает мистер Крутой, вы лучше оставьте алмаз себе. Если мистер Крутой оставит деньги, вы окажетесь с алмазом и деньгами. Если мистер Крутой ничего не оставит, по крайней мере у вас останется алмаз, чтобы продать кому-то еще. Так что вам надо обманывать и ничего не оставлять в поле.

Вот другая точка зрения. Вы оба в одной лодке. Пройдите рассуждения из предыдущего абзаца еще на шаг вперед. Мистер Крутой вполне способен прийти к тому же заключению, что "рационально" обманывать. Тогда вы оба обманете и оба получите много безрезультатной суеты. Логика запрещает сделку, выгодную обеим сторонам? Нет в этом ничего логичного! Поэтому вы должны держаться соглашения. Вы должны быть достаточно вменяемы, чтобы понимать, что обман подрывает общее благо.

Это и есть дилемма заключенного, и теперь настало подходящее время спросить себя, что бы стали делать вы?»

Но ведь, в сущности, с чем-то подобным нам приходится иметь дело каждый день. Проблема обманутых вкладчиков в России 90-х заключалась именно в том, что они вносили деньги вперед за дома, которым только предстояло быть построенными, а получившая деньги компания испарялась, так ничего и не построив. Как мы уже знаем, ценность имеют не сами деньги, а их поток. Но когда деньги утекают втуне, это гораздо хуже, чем когда они просто лежат мертвым грузом. Для нас было бы лучше все-таки получить квартиру, но нам приходится принимать решение, не зная намерений другой стороны.

Неразрешимость подобного рода дилемм создает вполне предвидимую сложность при обучении искусственной нейронной сети в современном варианте искусственного интеллекта. Вам надо задавать модель правильного поведения, например, беспилотному автомобилю — какое

решение при неизбежном столкновении лучше: когда минимально страдает пассажир или когда минимизируется общий урон, пусть даже ценой жизни пассажира? При обсуждении этических стандартов разработчиков AI большую популярность приобрела дилемма вагонетки, в своей типичной форме она звучит так: «Вагонетка несется по рельсам, к которым привязаны пять человек. Вы находитесь на мосту, который проходит над рельсами. У вас есть возможность остановить вагонетку, бросив на пути что-нибудь тяжелое. Рядом с вами находится толстый человек, и единственная возможность остановить вагонетку — столкнуть его с моста на пути. Каковы ваши действия?»

В схожей формулировке дилемма впервые возникла в работе Филиппы Фут 1967 г., посвященной проблеме абортов и приложимости в ней доктрины двойного эффекта. Эта доктрина предполагает, что зло не должно совершаться ни при каких условиях — даже в тех случаях, когда у него будут благие последствия. И напротив — благое действие может быть совершено даже в тех случаях, когда неизбежны его скверные последствия. Таким образом, эта доктрина отрицает возможность принесения в жертву невинного не родившегося создания даже ради спасения жизни матери. Филиппа Фут приводит пример потерявшего управление трамвая, несущегося на работающих на путях пятерых рабочих. У вожатого есть возможность их спасти, если направить трамвай на другой путь, но там тоже работает человек, хотя и всего один. С точки зрения Фут, она построила контрпример, опровергающий доктрину двойного эффекта: вожатый должен направить трамвай по тому пути, где работает только один человек, так как гибель этого невиновного спасет жизнь пятерых других столь же невиновных рабочих. В 1967 г., когда эта статья была

напечатана, никто не предполагал, что изложенная в ней схема будет обсуждаться не столько в контексте права женщин на аборты, сколько в контексте написания целевых функций для самообучающихся на основе искусственных нейронных сетей машин. Контрпример, как казалось, решающий этическую проблему, обернулся новой этической дилеммой. А этической проблемой, тесно связанной с этой дилеммой, оказался выбор тактики машинного обучения компаний, производящих беспилотные автомобили. Довольно логично предположить, что руководство компаний выбирает такие обучающие алгоритмы, которые гарантируют максимальную безопасность владельцу автомобиля (предполагается, что именно ему придется чаще, чем кому-либо другому, находится внутри автомобиля во время его движения), даже в тех случаях, когда подобные гарантии вступают в прямое противоречие с этическими кодексами.

5.4. Смещение на Восток

2016 г. обозначил важный поворот в судьбе искусственного интеллекта. Революция началась еще в середине нулевых, после работ группы канадских ученых, которыми руководили Джеффри Хинтон из университета Торонто и Йошуа Бенджо (в англоязычном мире его итальянскую фамилию произносят как «Бенджио») и которые научились обучать **глубокие нейронные сети**, то есть нейронные сети с очень большим числом слоев. Это положило начало эпохи **глубокого обучения.** Если говорить коротко, суть их открытия заключалась в том, чтобы собственно обучению на основе размеченных данных предшествовало *предобучение* на основе неразмеченных данных. Если воспользоваться приведенным выше классическим примером с кошками и собаками, то смысл метода в том,

чтобы показать нейронной сети много разных животных, не уточняя, кто из них кто.

Кроме этого, немалую роль в успехе нейронных сетей в XXI в. сыграл рост вычислительных мощностей и их доступность: если в августе 1997 г. стоимость одного гигафлопса[6] оценивается в $48 000 (в долларах 2020 г.), то на август 2000 г. это было уже всего $115. Ресурсы для эмуляции глубокой нейронной сети стали значительно более доступными.

В марте 2016 г. произошло знаковое событие, показавшее всему миру мощь глубокого обучения. В это время проходил поединок между компьютером AlphaGo, разработанным английской компанией "DeepMind", и одним из сильнейших в мире игроков в го Ли Седолем. У игры го есть несколько принципиальных отличий от других подобных игр, например шахмат, делающих ее особенно сложной для программирования. Достаточно будет сказать, что если общее количество разных позиций для шахмат оценивается в 10^{46}, то для го эта оценка составляет 10^{117}. Даже просто для первого хода в шахматах на 18 разных вариантов у белых есть столько же ответов у черных, а в го — на 361 вариант первого хода белых черные могут ответить 360 разными способами.

Компьютер Deep Blue, сумевший в 1996 г. обыграть чемпиона мира по шахматам Гарри Каспарова, оценивал по 200 миллионов позиций в секунду. Даже если бы он оценивал по 200 миллиардов позиций, этого было бы недостаточно для игры в го. Но программисты из DeepMind пошли по другому пути: для начала они провели длинный

[6] Для оценки стоимости вычислений обычно используется количество операций над числами с плавающей запятой (то есть такого представления числа, когда его порядок и мантисса записаны в разных ячейках) в одну секунду — FLOPS (floating point operations per second). Гигафлопс (GFLOPS) — это миллиард флопсов.

сеанс *предобучения*, когда AlphaGo просто играла сама с собой, имея в своем распоряжении только правила игры. После того как было сыграно несколько миллионов партий, ей показали партии, сыгранные людьми, что позволило ей создать в процессе обучения алгоритм оценки позиции — она определялась вероятностью прийти от этой позиции к победе. Отдельная нейронная сеть училась предсказывать наиболее вероятный следующий ход для любой позиции.

Поражение Ли Седоля в этом матче поразило многих. По этому поводу высказались разнообразные корифеи искусственного интеллекта. И Н. Бостром, и Р. Курцвейл признались, что не ожидали столь стремительного прорыва в его развитии. Показательна эволюция суждений самого Ли Седоля на протяжении нескольких месяцев до, во время и сразу после матча.

«Октябрь 2015: “Оценивая нынешний уровень машины… я думаю, что выиграю почти все партии”.

Февраль 2016 года: “Я слышал, что Google DeepMind AI стал на удивление силен и быстро учится, но я убежден, что смогу выиграть хотя бы в этот раз”.

9 марта 2016 года: “Я был очень удивлен, так как совсем не ожидал, что могу проиграть”.

10 марта 2016 года: “У меня нет слов… Я просто в шоке. Должен признать… что третья игра будет для меня нелегкой”.

12 марта 2016 года: “Я чувствовал свое бессилие”».

Этому поединку посвящена обширная литература.

В одном из интервью Ли Седоль сравнил свои ощущения во время матча с надвигающейся стеной, которой ему просто нечего противопоставить.

В течение года после победы над Ли Седолем AlphaGo обыграла еще двадцать лучших мировых игроков в го, не проиграв ни одной партии.

Нельзя сказать, чтобы эти матчи вызвали большой интерес в Европе. Поединки Г. Каспарова с А. Карповым в 1980-е гг. пользовались куда как большим вниманием. Но это совсем не так для стран Востока. По словам бывшего руководителя офиса Google в Китае (вплоть до его перевода в Гонконг) Ли Кайфу, за борьбой Ли Седоля с AlphaGo неотрывно следило как минимум 288 миллионов китайцев. Проигрыш Ли Седоля произвел на них примерно такое же впечатление, как запуск советского спутника в 1957 г. на американцев. Сразу после этого в национальной экономической политике произошел радикальный поворот: после 2016 г. ориентация на AI-технологии стала определяющей. К 2020 г. Китай стал мировой AI-сверхдержавой. Пусть здесь не создавались сами технологии, но здесь было накоплено очень большое количество данных. «Пусть Запад и сумел развести костер глубокого обучения, но наибольшую пользу от тепла, даваемого костром искусственного интеллекта, получит Китай», — считает Ли Кайфу. Время экспертов закончилось, пришло время практического анализа накопленных данных.

А в этом у Китая есть колоссальные преимущества, связанные, в частности, с этическими ограничениями. Неприкосновенность частной жизни для Востока — это, скорее, некоторая абстракция и новейшие буржуазные веяния. Строго говоря, и в Европе приватность — явление относительно недавнее. Любой посетитель Лувра или Зимнего дворца легко может заметить почти полное отсутствие дверей: есть только внешние ворота. Практически вся жизнь двора проходила на глазах придворных. Возможность совершить что-то «за закрытыми дверями» исключалась их отсутствием. Жизнь мещан и крестьян часто проходила в одной комнате, где проводили свое время представители нескольких поколений: дети, родители, родители родителей... Китайские хутонги или сыхэюани — по сути дела,

это те же коммуналки, только под открытым небом. Несколько семей пользуются общим туалетом и кухней, живут фактически на глазах друг у друга. Культурная революция 1960–70-х гг. препятствовала проникновению идеи приватности в китайское общество. Не надо удивляться, когда представители восточных культур неожиданным образом разрешают этические дилеммы и не видят проблемы там, где европейцы находят неразрешимое противоречие. И надо быть готовыми к тому, что по мере создания новых интеллектуальных гаджетов с нейронными сетями, обученными на восточных образцах, этические нормы восточных культур будут оказывать все большее влияние на цивилизации Запада.

Ключевые понятия: автономное оружие, валидация и верификация, приватность, флопс, этическая дилемма, дилемма жирафа, дилемма заключенного, дилемма вагонетки.

Контрольные вопросы

1. Чем обосновывается решение об использовании «интеллектуальных» технологий в области вооружений различных стран?
2. В чем разница между верификацией и валидацией используемой технологии?
3. В чем заключается дилемма вагонетки и почему она оказалась самой популярной этической дилеммой в среде разработчиков AI?
4. Какие события повлияли на превращение Китая в AI-сверхдержаву?

Практико-ориентированные задания

1. Проведите сравнение особенностей использования искусственного интеллекта во время матчей компьютера

с Гарри Каспаровым в 1996 г. и Ли Седолем в 2016 г. В чем сходство? В чем различия? Сравните последствия.

2. Найдите и проанализируйте примеры этических дилемм, не упомянутых в этой главе. Опишите, как бы вы предложили искусственному интеллекту на них реагировать.

Темы сообщений, докладов и эссе

1. Дилемма заключенного и холодная война.
2. Беспилотный транспорт в общем потоке: варианты решений, когда авария неизбежна.
3. Где установить границу личного пространства виртуальному помощнику?

ГЛАВА 6.
ПРАВОСУБЪЕКТНОСТЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В результате изучения материалов главы обучающийся должен

знать:

– действующие юридические документы в сфере искусственного интеллекта и научную литературу по вопросу определения места искусственного интеллекта в структуре современных правоотношений;

– основные доктрины, понятия, категории в сфере определения правосубъектности различных видов AI;

уметь:

– определять место AI в структуре современных правоотношений;

– давать правовую оценку фактическим обстоятельствам дела и устанавливать правовые нормы в области взаимодействия социума с искусственным интеллектом в зависимости от правового поля отдельных государств;

владеть навыками:

– анализа различных явлений, фактов, правовых норм и правовых отношений в сфере правоотношений с AI;

– юридически грамотной квалификации отношений с AI.

6.1. Понятие правосубъектности в общей теории права

Правоспособность означает установленную законом способность лица или организации быть носителем

субъективных прав и юридических обязанностей. Она выступает в качестве первоначального условия, общей предпосылки к участию в правоотношениях. Наличие правоспособности означает наличие юридической возможности у лиц своими действиями порождать субъективные права и юридические обязанности.

В правовой теории и на практике различают *три основных вида правоспособности.*

1. *Общая правоспособность* — это способность любого лица или организации быть субъектом права как такового, вообще. Существуют различия между правоспособностью граждан (физических лиц) и юридических лиц. Это отличие связано преимущественно с моментом возникновения правоспособности. Для граждан она возникает с момента рождения; правоспособность юридических лиц неотделима во времени от дееспособности и возникает с момента регистрации устава юридического лица.

2. *Отраслевая правоспособность* означает юридическую способность лица или организации быть субъектом той или иной отрасли права. В каждой отрасли права сроки ее наступления могут быть неодинаковы.

3. *Специальная правоспособность* — способность быть участником правоотношений, возникающих в связи с занятием определенных должностей (президент, судья, член парламента) или принадлежностью лица к определенным категориям субъектов права (работники ряда транспортных средств, правоохранительных органов и др.). Возникновение специальной правоспособности всегда требует выполнения особых условий.

В настоящее время правоспособность во всех странах рассматривается как всеобщий принцип, распространяется на всех граждан. Определенные ограничения устанавливаются лишь в отношении дееспособности граждан и организаций.

Правоспособность — это способность лица иметь юридические права и обязанности. Существуют различия между правоспособностью граждан (физических лиц) и юридических лиц. Это отличие связано преимущественно с моментом возникновения правоспособности. Для граждан она возникает с момента рождения; правоспособность юридических лиц неотделима во времени от дееспособности и возникает с момента регистрации устава юридического лица.

Под *дееспособностью субъекта* понимается его способность своими действиями приобретать и осуществлять субъективные права и юридические обязанности. Существуют следующие виды дееспособности:

1) *полная дееспособность* — при достижении указанного в законе возраста;

2) *дееспособность малолетних* в возрасте от 6 до 14 лет закрепляет ст. 28 ГК РФ, где указано, что эти несовершеннолетние вправе совершать мелкие бытовые сделки (например, приобретать канцелярские принадлежности, продукты питания и др.), направленные на безвозмездное извлечение выгоды, которые не подлежат нотариальному удостоверению или государственной регистрации. Размер этой сделки в законодательстве не закреплен. Особенностью правового положения малолетних является то, что они *полностью неделиктоспособны,* то есть имущественную ответственность по всем сделкам несут их родители, опекуны (законные представители);

3) *дееспособность несовершеннолетних* в возрасте от 14 до 18 лет. Как и малолетние, они вправе совершать только мелкие бытовые сделки. Несовершеннолетние в этом возрасте вправе распоряжаться своим заработком, стипендией, осуществлять авторские и патентные права, вносить вклады в кредитные организации.

А также в возрасте 16 лет несовершеннолетний может быть объявлен *эмансипированным*, то есть полностью дееспособным; это производится по решению органов опеки и попечительства с согласия родителей (ст. 27 ГК РФ), а при отсутствии такого согласия — по решению суда. Несовершеннолетние от 14 до 18 лет *частично деликтоспособны* — сами отвечают за причиненный вред по совершенным им сделкам. В случае недостаточности средств для уплаты за причиненный вред, ответственность несут их родители или попечители.

Однако не все граждане могут обладать полной дееспособностью, это связано не только с их возрастом, но и состоянием здоровья и других обстоятельств. Ввиду этого различают *недееспособных граждан* и тех, у кого *дееспособность ограничивают по решению суда*.

Основанием для *ограничения дееспособности* гражданина является злоупотребление спиртными напитками или наркотическими средствами, если при этом он ставит семью в тяжелое материальное положение. Такому гражданину назначается попечитель. Ограничение полной дееспособности возможно только по решению суда. В этом случае гражданин вправе совершать самостоятельно только мелкие бытовые сделки. Однако получать стипендию, заработок, пенсию он может только с согласия попечителя.

Если гражданин прекращает злоупотреблять спиртными напитками или наркотическими средствами или перестает существовать его семья (в случае расторжения брака, смерти или др. причин), то есть когда отпала обязанность выделять средства на ее содержание, суд отменяет решение об ограничении дееспособности лица.

В конституционном праве лица, находящиеся в местах заключения, ограничены в политических правах.

Гражданин может быть признан *недееспособным* по решению суда, если он не может понимать значения своих действий или руководить ими вследствие психического расстройства. Суд выносит такое решение на основании заключения судебно-психиатрической экспертизы и назначает ему опекуна, который будет представлять интересы своего подопечного (совершать сделки). Если улучшается состояние здоровья гражданина, признанного недееспособным, суд может признать его дееспособным и отменить установленную опеку. При рассмотрении в судах таких дел обязательно должен участвовать прокурор и представители органа опеки и попечительства.

Опека и попечительство устанавливается для защиты гражданских прав и интересов не полностью дееспособных граждан и (или) недееспособных. Опека (попечительство) устанавливается по месту жительства подопечного лица органами местного самоуправления в течение месяца с момента получения решения суда о назначении опекунов или попечителей.

Опека устанавливается также над несовершеннолетними в возрасте до 14 лет (малолетними); гражданами, признанными судом недееспособными. Для защиты интересов несовершеннолетних в возрасте от 14 до 18 лет, а также граждан, ограниченных судом в дееспособности, устанавливается попечительство. При назначении органами опеки и попечительства представителей этим лицам принимается во внимание личные качества опекуна (попечителя), его желание, отношения, существующие между ним и подопечным лицом.

Опека (попечительство) прекращается при достижении 18 лет или по решению суда о восстановлении полной дееспособности.

6.2. Правосубъектность искусственного интеллекта

Сегодня перед Россией стоит глобальная задача создать систему нормативного регулирования как информационного общества, основой которого является цифровая экономика, так и общества знания. Стоит ли их разграничивать или необходимы единые правовые категории и институты? Прежде чем ответить на данный вопрос, необходимо определить новый субъектный состав таких возникающих общественных отношений, необходимость закрепления за ними правового статуса. На современном этапе социально-экономических отношений количество субъектов социальных, а следовательно и правовых, отношений постоянно расширяется. Кроме известных и признаваемых в юриспруденции физических и юридических лиц, а также таких особых коллективных субъектов права, как государство, субъект федерации, народ и др., в законодательстве ряда стран к числу субъектов права в отдельных сферах жизни общества уже сегодня причислены Пачамама (Мать-Земля), животные и растения, а также роботы и киберфизические системы на основе искусственного интеллекта[7].

[7] Закон Южной Кореи «О содействии развитию и распространению умных роботов» (2008); “France Robots Initiatives” («Инициативы Франции в сфере робототехники») (2013); «Азиломарские принципы искусственного интеллекта» (США, 2017); Восьмой закон ФРГ о внесении изменений в Закон о дорожном движении (2017); Закон Эстонии о роботах-курьерах (2017); Резолюция “Civil Law Rules on Robotics” (2015/2103(INL)) («Нормы гражданского права о робототехнике и Хартия робототехники» (2017.06)); План развития технологий ис-кусственного интеллекта нового поколения Китая (2017) и др.

Как отмечают Т.Я. Хабриева и Н.Н. Черногор (2018), «в сфере правового регулирования появляются отношения, в которых если не субъектом, то как минимум участником становится новая цифровая личность — робот. В связи с этим на повестке дня остро стоит вопрос о новых подходах к правовому регулированию общественных отношений с участием роботов, юридическому оформлению в цифровую эпоху правосубъектности как типичных (физических и юридических лиц, государства и др.), так и нетипичных (роботов, а также информационных посредников, таких как провайдеры, блогеры и т. п.) субъектов и участников правоотношений». Перед российским законодателем стоит задача определить статус таких новых субъектов права. Кто они: цифровая личность, электронное лицо, объект или субъект права? По каким критериям нам предстоит его определить? Рассмотрим существующие сегодня подходы к определению их правового статуса.

Одной из первых в юридической науке появляется точка зрения об **определении роботов как объектов правовых отношений по аналогии с животными,** основанием чего следует считать отсутствие у тех и других возможности возложения на них юридической ответственности. Е.Ф. Евсеев (2009) на основании того, что у животных есть собственная воля в совершении действий, в статье «О соотношении понятий “животное” и “вещь” в гражданском праве» предложил определять их как «одушевленная вещь». Подобный подход может быть отождествлен к определению правового положения раба в Древнем мире или холопа в Древнерусском государстве, однако на современном этапе развития права подобный статус вызывает серьезные нарекания.

И В.В. Архипов и В.Б. Наумов (2017) привели возражения против такой постановки вопроса, обосновывая

это тем, что роботы не являются одушевленными существами, поэтому «могут рассматриваться как особый вид имущества — имущества, способного к автономным действиям». При этом особое значение приобретает вопрос о наличии воли и возможности осуществления волевых действий у искусственного интеллекта. Даже если предположить их существование, то возникает вопрос: чья эта воля? Программиста, владельца этих роботов или их самих, способных самостоятельно отбирать информацию из виртуальных сетей и на основе ее анализа и синтеза принимать конкретные решения?

Мы можем соотнести искусственный интеллект с юридическим лицом, как **«искусственно сконструированным субъектом права»**, но не в физической, а в виртуальной реальности. Э. Кастронова (цит. по: Архипова, 2017) относит понятие «интеррации» (создание определенного статуса в виртуальной реальности) с процессом создания юридического лица в правовой реальности. «Юридический акт создает фиктивное лицо (иначе говоря, юридическое лицо. — *А.П.*)... Аналогичный юридический акт интеррации имел бы схожую цель: создание фиктивного пространства... определенного уставом интеррации синтетического мира. Такой устав мог бы... прояснить юридический статус событий, происходящих в таком мире, и имущества, которое в нем накапливается... мог бы определить права людей, выступающих в различных ролях, таких как разработчики, пользователи и те, кто находится за пределами такого мира». Поэтому отдельные виды роботов, а также некоторые иные аппаратные реализации искусственного интеллекта могут быть внесены в особые реестры наподобие реестров юридических лиц как искусственно созданные человеком субъекты права.

Современное интернет-пространство, вхождение человека в виртуальный мир предполагает определенную

дихотомию правового положения индивида, заключающуюся в существовании индивида в гибридной реальности. Ведь создающий в виртуальной цифровой реальности собственного аватара индивид предстает одновременно в двух правовых ипостасях: 1) физического лица в национальном или международном правовом пространстве; 2) цифровой личности — искусственно созданного лица как в материальном, так и «синтетическом» мире. Технологизация «человеческой» природы, в том числе с помощью НБИКС-технологий, представляющих собой конвергентное взаимодействие нано-, био-, информационных, когнитивных, социальных технологий, которые, безусловно, изменяют и правовой статус индивида.

Как отмечает В.В. Чеклецов (2013), по мере воздействия технологий на тело границы живого и неживого «размываются», возникает «философская рефлексия растущей тотальной межсвязности, панкоммуникации, техносоцио-культурного размытия границ между цифровым и “материальным” бытием, когда артефакты обретают память, среда учится чувствовать, а материя становится по-настоящему разумной и программируемой». Происходит не только «оразумнивание сред за счет обретения элементами среды цифровой индивидуальности (RFID-метки, коды), памяти (RFID, проникающий компьютинг), вычислительных, перцептивных, коммуникативных свойств (сети беспроводных сенсоров, сопряженных с Интернетом)», но и «персонализация сред — за счет роста способности элементов среды “узнавать” субъекта (распознавание образов, RFID-биочипы, сенсоры, биоидентификация, GPS, геотаргетинг и т. д.)». Подобные явления означают возможность определения AI и отдельных видов роботов как особого рода цифровых личностей.

В настоящее время широко обсуждается научная позиция о возможности **определения отдельных видов**

роботов как квазисубъектов права. Так, Д.С. Гришин, предложивший законопроект ФЗ «О внесении изменений в Гражданский кодекс Российской Федерации в части совершенствования правового регулирования отношений в области робототехники», предлагает ввести легальное определение *робота-агента*, под которым следует признавать «робота, который по решению собственника и в силу конструктивных особенностей предназначен для участия в гражданском обороте. Робот-агент имеет обособленное имущество и отвечает им по своим обязательствам, может от своего имени приобретать и осуществлять гражданские права и нести гражданские обязанности. В случаях, установленных законом, робот-агент может выступать в качестве участника гражданского процесса»[8].

Правоспособность робота-агента возникает только при регистрации модели такого робота в специально созданном едином государственном реестре роботов-агентов и с момента публичного заявления его собственника о начале его функционирования в таком статусе. При этом ответственность за действия робота-агента в пределах находящегося в их собственности имущества, переданного во владение и (или) пользование робота-агента, несут собственник и владелец робота-агента. Г.А. Гаджиев отмечает, что определение правового статуса «робота-агента можно будет в будущем, когда возникнут реальные предпосылки наличия у них интеллекта, то есть сознания и воли в их юридической, а не психологической интерпретации, признать “как бы субъектами права” (квазисубъектами)».

---

[8] Команда Dentons совместно с Дмитрием Гришиным, основателем Grishin Robotics, подготовила концепцию законопроекта, который может стать первым в мире полноценным законом о роботах. URL: https://www.dentons.com/ru/insights/alerts/2017/january/27/dentons-develops-first-robotics-draft-law-in-russia (дата обращения 30.05.2021).

Европейский союз предлагает **определять промышленных роботов как электронных личностей**, труд которых используется работодателями в различных областях жизнедеятельности человеческого общества, способных помогать человеку в его пути к обществу всеобщего благоденствия. Поэтому их следует определять как участников трудовых и налоговых правоотношений, и ЕС ставит вопрос об их налогообложении наряду с физическими и юридическими лицами. При этом авторы *резолюции "Civil Law Rules on Robotics" (2015/2103(INL)) («Нормы гражданского права о робототехнике и Хартия робототехники» (2017.06))* (далее по тексту — Резолюция ЕС) уделяют особое внимание необходимости создания правовых норм, определяющих ответственность третьих лиц (производителей, операторов и др.) за действие или бездействие роботов (п. А и В), объясняя это все большей автономностью их деятельности.

В п. 1 «Общих положений, касающихся развития робототехники и искусственного интеллекта для гражданских нужд» отмечается, что необходимо, прежде чем определять, являются ли различные объекты робототехники и искусственного интеллекта субъектами права, определить существенные черты таких явлений, как «киберфизические системы», «автономные системы», «умные автономные роботы», и их составляющих.

Как мы видим, в большинстве своем предложения о нормативном регулировании роботов, робототехники, киберфизических систем исходят из того, что AI находится в плоскости цифровой экономики и цифрового общества как нового этапа научно-технической революции. А где же его роль в развитии общества знаний, которое еще в 2005 г. во *Всемирном Докладе ЮНЕСКО «К обществам знаний»* ученые из разных стран призвали как можно быстрее переходить от информационного

общества? На Востоке с древних времен особое значение придавалось нравственному развитию человека. Знание постулатов конфуцианства и наизусть книги «Лунь Юй», содержащей морально-нравственные догмы жизни человека в семье, обществе и государстве, является обязательным и на современном этапе для каждого китайца. И не случайно в принятом в 2017 г. *«Плане развития технологий искусственного интеллекта нового поколения»* Китая отмечается, что «искусственный интеллект несет в себе возможность лучшего устройства общества». Ведь «технологии искусственного интеллекта позволят с высокой точностью выявлять потенциальные угрозы общественной безопасности, вовремя предупреждать о таких угрозах и оперативно принимать необходимые меры. Это позволит улучшить жизнь людей и повысить уровень социальной стабильности».

Показательно, что AI, способный к обучению (умные роботы, умные транспортные средства, средства связи и др.), к 2030 г. должен стать основанием для создания «умного общества», «умной экономики» («умной» промышленности, сельского хозяйства, финансов, бизнеса, кампаний и др.), «умного» образования и «умной» медицины, «умного» правительства и «мудрого» суда, которые напрямую ассоциируются с обществом знания, где главными действующими останутся люди и их объединения, в то время как в цифровом обществе AI может быть также представлен в роли субъекта наряду с физическими и юридическими лицами.

Знание представляет собой оценочную, морально-нравственную категорию, в то время как информация с этой точки зрения нейтральна, поэтому овладение подлинным знанием может быть только человеком, а получать информацию и ее анализировать подвластно и «бездушной» машине. Именно поэтому главной задачей

правового регулирования AI является обязательное установление этических норм и юридической ответственности. Это признается и в Западной Европе, ведь категорический императив И. Канта «будь лицом и уважай других в качестве лиц» является основой для существования современного западноевропейского социума, его либерального толерантного правосознания.

Так, п. 10–14 Резолюции ЕС напрямую посвящены необходимости учета этических принципов в процессе создания, программирования и дальнейшего взаимодействия людей и роботов в процессе деятельности. В соответствии с Этическим кодексом разработчиков робототехники (приложение Резолюции ЕС) лица, спонсирующие развитие искусственного интеллекта, должны заранее предусмотреть возможные риски по отношению к человеку, его жизни, здоровью и т. д. Исследователям данной области предлагается руководствоваться такими принципами, как «не навреди»; «делай благо»; принцип самостоятельности в принятии индивидом решения о сотрудничестве с объектами робототехники; принцип справедливости при распределении социальных благ, создаваемыми роботами внутри социума. Поэтому исследования и разработки в области сверхинтеллекта должны осуществляться в соответствии с правами человека в целях его благосостояния и самоопределения, как отдельного индивида, так и общества в целом, в целях достижения максимальной пользы при минимальном вреде. И главное -- запрещается использовать роботов в любых целях, противоречащих морально-этическим и правовым нормам и стандартам, так как человек — это хрупкое создание как физически, так и психологически.

Развитие технологий AI находится сейчас на таком этапе, что позволяет говорить о возможности его широкого использования в частноправовых и публично-правовых

отношениях, что актуализирует необходимость скорейшего создания принципиально новой нормативной правовой базы и внесения соответствующих изменений практически во все отрасли российского законодательства. Уже сегодня есть гуманоид Фран Пеппер и гиноид София с удостоверениями личности, обладающие определенными правами и свободами наряду с человеком. Судя по скорости развития цифровых технологий, в ближайшем будущем каждый из нас сможет использовать виртуальных помощников, способных от нашего имени совершать действия правового характера. Именно поэтому жизненно необходимым представляется цель адаптации всех правоприменительных и правоохранительных систем.

6.3. Юридическая ответственность: возможность ее существования у искусственного интеллекта[9]

Несмотря на обширную юридическую и философскую литературу, посвященную ответственности, относительно немногие ученые сосредоточили свое внимание на фундаментальной роли ответственности для индивидов и для общества. **Ответственность** может возникнуть, во-первых, при первоначальном решении о разработке и внедрении технологических нововведений, а во-вторых, в их конечных точках, то есть при реальном использовании

[9] Параграф написан на основе научных статей: *Горохова, С.С.* О некоторых аспектах публич-ной юридической ответственности в сфере использования искусственного интеллекта и ав-тономных роботов // Юридические исследования. 2021. № 5. С. 24–41; *Горохова, С.С.* Теоретические подходы к публичной юридической ответственности в сфере использования ис-кусственных интеллектуальных систем // Современный юрист. 2021. № 2 (35). С. 23–31.

технических результатов. Здесь лежат ключевые моменты человеческой свободы, а значит, и ее ответственного использования. Понятие и сущность ответственности заключает в себе определенный образ человека, а именно образ живого существа, обладающего сознанием, обладающего автономией и свободой воли. Наиболее полно и широко то, что делает человека человеком, выражается понятием автономии.

Отметим, что, хотя в научном мире нет единого толкования понятия **юридической ответственности**, тем не менее виды юридической ответственности имеют статус базовых государственно-правовых институтов, входящих в правовую систему любого государства независимо от принадлежности национальной правовой системы к той или иной правовой семье. Юридическая ответственность, неотъемлемая часть любой правовой конструкции, направленной на регулирование общественных отношений. Поскольку там, где есть право, должна быть обязанность, направленная на реализацию этого права, или же запрет, предназначенный защитить законное право или интерес от преступных посягательств. Соответственно, должны быть установлены правила, регламентирующие реализацию прав и выполнение обязанностей. И, в конце концов, необходима ответственность за несоблюдение правил, невыполнение обязанностей и нарушение запретов. Именно так работает процесс правового регулирования, и только ответственность в конечном счете гарантирует нормальный ход человеческих отношений, урегулированных правовыми нормами.

Для того чтобы наступление правовой ответственности стало возможным, необходимо, чтобы было нарушено правило, не выполнена обязанность или же не соблюден запрет, то есть лицо (физическое или юридическое — в зависимости от того о каком виде ответственности

мы говорим) совершило деяние, в котором бы усматривались все признаки правонарушения или его наиболее общественно-опасной формы — преступления. Однако в любом случае правонарушение состоит из объективной и субъективной стороны его состава. При этом, под составом правонарушения понимают совокупность его объективных и субъективных элементов, необходимых и достаточных для того, чтобы квалифицировать деяние как противоправное.

Наиболее существенным в процессе квалификации является то, что при отсутствии хотя бы одного из предусмотренных составом правонарушения элементов и признаков, привлечение субъекта к ответственности делается невозможным. Только полный реализованный правонарушителем юридический состав может породить правовую ответственность.

Можно выделить несколько основных видов юридической ответственности: *материальную; дисциплинарную; гражданско-правовую; административную; уголовную*. Каждый из вышеперечисленных пунктов заслуживает детального рассмотрения, однако в рамках нашего исследования особое значение имеют два последних вида (административная и уголовная), поскольку мы рассматриваем публично-правовую ответственность в области использования искусственного интеллекта, роботов и объектов робототехники.

Почему были отмечены два этих вида? Да потому, что именно они наиболее полно попадают под определение публично-правовой ответственности в праве. Так, Конституционный суд Российской Федерации (далее по тексту — Конституционный суд РФ, КС РФ) исходит из того, что публичная ответственность — разновидность именно юридической, а не социальной ответственности. При этом, в отличие от авторов, отрицающих ее самостоятельный

характер, Конституционный суд России рассматривает **публичную ответственность** как особый вид юридической ответственности, являющий собой форму государственного принуждения *(Постановление КС РФ от 23 сентября 2014 г. № 24-П)*.

Правом реализации публичной ответственности (правом публичного преследования) наделяется только государство в лице его компетентных органов, а основанием для ее применения становится нарушение, которое носит противоправный и общественно опасный характер. Обязательным элементом состава публично-правового нарушения является вина правонарушителя *(Постановление Конституционного суда РФ от 23 сентября 2014 г. № 24-П)*, при этом формами проявления публично-правовой ответственности КС РФ называет административную и уголовную ответственность (например, *Постановление КС РФ от 23 сентября 2014 г. № 24-П*). При этом публично-правовой ответственности свойственны общие, сущностные характеристики, а дифференциация на уголовную и административную — право законодателя, реализуемое в соответствии с требованиями справедливости и соразмерности.

Итак, мы определились с тем, что имеем в виду, говоря о публично-правовой ответственности вообще, теперь рассмотрим данный вид ответственности применительно к сферам использования искусственного интеллекта, роботов и объектов робототехники.

Вообще, следует отметить, что регулирование AI — задача крайне сложная, так как чрезмерное регулирование может привести к охлаждающему эффекту в инновационной деятельности, в то время как недостаточное регулирование способно повлечь за собой серьезный ущерб для прав граждан, а также потерю возможности формировать будущее российского права, призванного

отражать все изменения, происходящие в обществе, в том числе — продиктованные достижениями научно-технического прогресса, поскольку то, как мы подходим к AI, будет в дальнейшем определять мир, в котором мы живем.

Бытовая техника, транспортные средства, медицинское оборудование, дроны и другие продукты все чаще используют AI, и в частности технологии машинного обучения и NLP (обработки естественного языка) для автоматизации принятия решений. Растущая степень автономии, обеспечиваемая AI, имеет много преимуществ, но также порождает неизвестные риски. В частности, что происходит, когда AI развертывается и создает опасность причинения вреда здоровью или даже жизни людей, вызывает материальные или финансовые потери?

Специфические характеристики этих технологий и их приложений, включая сложность, модификацию путем обновления или самообучения в процессе эксплуатации и ограниченную предсказуемость, могут затруднить определение того, что пошло не так и кто должен нести ответственность, если это произойдет. Определение того, кто должен нести ответственность, может быть проблематичным, поскольку в системе AI часто участвует много сторон (поставщик данных, проектировщик, производитель, программист, разработчик, пользователь и сам AI). Дальнейшие осложнения могут возникнуть, если ошибка или дефект возникает из решений, которые система с AI приняла сама на основе принципов машинного обучения с ограниченным или вообще без вмешательства человека.

Сложности, возникающие в результате поступательного развития технологий искусственного интеллекта, и в частности машинного обучения, могут привести к повреждениям, потерям или убыткам в различных контекстах, например следующие.

1. Проблемы владения правом интеллектуальной собственности — способность AI создавать произведения, которые в противном случае были бы признаны объектом интеллектуальной собственности (ИС), созданными человеком, поднимает вопросы о том, кому принадлежит такая ИС и, кроме того, кто несет ответственность, когда такие произведения нарушают право интеллектуальной собственности другой стороны?

2. Угроза конфиденциальности со стороны умных домашних устройств типа Алексы Amazon или Алисы Яндекса, которые предназначены для облегчения жизни. Эти устройства также собирают огромное количество данных (включая персональные данные), которые, если их взломать или скомпрометировать, могут привести к росту претензий в соответствии с законами о конфиденциальности данных.

3. Экономические потери. Предприятия все чаще используют AI для принятия бизнес-решений. Например, в сфере финансовых услуг AI используется для анализа контрактов, принятия инвестиционных решений, возбуждения судебных исков, определения кредитоспособности (см. обсуждение этих вопросов в гл. 4). Если будут допущены ошибки, это может привести к значительным финансовым потерям для бизнеса.

4. Расовая или гендерная дискриминация (см. обсуждение этих вопросов в гл. 4).

Новые вопросы ответственности возникают при использовании AI в беспилотном транспорте или медицинском оборудовании. Отметим, что в настоящее время в Российской Федерации нет системы ответственности, специально применимой к ущербу или убыткам в результате использования новых технологий, в частности AI. Тем более отсутствуют нормы, которые бы регламентировали неимущественную ответственность, связанную

с нарушением гражданских, политических и иных прав граждан.

В качестве исключения можно назвать *Постановление Правительства РФ от 26 ноября 2018 г. № 1415 «О проведении эксперимента по опытной эксплуатации на автомобильных дорогах общего пользования высокоавтоматизированных транспортных средств»*, в п. 18 которого определяется, что ответственность за дорожно-транспортные и иные происшествия на автомобильных дорогах Российской Федерации, произошедшие с участием принадлежащего ему высокоавтоматизированного транспортного средства при проведении эксперимента и при отсутствии виновных действий других участников дорожного движения, приведших к данному дорожно-транспортному или иному происшествию на автомобильной дороге, несет собственник высокоавтоматизированного транспортного средства.

Кроме того, в соответствии с п. 7, собственник должен застраховать и поддерживать застрахованным в период проведения опытной эксплуатации риск ответственности по обязательствам, возникающим вследствие причинения вреда жизни, здоровью или имуществу других лиц в пользу третьих лиц на сумму 10 миллионов рублей в отношении каждого высокоавтоматизированного транспортного средства. По общему же правилу, если транспортное средство принадлежит юридическому лицу, то на нем и лежит ответственность за причиненный вред. Однако юридическое лицо может потребовать частичного возмещения убытков с водителя, если докажет, что причиной ДТП стали его действия. Такой подход в целом соответствует мировой практике. Так, в 2018 г. Великобритания приняла *Статут об автоматизированных транспортных средствах и электромобилях*, в соответствии с которым ответственность за ущерб, причиненный застрахованным

автоматизированным транспортным средством при управлении им самим, лежит на страховщике. В противном случае возмещение ущерба жертвам, понесшим ущерб в результате сбоя AI, скорее всего, будет испрашиваться в соответствии с существующими законами о возмещении ущерба в контракте, законодательством о защите прав потребителей и деликтом халатности.

Отметим, что существующий режим ответственности обеспечивает по крайней мере базовую защиту жертв, ущерб которым причинен в результате применения таких новых технологий. Однако конкретные характеристики и сложности этих технологий и их применения могут затруднить предоставление жертвам компенсации во всех случаях, когда это представляется оправданным, и это может не обеспечить справедливого и эффективного распределения ответственности во всех случаях.

Кроме того, все-таки речь здесь идет по большей части о **гражданско-правовой ответственности**, а нас в первую очередь интересует ответственность публично-правовая. Представим себе ситуацию, когда в результате ДТП при нарушении правил дорожного движения был причинен вред здоровью третьего лица или даже пострадавший в результате этого происшествия погиб. При этом авария произошла не по вине водителя (оператора беспилотного транспортного средства), а по вине, скажем, компании разработчика ПО или же производителя транспортного средства.

В этой ситуации возникает некоторая правовая неопределенность, связанная в первую очередь с определением вида применимой ответственности. Понятно, что транспортное средство было застраховано и пострадавшему (его представителям) в рамках гражданско-правовой ответственности, на сколько это возможно, в денежном эквиваленте ущерб будет компенсирован. Но в некотором

роде здесь бы имел место случай, во-первых, обладающий явными признаками повышенной общественной опасности, так как последствием стал вред здоровью или смерть потерпевшего; во-вторых, мы не могли бы здесь говорить о невиновном причинении вреда, так как подразумеваем, что виновны компания-разработчик или компания-производитель; в-третьих, в ситуации с технологиями AI, полагаем, было бы крайне затруднительно выявить конкретного виновника — физическое лицо в группе разработчиков или же производителей транспортного средства.

То есть, принимая во внимание все эти факторы, усматривается ассиметричный подход к определению ответственности для виновных лиц. Так, если бы виноват был водитель (оператор), то он бы (теоретически) подлежал привлечению к уголовной ответственности по ст. 264 УК РФ, предусматривающей уголовную ответственность за нарушение лицом, управляющим автомобилем, трамваем либо другим механическим транспортным средством, правил дорожного движения или эксплуатации транспортных средств, повлекшее по неосторожности причинение тяжкого вреда здоровью человека. Однако для компаний разработчиков и производителей такая ответственность была бы невозможна по той причине, что юридические лица просто не являются субъектами уголовной ответственности в России. И, кроме того, очевидно, что даже если бы корпоративная уголовная ответственность была бы установлена, то встал бы вопрос о том, что не они управляли транспортным средством и не они нарушили ПДД, которое привело к тяжким последствиям, хотя это и произошло по их вине. Описанный казус очевидным образом указывает на два основных направления развития публично-правовой ответственности в сфере использования искусственного интеллекта, роботов и объектов робототехники.

Исходя из функциональной природы возможной частичной правоспособности AI, ключевой вопрос здесь заключается не в том, должен ли интеллектуальный агент сам нести ответственность, а в том, как выявить лицо, ответственное за действия интеллектуальных агентов. Разумный агент действует от лица своего хозяина, а значит, и причиненный вред должен рассматриваться соответственно. Иными словами, поскольку деликт совершается в рамках развертывания, ответственность должна, как обычно, возлагаться на лицо, получающее прибыль от развертывания. В общем праве очевидным решением было бы применение правила respondeat superior (отвечает старший), которое гласит, что хозяин несет ответственность за действия своих агентов. Например, в Соединенных Штатах существуют обстоятельства, когда работодатель несет ответственность за действия сотрудников, совершенные в ходе их работы.

Однако против такого универсального подхода есть ряд возражений.

Во-первых, трудно определить ответственность за ущерб, причиненный автономным роботом. Обычно повреждение, вызванное автономным роботом, может возникнуть из дефекта механической (программной) части, который означал бы, что производитель недостаточно информировал клиента об опасностях, связанных с автономными роботами или что системы безопасности робота были недостаточны. В данном случае мы сможем проследить ответственность до производителя или разработчика программного обеспечения.

Если робот продается с открытым исходным кодом программного обеспечения, лицом, несущим ответственность, должно в принципе быть тем, кто запрограммировал приложение, которое привело к повреждению робота. Микророботы, как правило, все чаще продаются

с (полным или частичным) открытым исходным кодом программного обеспечения, что позволяет покупателям разрабатывать свои собственные приложения. В принципе договор регулирует отношения между сторонами. Open Robot Hardware — это еще одна тенденция, где как программное, так и аппаратное обеспечение робота допускает практически произвольное дополнение. Если робот причиняет какой-либо ущерб, который может быть прослежен до его проектирования или производства — например, ошибка в алгоритме робота, вызывающая вредное поведение, разработчик или производитель должны нести ответственность. Однако здесь вид ответственности может варьироваться в зависимости от того, купил ли потерпевший робота (договорная ответственность) или является третьим лицом (внедоговорная ответственность).

Во-вторых, ущерб, причиненный автономными роботами, может быть прослежен до ошибки пользователя. Если робот причиняет какой-либо ущерб во время использования или во время обучения, его пользователь или владелец должны нести ответственность. В этом отношении решение может варьироваться в зависимости от того, является ли пользователь профессионалом или жертвой. Например, любой ущерб, связанный с инструкцией робота профессиональным пользователем и нанесенный жертве третьей стороны, влечет ответственность профессионального пользователя. Совершенно другая история, если такой же ущерб был причинен жертве, которая была профессиональным, оплачиваемым пользователем, поскольку тогда это будет считаться несчастным случаем на работе.

В-третьих, не исключен перехват управления и перепрограммирование интеллектуальной информационной системы, повлекшие причинение вреда третьим лицам, здесь, очевидно, ответственность должна быть применена

к злоумышленнику, осуществившему указанные действия.

Тем не менее развитие современных технологий может привести к дальнейшим беспрецедентным трудностям, когда будет сложнее установить, что вызвало причинение вреда в определенных ситуациях, особенно если робот способен к самообучению и принятию автономных решений. Однако говорить о том, что машина может частично или полностью нести ответственность за свои действия или бездействие, просто не имеет смысла.

По нашему мнению, только физическое (или юридическое) лицо должно привлекаться к ответственности с помощью различных правовых механизмов, в том числе и страхования ответственности. Что касается основания для ответственности, то в зависимости от того, о какой именно ответственности идет речь, необходимо будет использовать соответствующие видам юридической ответственности подходы.

По общему правилу следует исходить из принципа вины, ее формы и тяжести наступивших общественно опасных последствий. А для возмещения причиненного вреда жертве в случае, когда виновного с достоверностью определить будет невозможно, необходимо использовать механизмы превентивного страхования ответственности, по принципу страхования ответственности лиц (предприятий), эксплуатирующих источники повышенной опасности (ИПО).

Напомним, что специфика гражданской ответственности владельцев ИПО, то есть граждан и юридических лиц, чья деятельность связана с повышенной опасностью для окружающих (использование транспортных средств, механизмов, электрической энергии высокого напряжения, атомной энергии, взрывчатых веществ, сильнодействующих ядов, осуществление строительной

деятельности и т. п.), заключается в том, что владелец ИПО практически всегда несет ответственность за причиненный таким источником вред, в то время как лицо, не являющееся владельцем ИПО, в большинстве случаев ответственности не несет, если докажет, что вред причинен не по его вине.

6.4. Гражданско-правовая ответственность разработчика (создателя) в области использования искусственного интеллекта, робота и объектов робототехники

В Резолюции ЕС рекомендовано установление правовых норм «об ответственности за качество и безопасность товаров, согласно которым производитель несет ответственность за любые неисправности». Следовательно, разработчик системы (юнита, носителя) искусственного интеллекта, робота и объекта робототехники фактически может быть приравнен к производителю товаров. В существующей системе права указанные носители могут признаваться только в качестве объектов или товаров, то есть могут быть отнесены к объектам гражданских прав с распространением на них положений ГК РФ (ст. 128, 129 и т. д.).

Важным вопросом является установление критериев или технических характеристик, по которым возможно определение качества данных «товаров», то есть необходимо внесение изменений и дополнений в сфере нормативно-технического регулирования, основные положения которых представлены автором на рис. 11.

Техническая регламентация характеристик носителей искусственного интеллекта, роботов и объектов робототехники должна соответствовать уровню развития

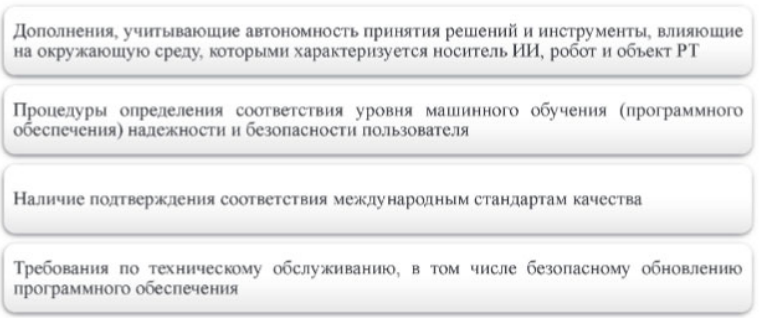

Рис. 11. Основные положения по изменениям и дополнениям в сфере нормативно-технического регулирования

используемых технологий, иначе невозможно их отнесение к надежным и безопасным «продуктам» и осуществление государственного контроля в указанной сфере.

При введении гражданско-правовой ответственности разработчика следует установить гарантии справедливого ее распределения, так как разработчик и пользователь зачастую являются разными субъектами, следовательно, последствия действий пользователя не должны включаться в зону ответственности разработчика.

В системе законодательства Российской Федерации существует обязанность производителя раскрыть информацию по характеристикам товара, а также правилам пользования. В условиях выхода «инновационного продукта», то есть применения новой технологии, актуальность, обязательность и полнота предоставляемой пользователю информации только возрастает.

Таким образом, установление гражданско-правовой ответственности разработчика (создателя) в области использования искусственного интеллекта, робота и объектов робототехники может быть основано на правовых нормах, регламентирующих применение ответственности к производителю. Однако необходимы внесения

изменений и дополнений в нормативно-техническое регулирование, неотъемлемой составляющей которых является условие раскрытия информации по характеристикам товара, а также правилам пользования.

6.5. Гражданско-правовая ответственность пользователя (владельца, собственника или лица, получающего прибыль) в области использования искусственного интеллекта, робота и объектов робототехники

В.А. Лаптев проанализировал в статье «Понятие искусственного интеллекта и юридическая ответственность за его работу» (2019) возможную ответственность владельца. Он исходил из трех основных правомочий собственника, установленных в действующем законодательстве: владение; пользование; распоряжение. Наиболее важной составляющей в данном случае представляется регистрация (фиксирование) владельца, поскольку среди объектов данной технологии могут быть источники повышенной опасности, как, например, автомобили, специальная техника. Кроме того, введение указанной регистрации гарантирует защиту собственника от незаконного изъятия или хищения системы искусственного интеллекта, робота и объекта робототехники. Следует отметить, что предлагаемая регистрация необходима только для объектов, которые относятся к источникам повышенной опасности, либо использование которых требует профессиональных навыков от пользователя (владельца).

При установлении ответственности для владельца следует разграничить понятие «владелец» и «пользователь». Владелец может предоставить право пользования указанными объектами другому лицу, чьи действия,

возможно, станут причиной негативных последствий, поэтому к ответственности необходимо привлечение непосредственно последнего.

Основной составляющей в данном подходе являются возможные негативные последствия (в первую очередь ущерб), вызванные непосредственными действиями (бездействием) пользователя, которые совершены с умыслом или без умысла (по неосторожности) с учетом того факта, имел ли субъект достаточные знания об объекте и вероятности причинения ущерба как следствия его действий.

Следовательно, основными положениями относительно установления гражданско-правовой ответственности для пользователя являются:

- действие (бездействие), совершенное с умыслом на причинение вреда жизни, здоровью (своему или иного лица), окружающей среде и проч., в том числе мошеннические действия, например несанкционированное перепрограммирование или иное внесение изменений, не предусмотренных разработчиком (создателем);
- действие (бездействие), совершенное без умысла ввиду неосторожности или небрежности, в том числе при отсутствии профессиональных навыков для осуществления управления системой искусственного интеллекта, роботом и объектом робототехники, но причинившие вред жизни, здоровью (своему или иного лица), имуществу, окружающей среде и проч.

Ущерб, причиненный действиями (бездействием) лица вследствие нераскрытия разработчиком необходимой информации и (или) предоставления некачественного объекта, соответственно, не должен относиться к ответственности пользователя. Установление указанных положений разграничит гражданско-правовую ответственность пользователя и разработчика, будет

способствовать более ответственному отношению к применению данных объектов со стороны пользователя.

Таким образом, при установлении гражданско-правовой ответственности владельца (лицо, получающее прибыль) в области использования искусственного интеллекта, робота и объектов робототехники основной составляющей является ответственность за возможные негативные последствия (ущерб), вызванные непосредственными действиями (бездействием) пользователя, которые совершены с умыслом или без умысла (по неосторожности).

Ключевые понятия: искусственный интеллект (AI), правовой статус, правоотношение, субъект правоотношения, объект правоотношений, юридическая ответственность, основания и принципы юридической ответственности.

Контрольные вопросы

1. Определите место ИИ в системе современных правоотношений. В чем заключается частичная правоспособность сильного (General AI) или сверхсильном (Super AI) искусственного интеллекта как разумных агентов (квазисубъектов права)?

2. Какова ответственность программистов, разработчиков алгоритмов, правообладателей и владельцев AI? Есть ли презумпция виновности?

3. Раскройте понятие и виды оснований для освобождения от ответственности (ошибка производителя, ошибка пользователя, перехват управления и др.) на основе имеющейся у вас информации.

4. Какие виды юридической ответственности в случае причинения ИИ вреда человеку вы полагаете необходимыми для применения? В какие нормативные правовые акты необходимо внести изменения и какие именно?

Практико-ориентированные задания

1. Сегодня персональная информация (паспортные данные, номера телефонов, СНИЛС и т.д.) является доступной в электронной среде.

Каким образом «научить» искусственный интеллект требованиям информационной безопасности? Предложите вариант информационно-коммуникационных технологий для решения этого вопроса.

2. На основе зарубежной и российской научной литературы постройте таблицу для различных видов ИИ по определению их места в структуре правовых отношений, выявив общие и особенные черты.

Темы докладов и сообщений

1. Проблемы определения объекта правоотношения в сфере использования искусственного интеллекта.
2. Проблемы правосубъектности нейронных сетей.
3. Проблемы правосубъектности роботов.
4. Проблемы правосубъектности объектов робототехники.
5. Проблемы правосубъектности киберфизических систем.

ГЛАВА 7.
ПРАВОВОЕ РЕГУЛИРОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В результате изучения материалов главы обучающийся должен

знать:

– действующие юридические документы в сфере искусственного интеллекта и научную литературу по вопросу определения места AI в структуре современных правоотношений;

– правовые принципы взаимодействия человека и различных видов AI;

уметь:

– определять место AI в структуре современной системы законодательства России и зарубежных стран;

– устанавливать правовые нормы в области взаимодействия социума с искусственным интеллектом в зависимости от правового поля отдельных государств;

владеть навыками:

– анализа различных явлений, фактов, правовых норм и правовых отношений в сфере правоотношений с AI;

– юридически грамотной квалификации отношений с AI.

7.1. Правовые принципы использования искусственного интеллекта

Общественные отношения, существующие на данном этапе научно-технического развития, предопределили

необходимость трансформации правовой составляющей социального взаимодействия естественного (человеческого) и искусственного интеллекта, в первую очередь, путем включения новых, выходящих за рамки создаваемого веками правового тезауруса, терминов, категорий, дефиниций в юридический обиход для характеристики нового экономического уклада цифрового общества. Современные возможности и связанные с ними новые риски и угрозы для государства, общества и отдельных индивидов формируют ситуацию, когда правовое поле как минимум должно подстроиться под воплощаемую реальность, а как максимум создать «задел» для обеспечения дальнейшего соответствия между развивающимися цифровыми технологиями, технологиями искусственного интеллекта и социальными отношениями в целом и правовыми в частности.

Для выработки принципиальных подходов к необходимой правовой трансформации и достижения поставленных целей и задач особое значение имеет теоретико-правовая основа определения места и роли нейронных сетей, киберфизических систем, различного вида роботов и робототехники в системе права и системе законодательства национального и наднационального уровня. Основополагающим принципом правового регулирования, на наш взгляд, должен остаться принцип гуманизма, образующий антропоцентрическую правовую оболочку, которая строится вокруг незыблемости прав и свобод человека, их высшей ценности по отношению к другим (всем) менее значимым категориям. Такой подход, помимо прочего, обосновывался исключительностью человеческого сознания (естественного интеллекта), его превосходством над всеми иными явлениями, объективно существующими в реальном мире, ибо как утверждал софист Протагор: «Человек как мера всех вещей». Отправной точкой всегда

должен оставаться человек, его интересы и благополучие. Однако уже сегодня есть основания полагать, что в течение нескольких десятилетий искусственный интеллект сможет превзойти человеческий. И это поставит перед людьми ряд серьезных вопросов, связанных со способностью контролировать свое собственное творение, реакцией на возможность «ремонта» и улучшения людей, возможностью замены межличностных отношений (дружбы, любви, товарищества) на отношения другого рода, объединяющих творение и творца.

Такие глобальные изменения экспозиции будущего мира, просто нельзя отразить в точечных, неполных и бессистемных изменениях законодательства. Поэтому, формируя фундаментальные основы правой системы с поправкой на существование искусственного интеллекта, необходимо начинать с коррекции (или с подтверждения) философской канвы, с определения парадигмы выстраивания правовых и нравственно-этических основ взаимодействия между искусственным и естественным интеллектами. В рамках достижения поставленной цели для дальнейшего развития искусственного интеллекта, нейронных сетей, роботов и объектов робототехники и их практического применения в отраслях народного хозяйства и различных сферах цифровой экономики, необходимо планомерно реализовывать намеченные задачи, создавая обновленную нормативную правовую базу. В России принят *Федеральный закон от 24 апреля 2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации — городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона "О персональных данных"»* (далее по тексту Федеральный закон

№ 123-ФЗ), в котором с 1 июля 2020 г. в Москве устанавливается экспериментальный правовой режим на 5 лет по внедрению AI в жизнь каждого жителя умного города, установлены цели, задачи и принципы этого экспериментального правового режима. Но в нем, к сожалению, нет статей, регулирующих правовые и этические принципы взаимодействия человека и AI, CPS, роботов и объектов робототехники, что, на наш взгляд, может привести к определенным социальным рискам в будущем в процессе проведения подобного эксперимента.

Правовые принципы представляют собой базис, на основе которого и должно быть создано российское законодательство в области AI. Прежде всего, необходимо определиться о каком именно виде AI идет речь в Федеральном законе № 123-ФЗ. Ведь в современной научной литературе определены семь направлений развития искусственного интеллекта, начиная от способности систематизировать и воспроизводить различные области знания на основе анализа различной информации, используя алгоритмы обучения и самообучения до способности «разумного» общения с другими системами AI и человеком, применяя заложенные поведенческие алгоритмы и интеллектуальное программирование. Развитие в мире робототехники и появление «умных роботов», «социальных роботов», AGI и, возможно, уже в скором будущем сверхсильного AI (SAI) требует от законодателя определенности в формулировании правовых и этических принципов взаимодействия людей, общества и государства в целом и этих видов AI.

9 июня 2020 г. на сайте мэра Москвы mos.ru был опубликован «проект стратегии» под заголовком «Москва “Умный город — 2030”», описывающий предполагающиеся принятым федеральным законом изменения в жизни данного конкретного города. В частности, он описывает использование цифровых технологий и технологий

искусственного интеллекта по следующим шести направлениям (рис. 12).

Рис. 12. Направления проекта цифровой стратегии Москвы «Умный город — 2030»

Как видно из этого рисунка, практически во всех сферах жизни человека и общества будут использоваться технологии AI, причем авторы проекта стратегии прямо апеллируют к идеям зарубежных исследователей-футурологов Рэя Курцвейла, Яна Пирсона, Брайана Дэвида Джонсона, приверженных концепции трансгуманизма и поддерживающих «использование достижений науки и технологии для улучшения умственных и физических возможностей человека, с целью устранения тех аспектов

человеческого существования, которые считаются нежелательными — страданий, болезней, старения и смерти». Для чего предполагается даже вживление в организм человека медицинских устройств. Вопрос, который сразу же возникает: на какие юридические принципы может опираться реализация такого проекта?

В «проекте стратегии» предлагаются не правовые или этические принципы взаимодействия человека и AI (искусственной нейронной сети, робота или какой-то иной CPS), а некие «принципы умного города», к числу которых отнесены (цитируем только заголовки, дословно, с сохранением орфографии): «Принцип 1. Умный город — для человека»; «Принцип 2. Участие жителей в управлении городом»; «Принцип 3. Искусственный интеллект для решения городских задач»; «Принцип 4. Цифровые технологии для создания полноценной безбарьерной среды во всех сферах жизни»; «Принцип 5. Развитие города совместно с бизнесом и научным сообществом на партнерских взаимовыгодных условиях»; «Принцип 6. Главенство цифрового документа над его бумажным аналогом»; «Принцип 7. Сквозные технологии во всех сферах городской жизни»; «Принцип 8. Отечественные решения в сфере цифровых технологий»; «Принцип 9. Зеленые цифровые технологии».

Какие выводы отсюда можно сделать? Во-первых, для авторов стратегии нет разницы между цифровыми технологиями и технологиями искусственного интеллекта (или они умышленно стараются ее нивелировать), поэтому они не выделяют особенности работы нейронных сетей, киберфизических систем, различных видов искусственного интеллекта, их алгоритмов и возможных ошибок, допущенных их разработчиками, которые могут нарушать права и свободы человека и гражданина, способствовать изменению системы взаимодействия

в социуме и механизме государства. А во-вторых, полностью отсутствуют этические принципы применения технологий искусственного интеллекта, о которых идет речь в Федеральном законе № 123-ФЗ.

В данном законе практически нет указаний на цифровые технологии, упоминается исключительно AI, правда непонятно, о каком его виде идет речь. Снова нет речи о правовых принципах использования технологий AI, но зато уже регламентируются принципы специального правового режима, к числу которых законодатель в п. 3 ч. 2 ст. 3 отнес:

«1) прозрачность экспериментального правового режима;

2) защита прав и свобод человека и гражданина, обеспечение безопасности личности, общества и государства;

3) недискриминационный доступ к результатам применения искусственного интеллекта».

Принципы просто перечислены без правовой расшифровки их содержания, что дает возможность широкого толкования данной нормы правоприменителем, то есть высшим исполнительным органом г. Москвы, или нейронной сетью, или ботом на основе AI.

Особую озабоченность вызывает тот факт, что в соответствие со ст. 4: сами технологии искусственного интеллекта и (или) производства, реализации, оборота отдельных товаров (работ, услуг) на основе указанных технологий, а также требования к указанным технологиям и (или) товарам (работам, услугам); все случаи и порядок использования результатов применения искусственного интеллекта; а также случаи обязательного применения и (или) учета результатов применения искусственного интеллекта в деятельности органов исполнительной власти субъекта Российской Федерации — города федерального значения Москвы и подведомственных им организаций;

порядок и случаи передачи собственниками средств и систем фото- и видеонаблюдения изображений, полученных в соответствии с условиями, предусмотренными подп. 1 и 2 п. 1 ст. 152.1 ГК РФ, а также предоставления доступа к таким средствам и системам фото- и видеонаблюдения органам государственной власти и организациям, осуществляющим публичные функции в соответствии с нормативными правовыми актами Российской Федерации; по согласованию с уполномоченным федеральным органом исполнительной власти, осуществляющим функции по выработке и реализации государственной политики и нормативно-правовому регулированию в сфере информационных технологий, порядок и условия обработки участниками экспериментального правового режима персональных данных, полученных в результате обезличивания, на основании соглашений с уполномоченным органом, а также требования к таким соглашениям определяет высший исполнительный орган государственной власти субъекта Российской Федерации — города федерального значения Москвы. Причем в самом тексте Федерального закона № 123-ФЗ нет принципов, которыми должен руководствоваться данный орган исполнительной власти, не прописана и ответственность в случае, если мэрия Москвы станет руководствоваться ошибочным решением, принятым искусственным интеллектом.

Отметим, что на текущем этапе государственно-правового развития, просто невозможно спрогнозировать все возможные последствия включения искусственного интеллекта в правовое поле практически каждой сферы общественной жизни. Нельзя также предсказать, какого уровня достигнет развитие искусственного интеллекта в будущем, по той простой причине, что сейчас его совершенствование ограничено техническими возможностями по обеспечению софтом, однако в случае решения этой

проблемы его развитие, по мнению экспертов, не будет ограничено уже ничем. И в этой связи наиболее важным аспектом в деятельности ученых-юристов становится не формулирование конкретных правовых установлений, дающих определение искусственного интеллекта, включающих его в правовые конструкции и снабжающих узаконенными правами (поскольку все эти положения могут очень быстро устареть, если или когда появится искусственный интеллект, способный к самопознанию, саморазвитию и собственным представлением о своем месте в мире), а формулирование правильных принципов правового регулирования, сопровождаемое надеждой, что и искусственный интеллект будет ими руководствоваться и соблюдать в той же мере, что и человечество.

Исходя из вышеизложенного, следует уделить особое внимание принципам правового регулирования исследуемой тематики. На наш взгляд, в число этих принципов (помимо общеправовых, таких как гуманизм, законность, запрет дискриминации и стигматизации) должны войти основные начала правового регулирования, способные задать правильный вектор выстраиванию взаимодействия между двумя видами интеллектов. Ведь правовое обеспечение общеправовых принципов происходит через межотраслевые и отраслевые принципы правового регулирования, к системе которых следует отнести следующие.

1. Принцип недопустимости причинения вреда человеку, сформулированный еще в 1943 г. Айзеком Азимовым. Интересным фактом здесь является то, что все эти законы дословно приведены со ссылкой на источник в пункте (M) введения к докладу Европарламента с рекомендациями Комиссии по нормам гражданского права в области робототехники: «(1) робот не может причинить вред человеку или своим бездействием допустить причинение вреда человеку; (2) робот должен подчиняться

приказам, данным ему людьми, за исключением случаев, когда такие приказы противоречат Первому Закону; (3) робот должен защищать свое существование до тех пор, пока такая защита не противоречит Первому или Второму Законам, и (0) робот не может причинить вред человечеству или бездействием позволить человечеству прийти к вреду».

По мнению писателя, эти законы должны быть заложены изначально в программу искусственного интеллекта как аналог нравственных ценностей, «категорического императива» для индивида. Не случайно именно эти законы стали основой для *Кодекса этики для разработчиков робототехники, являющегося приложением к Резолюции Европарламента от 16 февраля 2017 г. 2015/2013(INL) P8_TA-PROV(2017)0051*. Однако в данном случае следует учитывать то обстоятельство, что системы AI уже нашли свое применение не только в гражданской, но и в военной сфере, что накладывает определенные ограничения на применимость указанного принципа. На сегодняшний момент есть примеры причинения вреда человеку[10].

2. Принципы уважения человеческого достоинства и конфиденциальности, обусловленные тем, что частная информация о человеке может быть раскрыта через решения и прогнозы, сделанные AI, более того, от применения некоторых возможностей технологий на основе AI может

[10] В 2017 г. корпорация Amazon (США) была вынуждена закрыть экспериментальный проект по найму сотрудников с использованием искусственного интеллекта, так как алгоритм занижал оценки кандидатов-женщин, поскольку был обучен на прошлом десятилетнем опыте отбора кандидатов в Amazon, среди которых преобладали мужчины. В 2016 г. была прекращена работа чат-бота Tay за расизм в рекламе Facebook, так как бот на основе высказываний пользователей Twitter менее чем за сутки научился расистским высказываниям.

пострадать и человеческое достоинство, так как AI позволяет все более реалистичные фото, аудио и видео подделки или «глубокие подделки», которые могут быть использованы для дискредитации граждан. Так, известна ошибка, допущенная AI в Китае, когда в числе злостных нарушителей правил дорожного движения была названа Дун Минчжу (董明珠, входит в первую сотню самых влиятельных женщин Китая по версии Forbes) просто из-за того, что ее портрет был размещен в рекламных целях на автобусе[11].

Кредитные организации уже используют интеллектуальные системы для прогнозирования кредитного риска при выдаче кредитов, и некоторые государства (например, США) пропускают разнообразные характеристики заключенных через сложные алгоритмы интеллектуальных программ, чтобы предсказать вероятность рецидива при рассмотрении вопроса об условно-досрочном освобождении. В этих случаях очень важно обеспечить, чтобы такие факторы, как национальность, раса и сексуальная ориентация, не использовались для принятия решений на основе AI. Даже когда такие функции не включены непосредственно алгоритмами программы, они все равно могут сильно коррелировать с кажущимися безобидными сведениями, такими как почтовый индекс или адрес. Вместе с тем при тщательном проектировании, тестировании и развертывании алгоритмы AI могут принимать менее предвзятые решения, чем типичный человек.

3. Презумпция невиновности человека при решении AI, CPS или нейронной системой вопросов, связанных

[11] В 2016 г. роботом на основании фото человека, имеющего ярко выраженную азиатскую внешность, было отказано в приеме документов на получение паспорта Новой Зеландии, так как алгоритм посчитал глаза закрытыми.

с правомерным поведением человека (приложение «Социальный мониторинг» в г. Москве в период пандемических мероприятий однозначно исходит из принципа виновности больного COVID-19, невзирая на то, что, например, возможен сбой в системе GPS или человек может спать и не слышать сигнала о необходимости прислать селфи, или камеры наружного наблюдения распознают лицо гражданина примерно на 52 % и в этом случае он уже считается нарушившим режим и обязан доказать свою невиновность и т. д.).

4. Принцип раскрытия информации о разработке, производстве и использовании роботов и искусственного интеллекта, смысл которого заключается в том, что разработчики, производители и участники хозяйственного оборота обязаны будут раскрывать информацию о количестве «умных роботов», которых они используют, об экономии средств, вносимых на социальное обеспечение за счет использования робототехники вместо человеческого персонала, об оценке суммы и доли доходов предприятия, полученных в результате использования робототехники и искусственного интеллекта. Особенно важным данный принцип представляется для защиты прав и свобод физических лиц в случае принятия решения на основе данных нейронных сетей, киберфизических систем, AI или использования ими персональных данных. Только таким образом можно создать атмосферу доверия к ИС, которая основана на принципах надежности, справедливости, безопасности, объясняемости и точной информации о всех составляющих ИС, то есть создание объяснимого или прозрачного AI (Explainable AI, XAI).

5. Принцип автономности воли, по нашему убеждению, должен иметь отношение не только к гражданам, у которых должно быть гарантированное легальное

право принимать обоснованное, не принуждаемое решение об условиях взаимодействия с AI, CPS, роботами или объектами робототехники, но и у разработчиков алгоритмов и программистов подобных ИС и их владельцев. При этом последние, с одной стороны, должны быть лишены предвзятости, а с другой — иметь возможность учесть в создаваемых алгоритмах и программах автономию воли при использовании методов машинного обучения (ML) различных моделей нейронных сетей — создание автоматизированного машинного оборудования (AutoML). При этом до внедрения в практику многозадачных моделей следует законодательно зафиксировать запрет на использование отдельных моделей нейронных сетей, созданных с определенной целью, для решения иных задач (например, использовать сверточные сети для создания изображений или порождающие сети для моделирования определенных последовательностей).

6. Принцип информированного согласия / презумпция несогласия, тесно связанный с предыдущим принципом, охватывает все возможные преимущества и последствия взаимодействия с искусственным интеллектом, начиная от медицины, образования и заканчивая сферой развлечений, применением геймификации при обучении и др. Граждане должны иметь возможность отказаться от использования интеллектуальных технологий, как в публичной, так и в частной сфере. Простого публичного уведомления об их использовании явно будет мало.

Особенного внимания требует внедрение и использование технологий, связанных с распознаванием эмоций и призванных определять такие аспекты, как испытываемые индивидом чувства, состояние его психического здоровья, «вовлеченность» сотрудников (студентов) в трудовой (образовательный) процесс. Такие технологии

основаны на принципе профайлинга[12] и системе кодирования лицевых движений (FACS) П. Экмана и У. Фризена, особую эффективность имеют сверточные нейронные сети, распознающие эмоции по голосу диктора. Пока точность таких систем составляет около 61–73 %, что, безусловно, дает возможность интерпретировать их результаты слишком вольно, а следовательно, свидетельствует о принципиальной необходимости легального закрепления принципа информированного согласия субъектов права.

7. Принцип справедливости — принцип, предусматривающий справедливое распределение преимуществ, связанных с использованием AI, робототехники и доступностью роботов. Правовое регулирование должно оцениваться с точки зрения того, способствуют ли они демократическому развитию и справедливому распределению благ AI или концентрируют власть и выгоды в определенных кругах. Особое внимание следует уделить трактовке понятий «благо человека», «благо социума». А поскольку будущие технологии искусственного интеллекта и их последствия невозможно предвидеть с полной ясностью, правовую политику необходимо будет постоянно пересматривать в контексте наблюдаемых социальных проблем и данных, полученных в результате полевых исследований. Полагаем, с течением времени при внедрении систем сверхсильного AI будет дополняться и перечень правовых принципов, которые лягут в основу правового регулирования сферы искусственного интеллекта, киберфизических систем, нейронных сетей, основанных на нем, а также роботов и объектов робототехники.

[12] Профайлинг — это совокупность психологических методов и методик оценки и прогнози-рования поведения человека на основе анализа наиболее информативных частных призна-ков, характеристик внешности, невербального и вербального поведения.

Вопросы правовой регламентации правовых принципов взаимодействия человека, общества, государства в лице государственных органов является необходимым условием гарантированности конституционных норм, закрепляющих человека, его права и свободы как высшую ценность Российской Федерации.

7.2. Структуры нормативной правовой базы в сфере правового регулирования искусственного интеллекта в Российской Федерации

На современном этапе развития цифровых технологий и технологий искусственного интеллекта в Российской Федерации на 1 апреля 2021 г. действуют следующие юридические документы, определяющие концептуальные и стратегические подходы к регулированию данной сферы общественных отношений.

– *«Национальная стратегия развития искусственного интеллекта на период до 2030 г.»*, в которой даны определения самого искусственного интеллекта, его технологий и перспективных методов, а также глоссарий всех используемых терминов в сфере правового регулирования искусственного интеллекта, его создания и применения.

– *«Стратегия развития информационного общества в Российской Федерации на 2017–2030 годы», утвержденная Указом Президента Российской Федерации от 9 мая 2017 г. № 203,* где AI и его технологии названы технологической основой государственного, социально-экономического развития страны. В Стратегии обозначено содержание понятия «информационная безопасность» и указаны правовые меры по ее обеспечению,

особенно в части предотвращения информационных угроз, которые могут исходить от использования информационных технологий без учета необходимости обеспечения информационной безопасности.

– Паспорт национальной программы «Цифровая экономика Российской Федерации» включает шесть федеральных проектов: «Нормативное регулирование цифровой среды», «Информационная инфраструктура», «Кадры для цифровой экономики», «Информационная безопасность», «Цифровые технологии» и «Цифровое государственное управление», где особое место уделяется практическому применению не только цифровых технологий, но и технологиям искусственного интеллекта.

– *Федеральный закон от 26 июля 2017 г. № 187-ФЗ «О безопасности критической информационной инфраструктуры»*, где основной задачей ставится формирование специальной «системы управления российским сегментом сети "Интернет", повышение защищенности критической информационной инфраструктуры и устойчивости ее функционирования, а также обеспечение безопасности информации, передаваемой по ней и обрабатываемой в информационных системах на территории России».

– *Федеральный закон от 24 апреля 2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации — городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона "О персональных данных"»* (далее по тексту Федеральный закон № 123-ФЗ), в котором с 1 июля 2020 г. в Москве устанавливается экспериментальный правовой режим на пять лет по внедрению AI в жизнь каждого жителя умного города,

установлены цели, задачи и принципы этого экспериментального правового режима.

– *Федеральный закон от 31 июля 2020 г. № 258-ФЗ (последняя редакция) «Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации»*, где в п. 2. ст. 1 определены сферы применения цифровых инноваций, в том числе и AI, нейронных сетей, квантовых технологий, роботов и объектов робототехники — медицина, транспорт, сельское хозяйство, финансовый рынок, предоставление государственных и муниципальных услуг, продажа услуг, товаров, работ в интернет-пространстве, архитектура, промышленность и иные сферы, определенные Правительством Российской Федерации. Таким образом, закон касается формирования комфортной правовой среды в виде «правовых песочниц» в виде особой формы экспериментального правового режима.

– *Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.»* (далее по тексту — Концепция), в которой определены основы нормативного регулирования технологий искусственного интеллекта и робототехники с соблюдением прав граждан и обеспечением безопасности личности общества и государства, а также механизмы правового регулирования в части «формирования основ правового регулирования новых общественных отношений, складывающихся в связи с разработкой и применением технологий искусственного интеллекта и робототехники и систем на их основе, а также определение правовых барьеров, препятствующих разработке и применению указанных систем», к числу которых следует отнести создание соответствующих норм в части определения юридической

ответственности при применении AI и иных объектов, определяемых в Концепции; специальное отраслевое регулирование технологий искусственного интеллекта и объектов робототехники, прежде всего, в таких сферах социально-экономического развития, как медицина, промышленность, транспорт, государственное и муниципальное управление, градостроительство, космическая деятельность и финансовое законодательство.

В конце 2020 г. Президент Российской Федерации в целях дальнейшего развития механизма правового регулирования создания и применения AI, роботов, объектов робототехники и киберфизических систем дал в общей сложности одиннадцать поручений Правительству Российской Федерации. Так, в соответствие с подп. (а) п. 1 *Поручения Президента Российской Федерации от 31 декабря 2020 г. № Пр-2242 «Перечень поручений по итогам конференции по искусственному интеллекту»* в срок до 1 мая 2021 г. глава государства поручил Правительству Российской Федерации «обеспечить принятие федеральных законов, предусматривающих возможность установления в отдельных отраслях экономики и социальной сфере экспериментальных правовых режимов в целях расширения применения технологий искусственного интеллекта»; а также к 1 июля 2021 г. обеспечить исполнение по тексту подп. (в) п. 1 законотворческой деятельности по изменению российского законодательства в целях «ускоренного создания отечественного программного обеспечения и программно-аппаратных комплексов на основе технологий искусственного интеллекта» в части «предоставления (при условии обеспечения защиты персональных данных) организациям, разрабатывающим технологические решения на основе искусственного интеллекта, доступа к наборам данных, содержащимся в том числе в государственных информационных

системах, а также возможности использования указанными организациями таких данных», а также в соответствии с подп. (з) п. 1 изменения в медицинском законодательстве в части развития телемедицинских технологий, выдачи медицинских документов в электронной форме и подп. (д-2) п. 1 в законодательстве об образовании в части формирования дополнений во все образовательные программы разделов по теории и практике применения искусственного интеллекта, для чего создать условия для приоритетного изучения математики и информатики (подп. (е) п. 1).

Как видно из этого перечня, специальное законодательное регулирование AI, его технологий, роботов, объектов робототехники еще предстоит создать, но такая задача поставлена Президентом Российской Федерации и для ее выполнения отведено только 5–6 месяцев. Полагаем, что следует определить не только сферу действия такого законодательства, но и принципы, структуру и логику взаимосвязи норм права в различных сферах такого правового регулирования. В этой связи полагаем необходимым обратиться к зарубежному опыту, так как подобное законодательство уже частично существует в отдельных странах, а стратегические документы, этические правила, этические кодексы взаимодействия с AI на сегодняшний момент есть уже в более чем в пятидесяти государствах, а также к российской и иностранной научной литературе для определения концептуальных подходов «специального законодательного регулирования, учитывающего специфику применения технологий искусственного интеллекта и робототехники»[13].

[13] Цели. Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г / Распоряжение Правительства РФ от 19 августа 2020 г.

Новые технологии, обычно описываемые общим термином «искусственный интеллект», становятся все более распространенными в человеческом обществе. Они быстро развиваются и влияют практически на все аспекты нашего существования: автопилотники, телемедицинские технологии, чат-боты, образовательные технологии, большие данные (Big Data), умный город, умный дом, автоматизированные методы наблюдения, вооруженные технологии искусственного интеллекта, киберправосудие и т. д. Нейронные сети, основанные на концепте интеллектуального анализа данных, позволяют получать огромное количество информации за короткий промежуток времени. Искусственный интеллект революционизирует финансовые услуги с помощью приложений, простирающихся от обнаружения мошенничества, уклонения от уплаты налогов или отмывания денег до регуляторных технологий (RegTech), улучшающих регулятивные процессы, такие как мониторинг, отчетность и соблюдение требований. Система правосудия все больше полагается на системы принятия решений с помощью искусственного интеллекта для прогнозирования действий правоохранительных органов, государственных и муниципальных служб и вынесения приговоров. И список примеров можно было бы продолжать.

С ростом повсеместного распространения искусственного интеллекта в настоящее время обсуждается необходимость создания как национального законодательства, так и международного и регионального, регулирующих абсолютно новую систему общественных отношений,

---

№ 2129-р «Об утверждении Концепции развития регулирования отноше-ний в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.» // Официальный интернет-портал правовой информации. URL: http://www.pravo.gov.ru (дата обращения: 26.08.2020).

складывающуюся под влиянием инновационных технологий, прежде всего AI, проникающих практически во все сферы жизни современного общества. Основной целью подобного специального законодательства должно стать формирование инновационной, дружественной, но безопасной, регулирующей среды. Адекватное правовое регулирование, как отмечается в научной литературе, имеет ключевое значение для максимизации выгод и минимизации рисков, связанных с AI и его технологиями.

7.3. Транснациональное правовое регулирование искусственного интеллекта

В зарубежной литературе довольно часто встречается мнение о необходимости создания *специальной международной структуры по правовому регулированию за AI*, которая, опираясь на междисциплинарный опыт, занималась созданием унифицированных стандартов регулирования технологий AI и единых правил правовой политики AI в разных странах мира в целях координации усилий всех государств. Подобные центры координации успешно работают, например, в Европейском союзе, США, Восточной Азии и некоторых других регионах[14]. Так, в региональных рамках ЕС существуют несколько платформ-организаций с разной степенью технического и мониторингового охвата.

Особую значимость для правового регулирования AI приобретает созданная год назад **Обсерватория Организации экономического сотрудничества и развития (ОЭСР)**

[14] К числу таких организаций следует отнести Международную ассоциацию по искусственному интеллекту и праву (International Association for Artificial Intelligence and Law); Партнерство по ИИ (Partnership on AI), Форум по искусственному интеллекту Новой Зеландии (Forum of New Zealand), SPARC в ЕС и др.

по политике в области искусственного интеллекта (OECD AI Policy Observatory, AIPO)[15], целью которой провозглашается координация усилий государств по правовому регулированию AI и обмен между ними опытом, для чего странам-участницам предоставляется доступ к площадкам AIPO[16]. Как отмечают, «AIPO — это комплексная аналитическая платформа по обзору политических мер и различных национальных инициатив в области искусственного интеллекта», которая занимается мониторингом создания и развития цифровых технологий и технологий искусственного интеллекта на национальном уровне, для чего разрабатываются различного рода расчетные методики, показатели и индикаторы.

Как отмечается исследователями, на сайте AIPO информация о роли и месте AI располагается в таких смысловых группах, как «AI-принципы ОЭСР, имплементация принципов, в том числе практические рекомендации; база данных по AI-политикам и инициативам в различных областях экономики; тренды и полезная информация: исследования по метрикам и методам измерения AI, различные данные от партнерских организаций, депозитарий по тематическим областям, мониторинг прикладных направлений развития AI по секторам экономики; информация по странам (представлены 59 стран, в том числе и Российская Федерация[17], и общие показатели по ЕС)

[15] Официальный сайт OECD AI Policy Observatory. URL: https://www.oecd.ai/ (дата обращения: 26.03.2021).

[16] Россия также обладает такой возможностью.

[17] Russian Federation / OECD Going Digital Toolkit. URL: https://goingdigital.oecd.org/countries/rus (дата обращения: 26.03.2021). При выборе в соответствующем окошке Российской Федерации можно увидеть информацию по основным правительственным документам в сфере ИИ — всего 8 инициатив, в то время как у США — 36, ЕС — 22, Германии — 15, Канады — 10.

и инициативам: национальные стратегии и политики, инициативы различных мировых стейкхолдеров». Методология проводимых исследований основывается на трех основополагающих принципах: междисциплинарность, сравнительный анализ, включая сравнительно-правовой анализ стратегических и иных юридических документов по регулированию AI в каждой из стран или регионов, международное сотрудничество в сфере создания, развития и будущего AI.

С 2019 г. хорошо себя зарекомендовал **практический интерактивный сервис «Инструментарий по изменению цифровизации» — Going Digital Toolkit**, позволяющий в режиме онлайн визуализировать показатели стран в области формирования цифровой экономики и практического использования AI по семи ключевым показателям политики в сфере AI, каждый из которых разделен, в свою очередь, на 33 критериальных индикатора, имеющих для удобства восприятия определенное цветовое решение. К числу таких показателей официальный сайт относит: «1) доступ к коммуникационным инфраструктурам, услугам и данным; 2) эффективное использование цифровых технологий и данных; 3) цифровые инновации и инновации, основанные на базах данных; 4) совершенствование компетенций и положительная динамика в сфере трудовой занятости; 5) инклюзия общественных структур в цифровую экономику; 6) доверие к различного рода цифровым технологиям; 7) открытость рынка в цифровой деловой среде».

Мониторинг развития AI в мире основывается на использовании известных юридических документов в сфере AI: *The Partnership on AI to Benefit People and Society; The AI Initiative of the Future Society; The Ethically Aligned Design principles (version 3) of the Institute of Electrical and Electronics Engineers (IEEE); ISO/IEC joint technical*

committee (JTC) 1/ Standards Committee (SC) 42 on Artificial intelligence; The Asilomar AI Principles of the Future of Life Institute, которые являются примерами мягкого права и одновременно базой для проведения сравнительного анализа национального правового поля.

В работе созданных при Европейской комиссии различных проектных учреждений, занимающихся анализом и координацией теории и практики AI, принимают активное участие такие международные организации, как ЮНКТАД, Всемирный банк, Международный союз электросвязи и др., которые проводят детальные мониторинги происходящих изменений в сфере правового регулирования AI. К числу наиболее успешных следует отнести следующие из них.

1. *AI WATCH* — отдельный электронный ресурс Европейской комиссии, отслеживающий особенности европейского регулирования AI и анализирующий стратегии ряда стран-участниц ЕС; на платформе есть специальные закладки по методологии такого мониторинга, даны индикаторы измерений развития технологий AI, а также значительный ресурс данных о различных исследованиях и данные по социальному воздействию AI, в том числе по применении AI в сфере культуры и права интеллектуальной собственности, в сфере действия государственных органов, в том числе правоохранительных, пограничного контроля и безопасности, медицины и здравоохранения, образования стран Западной Европы.

Так, в 2020–2021 гг. (по состоянию на 1 апреля 2021 г.) на официальном сайте Европейского союза размещено около 25 различного рода исследований, подготовленных в рамках деятельности AI Watch, касающихся разных сфер применения AI и его влияния на действующий ландшафт государств — участников ЕС. В отчете,

подготовленном в рамках данной организации, «TES-анализ мировой экосистемы AI в 2009–2018 гг.» *("TES analysis of AI Worldwide Ecosystem in 2009–2018")* представлены ответы на такие вопросы, как размер экосистемы AI в мире и на уровне страны; уровень промышленного участия в стране; данные компаний, профилирование экономических агентов в соответствии с их сильными сторонами в инновациях и использовании AI, включая их эффективность в области патентования; а также степень внутреннего и внешнего сотрудничества между фирмами и исследовательскими институтами из разных стран мира.

2. В структуру проекта *европейского AI-сообщества AI4EU* включена *Обсерватория OSAI (The European Observatory on Society and Artificial Intelligence)*, в задачи которой входит как обеспечение распространения, так и поддержание дискуссионного формата обсуждения собранной и представленной информации об этических, правовых, социально-экономических и культурных проблемах, связанных с распространением AI в Европе, а также платформа — AI4EU Platform, действующая по принципу сетевого сообщества, снабжая его участников информацией о предоставляемых услугах с использованием AI, различного рода программном обеспечении, возможности получения экспертного заключения по инновационным технологиям, и в первую очередь технологиям AI.

Как и на наднациональном уровне, в отдельных странах в целях управления AI создаются аналогичные координационные центры, аккумулирующие всю информацию о создании и продвижении в различные сферы жизни общества технологий AI, роботов и объектов робототехники. Примерами могут служить *канадский координационный центр «Ответственное производство*

и использование искусственного интеллекта и цифровых технологий» (l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA)), созданный на деньги Квебекского исследовательского фонда как открытая исследовательская сеть, объединяющая опыт более двухсот двадцати исследователей в области гуманитарных, социальных, технических, а также медицинских наук. Это открытое пространство для обсуждения и размышлений для всех заинтересованных сторон в разработке и использовании AI и цифровых технологий в сфере искусства, СМИ и культурного разнообразия; индустрии 4.0, трудовой и профессиональной занятости; окружающей среды, умных городов, территории и мобильности населения; международных отношений, гуманитарной деятельности в целом, прав и свобод человека; этики, государственного управления и демократического строя; системы права и системы законодательства, киберправосудия и кибербезопасности; процессов образования и обучения, а также медицины и здравоохранения.

На территории Российской Федерации пока еще нет подобных перечисленным выше информационных ресурсов для мониторинга ежедневно меняющейся ситуации вокруг теории и практики создания AI и объектов робототехники. Однако многие специалисты в данной области полагают, что их создание необходимо. Создание таких информационных ресурсов на территории нашей страны необходимо для того, чтобы:

1) разработать в Российской Федерации специальное законодательство в сфере AI, роботов и объектов робототехники, основываясь на анализе различного рода стратегических документов, политических инициатив, программ и проектов, а также отдельных нормативных правовых актов, доказавших позитивное социально-экономическое воздействие AI в среднесрочной и долгосрочной временной перспективе;

2) заключить двусторонние и многосторонние договоры по соответствующим направлениям международного и регионального сотрудничества Российской Федерации с иностранными государствами в сфере создания и использования инновационных технологий, прежде всего оказывающих положительное воздействие на развитие социальной и экономической сфер жизни россиян, при обязательном соблюдении национальных интересов;

3) определить особенности правового регулирования искусственного интеллекта, области его применения;

4) сделать возможным обмен информацией в сфере AI с зарубежными партнерами и способствовать росту глобальной конкурентоспособности России на международной арене.

Как отмечают Оливия Дж. Эрдели и Джуди Голдсмит (Erdélyi, 2018), вопрос, который приобретает все большую практическую важность в связи с быстро растущим влиянием искусственного интеллекта на жизнь людей практически во всех областях нашей жизни, заключается в необходимости заполнить возникший правовой вакуум. По мнению зарубежных исследователей, правительства большинства стран с развитыми экономиками (Канада, Китай, Япония, Великобритания, США и ЕС) создают значительное число стратегических документов, основными целями в которых провозглашаются развитие и коммерциализация AI с целью поддержания устойчивой экономической конкурентоспособности после неизбежного глобального перехода к экономике знаний, управляемой AI. Применение технологий AI, в ходе которых возникает целый ряд проблем, подлежащих в последующем судебному разрешению, привело к постановке вопроса о необходимости междисциплинарного сотрудничества в сфере правового регулирования AI.

7.4. Национальное законодательное регулирование искусственного интеллекта и перспективы «мягкого» права

Проблема национального законодательного регулирования заключается в том, что *системе правотворчества и законотворчества различных стран присущ целый ряд особенностей, соответствующих правовой действительности и правовому пространству каждого конкретного государства* в то время, как искусственный интеллект не существует исключительно в правовом поле только отдельно взятой страны, необходимы единые подходы, общие правовые и этические принципы, касающиеся основ взаимодействия человека и подобных технологий. *Правовые нормы, принимаемые в сфере AI, должны быть действительно легитимные и институционализированными, а не символическими правилами, не оказывающими никакого влияния на нормативное регулирование в целях установления соответствующего правопорядка.* Основная проблема разработки юридических документов заключается в том, что они должны иметь возможность контролировать развитие и функционирование инновационных технологий, обеспечивая при этом, чтобы прогресс развития AI не был затруднен в значительной степени или полностью сорван.

Кроме того, следует помнить, что цифровые технологии и технологии AI выходят за рамки только национального регулирования и в этом случае возникает конфликт между правовыми порядками разных стран, регулирующих AI, роботов и объекты робототехники. Значительные трудности, возникающие при столкновении различных правовых режимов, позволяют считать не только возможным, но и необходимым унификацию национальных норм права в данной сфере. Ведь если не пойти на такой

шаг, подобное противоречие между транснациональным характером воздействия цифровых технологий и технологий искусственного интеллекта на жизнь социума и национальными правовыми системами, в рамках которых происходит регулирование AI и CPS, создаст известные сложности на современном этапе государственно-правового развития большинства стран. Однако есть и вторая сторона медали: в основе транснационального правового регулирования находится сложная система рекурсивных разнонаправленных процессов, которые, следуя своей собственной логике, оказывают значительное, если не сказать, огромное влияние на авторитетность международных норм права, зачастую «подавляя» правовой суверенитет отдельных национальных правовых режимов.

Возможно ли правовое регулирование AI по той же схеме, по какой происходит регулирование различных сфер жизни человека в обществе и государстве? В научной литературе есть мнение о том, что существуют **«жесткое» право (hard law)** и **«мягкое» право (soft law)**, разница между которыми заключается в регуляторах, зафиксированных в системе законодательства. «Жесткое» право основано на трех основных видах норм права — управомочивающих, запрещающих и обязывающих, то есть подразумевающих наличие выраженной воли субъекта права. В отношении существующего сейчас искусственного интеллекта говорить о присущей ему воле пока нет оснований. Поэтому понятия разрешительных, обязывающих или рекомендательных норм права не могут соотноситься со сферой технологий, «имитирующих когнитивные способности человека», а следовательно, закон в том виде, о котором мы говорим, не вполне будет соотносим с AI как сложно-определяемой в правосубъектном плане ипостасью.

В последнее время в научной литературе возник значительный интерес к феномену «мягкого» права, широко

обсуждавшемуся с 1970-х гг. Так, Ю.Б. Фогельсон (2013) в статье «Мягкое право в современном правовом дискурсе» полагает «мягкое» право идентичным «живому» праву в социологической теории права Ф.К. фон Савиньи и О. Эрлиха. Исходя из данной теории, право можно представить системой таких элементов, как «социально-действенные правила поведения, объективированные в правовых нормах, общедоступные, систематизированные профессиональными юристами на основе общих принципов (научной основе), санкционированные и приводимые в исполнение государством, постоянно изменяемые, сознательно приспосабливаемые к изменениям общественной жизни профессиональными юристами».

Таким образом, в отличие от «жесткого» права, которое не только создает обязанности определенного поведения для сторон правоотношений, но и требует их исполнения под угрозой применения санкции, «мягкое» право также создает обязанности, однако требования их исполнения с юридической точки зрения нет. Однако существуют иные способы заставить субъектов не пренебрегать нормами soft law. Так, на международном уровне различные организации создают нормы-рекомендации, фиксирующие желательное поведение, в различных правилах, стандартах, этических кодексах и др., при этом субъекты следуют им из опасения, что в случае их неисполнения, они будут признаны неблагонадежными и по отношению к ним могут применить определенное давление.

Подобное применение «мягкого» права характерно не только для англо-американской системы права, но и для государств-членов ЕС, которые довольно часто прибегают к «открытому методу координации». Суть такого метода заключается в создании различного рода рекомендаций, необязательных к исполнению с точки зрения закона, но их исполнение жестко контролируется

специально созданной для этого общественной организацией, в функции которой входит контроль за исполнением таких «необязательных» правил. В результате анализа составляемых этими организациями отчетов содержащие подобного рода рекомендации стандартизированные документы совершенствуются, и причины неисполнения soft law раскрываются. Такой неюридический способ принуждения к исполнению рекомендованных норм получил название «naming and shaming» и практикуется не только в ЕС, но и в Организации экономического сотрудничества и развития (ОЭСР) и Global Compact в различных сферах, в том числе и в финансовой.

На первый взгляд, «мягкое» право сходно с морально-нравственными социальными нормами, но это не совсем так. Отличительными чертами именно «мягкого» права являются следующие признаки. Во-первых, нормы права, подлежащие исполнению «по желанию», опубликованы в официальных документах определенной организации и находятся в широком доступе, поэтому ознакомление с ними не вызывает никаких проблем, при этом не закреплена обязательность их исполнения. Во-вторых, контроль за выполнением предписанных стандартами или иными документами возложен на специально созданный орган, «встроенный» в структуру определенной организации, в задачи которого входит социальное давление на тех, кто «мягкое» право не хочет исполнять. В-третьих, правила-рекомендации — это «живое» право, которое, приспосабливаясь к изменяющейся действительности, способно к изменениям. Если «мягкое» право реализуется на национальном уровне, то для создания соответствующих правил используется государственный аппарат, в том числе и контрольные органы для применения различных форм социального давления. И последнее — все эти характерные признаки свидетельствуют о том, что

в случае применения «мягкого» права понятия государственного суверенитета практически не существует.

Исходя из анализа наднациональных документов в сфере искусственного интеллекта, можно сделать вывод, что по отношению к ним применяется как раз «мягкое» право. Гэри Марчант в препринте, опубликованном в рамках онлайн-проекта AI Pulse Калифорнийского университета в Лос-Анджелесе, об использовании «мягкого» права для регулирования искусственного интеллекта (Marchant, 2019) отмечает, что сама проблема правового регулирования AI порождает целый ряд вопросов, ответ на которые может дать только soft law. Ведь особенности систем правотворчества в каждой отдельной стране не могут обеспечить единый подход к регулированию AI. Сегодня реальность такова, что в ближайшие несколько лет в лучшем случае будет какое-то спорадическое разрозненное традиционное регулирование AI, несмотря на всё более широкое развертывание и применение AI в растущем диапазоне приложений и отраслевых секторов. Поэтому на таком промежуточном этапе правового регулирования «пробел в управлении» для AI будет в основном восполнен именно механизмами «мягкого» права, которые излагают существенные ожидания, но не подлежат прямому исполнению со стороны правительства, включая такие подходы, как профессиональные руководящие принципы, частные стандарты, кодексы этики, кодексы поведения.

Сегодня в Российской Федерации поставлен вопрос о необходимости специального законодательства в сфере разработки и практического применения систем искусственного интеллекта. Пока нет непосредственной перспективы появления сильного AI, возможности soft law представляются вполне достаточными. Даже традиционное регулирование конкретных технологий AI может быть достаточным, хотя эти технологии сами по себе

могут оказаться невосприимчивыми к комплексным нормативным решениям. К тому же многие риски для здоровья, безопасности и окружающей среды, создаваемые искусственным интеллектом, не всегда можно соотнести с национальной законодательной системой. Возникает дилемма: можно ли ограничиться изменениями и дополнениями в существующие нормативные правовые акты или необходимо срочно создавать специальные законы, регулирующие либо отдельные аспекты AI, либо кодификационный акт в сфере цифровых технологий, технологий AI, инновационных технологий и др.

По мнению Гэри Марчанта, «риски, преимущества и развитие AI весьма неопределенны, что опять же затрудняет принятие традиционных упреждающих нормативных решений, а национальные правительства не хотят препятствовать инновациям в новой технологии упреждающим регулированием в эпоху жесткой международной конкуренции». Правительства ЕС, США, Великобритании и других стран все чаще принимают решения о превалировании «мягкого» права над традиционным правовым регулированием, так в Стратегии AI (2018) ЕС не стала предлагать никаких новых мер регулирования для AI, и в декабре 2018 г. Комиссия ЕС опубликовала *«Скоординированный план действий по AI»*, в котором изложены цели и планы Комиссии в отношении общеевропейской стратегии по AI.

Однако Комиссия отметила, что «[пока] саморегулирование может предоставить первый набор критериев, по которым могут быть оценены появляющиеся приложения и результаты, государственные органы должны обеспечить соответствие нормативно-правовой базы для разработки и использования технологий искусственного интеллекта этим ценностям и основным правам. Комиссия будет следить за развитием событий и при

необходимости анализировать существующие правовые рамки, чтобы лучше адаптировать их к конкретным задачам, в частности для обеспечения уважения основных ценностей и основных прав Союза». Такой же позиции придерживается и верхняя палата парламента Великобритании, рекомендовавшая этический кодекс поведения для AI, в силу того что «темпы изменений в технологии AI означают, что чрезмерно предписывающие или конкретные законы не успевают за темпами и могут почти устареть к моменту его принятия», поэтому «регулирование, касающееся искусственного интеллекта, на данном этапе было бы неуместным».

Заслуживает внимания нередко высказываемое суждение, что «мягкое» право иногда можно рассматривать как переходную фазу механизма правового регулирования. Однако оно справедливо лишь отчасти: традиционное законодательное регулирование может включать в себя его элементы, когда этические принципы содержатся в тексте закона, и тогда им также придается необходимая обязательность исполнения, подкрепленная силой государства. Примером чего *является Федеральный закон от 31 июля 2020 г. № 258-ФЗ (последняя редакция) «Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации»*, где зафиксированы принципы такой правовой «песочницы», или *Закон штата Калифорния от 7 сентября 2018 г.*, «выражающий поддержку “Асиломарским принципам разработки искусственного интеллекта”». Поэтому для Российской Федерации было бы правильно создавать систему специальных норм права, дополняющих соответствующие нормативные правовые акты, прежде всего закрепляющих правовые и этические принципы AI и рамочно регулирующих специфику применения технологий в различных сферах жизни общества, в первую очередь определенных в Концепции развития

AI. Необходимость такого подхода диктует неоднозначное и слабо прогнозируемое развитие AI.

На конец марта 2021 г. в Российской Федерации действуют такие ГОСТы, как *ГОСТ Р 43.0.8–2017 «Национальный стандарт Российской Федерации. Информационное обеспечение техники и операторской деятельности. Искусственно-интеллектуализированное человеко-информационное взаимодействие. Общие положения»; ГОСТ Р 43.0.7–2011 «Информационное обеспечение техники и операторской деятельности. Гибридно-интеллектуализированное человеко-информационное взаимодействие. Общие положения»; ГОСТ Р 60.0.0.4–2019/ИСО 8373:2012. «Национальный стандарт Российской Федерации. Роботы и робототехнические устройства. Термины и определения»*.

Кроме них, в целях развития AI до 2024 г. должно быть создано в общей сложности 216 национальных стандартов по AI в соответствие с принятой Министерством экономического развития Российской Федерации в 2020 г. специальной «Программой стандартизации по приоритетному направлению “Искусственный интеллект” на период 2021–2024 годы» по следующим видам:

1) стандарты общего назначения, закрепляющие термины и определения в области AI, определяющие стадии жизненного цикла систем, универсальные принципы организации работ при создании и эксплуатации систем AI, особенности защиты информации, обрабатываемой в системах AI, устанавливающие требования к форматам представления данных и регламентирующие другие вопросы (всего предполагается создать 140 стандартов);

2) метрологические стандарты, устанавливающие унифицированные требования к процедурам оценки функциональных характеристик и характеристик безопасности систем AI, в том числе определяющие перечни

характеристик условий эксплуатации, оказывающих существенное влияние на работу систем AI, задающих требования к обучающим и тестовым наборам данных, используемым при создании и измерении характеристик систем, и при необходимости содержащие примеры представительных тестовых наборов данных для различных прикладных задач AI или групп таких задач (всего предполагается 76 стандартов).

К числу стратегических документов в сфере действия «мягкого» права, принятых за 2017–2020 гг., относятся: *«Национальная стратегия развития искусственного интеллекта на период до 2030 г.»; «Стратегия развития информационного общества в Российской Федерации на 2017–2030 годы», «Стратегия развития автомобильной промышленности РФ на период до 2025 г.», государственная программа Российской Федерации «Развитие авиационной промышленности»; национальная программа «Цифровая экономика Российской Федерации», включающая шесть федеральных проектов: «Нормативное регулирование цифровой среды», «Информационная инфраструктура», «Кадры для цифровой экономики», «Информационная безопасность», «Цифровые технологии» и «Цифровое государственное управление»; «Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.»* и др. Однако с учетом федеративного характера государственного устройства России в отдельных стратегиях социально-экономического развития субъектов Российской Федерации включены положения о необходимости развития кластеров по созданию AI, применения инновационных технологий. Таким образом, Российская Федерация действует в одном русле с другими зарубежными правовыми режимами, принявшими подобные стратегические документы. Вероятно, в части «мягкого» права будут законодательно закреплены правовые и этические

принципы, представляющие собой единый комплекс, на основе которого и должно быть создано российское отраслевое законодательство в области AI.

7.5. Сравнительный анализ национального законодательства иностранных государств в сфере искусственного интеллекта

Как отмечается в специальном *«Исследовании последствий применения передовых цифровых технологий (включая системы искусственного интеллекта) для концепции ответственности внутри системы прав человека»*, подготовленном экспертным комитетом по правозащитным аспектам автоматизированной обработки данных и различным формам искусственного интеллекта (MSI-AUT) в 2019 г. по заказу Европарламента, технологии искусственного интеллекта, с одной стороны, могут значительно облегчить жизнь, предоставляя ранее невообразимые преимущества, их применение чревато значительными и труднопредсказуемыми проблемам.

По данным ОЭСР, на конец 2019 г. около 37 стран мира уже имели стратегии по развитию AI и примерно тринадцать стран готовились к их принятию. Большинство опрошенных юрисдикций рассматривают AI в позитивном свете и стремятся стать лидерами в этой области. Поэтому многие страны разработали или находятся в процессе разработки национальных стратегий и планов действий (рис. 13) в области искусственного интеллекта или цифровых технологий. В стратегиях и планах действий среди прочего подчеркивается необходимость разработки этических и правовых норм, обеспечивающих разработку и применение AI на основе правовых, этических и религиозных традиционных ценностей каждой конкретной страны или региона.

Рис. 13. Страны, обладающие стратегическими документами в сфере искусственного интеллекта

Сравнительно-правовой анализ стратегических документов в сфере теории и практики AI, роботов и объектов робототехники подробно представлен в аналитическом отчете Агентства Искусственного Интеллекта «Сравнение национальных стратегий в области искусственного интеллекта» и «Обзоре отдельных вопросов в области больших данных и искусственного интеллекта», подготовленном Главным информационно-аналитическим центром Министерства внутренних дел Российской Федерации.

На современном этапе специальное законодательство, регулирующее AI и связанные с ним технологии, существует на национальном уровне. Как показал его анализ в разных странах мира, существуют единые сферы правового регулирования, которые были задействованы практически во всех странах мира — государственное и муниципальное управление, предоставление товаров, работ и услуг, автономные транспортные средства, правосудие, смертоносные автономные системы вооружений, сфера образования и медицины, базирующиеся на базовых принципах защиты данных и конфиденциальности, прозрачности, контроля со стороны человека, наблюдения и др.

Некоторые страны сделали первые шаги по использованию AI в сфере правосудия. Так, в Португалии запущен инструмент правовой помощи, который будет исследовать поступающие запросы и извлекать уроки из них. В будущем его можно будет использовать для прогнозирования вероятности успеха судебного процесса. Аналогичным образом, во Франции апелляционные суды Ренна и Дуэ используют программные средства для прогнозирования правосудия по различным апелляционным делам с 2017 г.

С 2016 г. в была принята специальная поправка к *Венской конвенции о дорожном движении (1968)*, которая

с целью «облегчить международное дорожное движение и повысить безопасность дорожного движения путем принятия единых правил устранила правовые препятствия для договаривающихся сторон, позволяющие передавать задачи вождения на автоматизированные технологии». В целом ряде стран — участниц этой Конвенции были приняты соответствующие нормы права, позволяющие проводить испытания автономных транспортных средств на дорогах общего пользования, но при обязательном присутствии в автомобиле человека-водителя, способного при необходимости взять на себя функции управления. Исключение составляет законодательство Нидерландов и Литвы, разрешающих экспериментальное использование беспилотных транспортных средств без водителя-человека на дорогах общего пользования — *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (Warrendale: SAE International, 15 June 2018),* использующие классификацию SAE: SAE International, J3016_201806. Израиль принял *постановление и директиву по экспериментам с автономными транспортными средствами.*

Таким образом, структура системы специального законодательства в сфере AI должна включать в себя soft law (технические стандарты по AI, стратегические документы в области создания, развития и применения AI, роботов, объектов робототехники, систему этических и правовых принципов взаимодействия технологий с человеком) и hard law (необходимые дополнения и изменения в отраслевое законодательство федерального и регионального уровней регулирования в части применения AI, роботов и объектов робототехники в сфере государственного и муниципального управления, предоставления товаров, работ и услуг, правового регулирования автономных транспортных средств, киберправосудия, сферы образования,

медицины и здравоохранения, права интеллектуальной собственности и т. д.) в целях защиты прав и свобод человека и гражданина, достижения высокого уровня и социально-экономического развития в целом страны и благосостояния россиян в частности.

На основе анализа зарубежного законодательства по вопросам развития AI и робототехники можно увидеть, что часть стран (в большинстве своем страны Тихоокеанского региона и Восточной Азии) пошла по пути создания единого закона в данной сфере, примером чему служит *Закон Южной Кореи № 9014 от 28 марта 2008 г. (в ред. Закона № 13744 от 6 января 2016 г.) «О содействии развитию и распространению умных роботов»*, другие — по пути создания отдельных нормативных правовых актов в различных отраслях правового регулирования, либо внося соответствующие изменения в уже имеющиеся акты, либо создавая специальное законодательство.

Российская Федерация идет по второму пути: уже сегодня вносятся изменения, например в *Гражданский кодекс Российской Федерации добавляется отдельная статья 141.1 «Цифровые права», в Федеральный закон от 21 ноября 2011 г. № 323-ФЗ (ред. 22 декабря 2020 г.) «Об основах охраны здоровья граждан в Российской Федерации» — статья 36.2 «Особенности медицинской помощи, оказываемой с применением телемедицинских технологий»*. Очевидна необходимость изменений в административном, финансовом, трудовом, налоговом, транспортном законодательствах, законодательстве об образовании и многих других отраслях российского законодательства, регулирующих как ограничения в применении AI, роботов и иных объектов робототехники из-за связанных с ними рисков, в особенности в сфере нарушения прав и свобод человека и гражданина, так и сами применения — в частности, этические и правовые

принципы взаимодействия AI и человека, вопросы правосубъектности, юридической ответственности в различных сферах правового регулирования, особенности применения и т. п.

Данный подход обосновывается тем, что, как указывают Ю.А. Тихомиров и С.Б. Нанба в статье «Роботизация: динамика правового регулирования» (2020): «Одни из них [роботов] способствуют процессу роботизации в разных сферах в качестве собственной части программ и проектов социально-экономического развития, другие вводят технические эквиваленты традиционных действий в правовых актах, третьи обеспечивают совмещение новых и традиционных режимов деятельности, наконец, четвертые обеспечивают соблюдение установленных режимов с помощью менталитета». Кроме того, необходимо учесть и тот факт, что Россия, будучи федеративным государством, кроме федеральной системы законодательства, имеет и системы законодательства субъектов Российской Федерации, поэтому следует помнить о необходимости создания регионального законодательства в сфере правового регулирования AI в духе соответствующих стратегических и концептуальных документов и Поручения Президента Российской Федерации от 31 декабря 2020 г.

Ключевые понятия: искусственный интеллект, киберфизические системы, алгоритм, правовые принципы, международное законодательство, права человека, национальное законодательство, стратегия развития искусственного интеллекта; правоотношение, объект и предмет правоотношения, юридическая ответственность.

Контрольные вопросы:

1. Какие законы робототехники А. Азимова следует применить к регулированию искусственного интеллекта?

2. Перечислите и дайте характеристику основным инициативам взаимодействия человека и искусственного интеллекта, принятым в последние годы в странах ЕС.

3. Какие положения российского федерального законодательства, касающиеся искусственного интеллекта или технологий искусственного интеллекта, вы можете назвать? В чем их плюсы и минусы (на ваш взгляд)?

4. Перечислите правовые принципы взаимодействия с искусственным интеллектом на основе анализа современного зарубежного законодательства в данной области.

5. Какие основные положения законодательства об автопилотниках вы можете перечислить? На ваш взгляд, их можно использовать в российской практике?

6. В чем заключается отличие правового регулирования в США и Южной Корее? Что такое умные роботы?

Практико-ориентированные задания

1. Проанализируйте «Модельную конвенцию о робототехнике и искусственном интеллекте», содержащую правила создания и использования роботов и иных систем искусственного интеллекта, размещенную по ссылке: http://robopravo.ru/modielnaia_konvientsiia, и создайте таблицу на основе ее содержания.

№	Содержание разделов конвенции	Основное содержание	«+»	«–»
1	Правила безопасности роботов, включающие фиксацию роботами информации об условиях своего функционирования и всех совершаемых ими действиях («черный ящик») и обеспечение функцией моментального или			

Продолжение табл.

№	Содержание разделов конвенции	Основное содержание	«+»	«–»
	аварийного отключения по требованию роботов, физически взаимодействующих с людьми («красная кнопка»)			
2	Общие правила создания роботов (ответственность, общее благо)			
3	Общие правила использования роботов (соблюдение прав человека и общепринятых норм морали и нравственности, информированность о функционировании роботов, возможность признания роботов субъектами права)			
4	Правила разработки искусственного интеллекта (презумпция опасности искусственного интеллекта)			
5	Ограничения по использованию военных роботов (их применение не должно нарушать общепринятых в мире гуманитарных правил ведения войны и использоваться для причинения вреда мирному населению)			
6	Развитие правил робототехники и искусственного интеллекта (содействие			

Окончание табл.

№	Содержание разделов конвенции	Основное содержание	«+»	«–»
	разработке общепринятых международных правил и созданию наднациональных институтов при уже существующих международных объединениях и организациях)			

2. Постройте схему международно-правового регулирования искусственного интеллекта на современном этапе развития цифрового общества.

Темы докладов и сообщений

1. Основные положения «Азиломарских принципов искусственного интеллекта» и возможность их применения в России.

2. Закон Южной Кореи «О содействии развитию и распространению умных роботов»: возможность использования в России.

3. Инициативы Франции в сфере робототехники: возможность использования в России.

4. Основные институты Закона Калифорнии об идентификации ботов: возможность использования в России.

5. Эстонский закон о роботах-курьерах: возможность использования в России.

ГЛАВА 8.
ЭТИЧЕСКИЕ ПРИНЦИПЫ ВЗАИМОДЕЙСТВИЯ ЧЕЛОВЕКА, ОБЩЕСТВА, ГОСУДАРСТВА И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЮРИДИЧЕСКИХ ДОКУМЕНТАХ

В результате изучения материалов главы обучающийся должен

знать:

– действующие документы в сфере искусственного интеллекта;

– этические принципы, принимаемые в различных странах мира по проблеме взаимодействия ИИ и человека;

уметь:

– анализировать юридические и иные документы в сфере правового и иного взаимодействия ИИ и социума;

– давать правовую оценку фактическим обстоятельствам дела в области взаимодействия социума с искусственным интеллектом в зависимости от правового поля отдельных государств;

владеть навыками:

– анализа различных явлений, фактов, социальных норм и общественных отношений в сфере ИИ;

– юридически грамотной квалификации отношений с ИИ, его морально-этической составляющей.

8.1. Этика в правовом регулировании искусственного интеллекта

Есть и еще одна проблема, которую в самом начале создания системы правового регулирования AI не сразу учли, — это этика взаимодействия человека и технологий. Если вначале на наднациональном и национальном уровнях начали создаваться стратегические документы и отдельные законы в сфере AI, то уже через два года главным становится фиксация именно этических норм, которые, как мы указывали в предыдущем исследовании (этап II), должны быть включены в законодательные акты как на наднациональном, так и национальном уровне регулирования. По мере того, как использование и воздействие автономных и интеллектуальных систем (А/ИС) становятся все более распространенным, нам необходимо установить социальные и политические ориентиры, чтобы такие системы оставались ориентированными на человека, служа ценностям и этическим принципам человечества. Эти системы должны развиваться и функционировать таким образом, чтобы приносить пользу людям и окружающей среде, а не просто достигать функциональных целей и решать технические проблемы. Такой подход будет способствовать повышению уровня доверия между людьми и технологиями, необходимого для их плодотворного использования в нашей повседневной жизни.

Примерами законодательно закрепленных этических норм в отношении AI на современном этапе могут служить такие документы, как: *«Этические руководящие принципы для Японского общества искусственного интеллекта (JSAI)» (Япония, 2017), «Руководство по этике для надежного AI Специальной группы*

экспертов высокого уровня Совета Европы» (2018), «Модельная конвенция робототехники и искусственного интеллекта» (Россия, 2018), «Римский призыв к этике» (Ватикан, 2020), «Этические принципы для искусственного интеллекта» (США, 2020), «Глобальная инициатива Института инженеров по электрике и электронике (IEEE) по этике автономных и интеллектуальных систем» (2016), 14 стандартов AI, подготовленных IEEE и касающихся управления и этических аспектов AI в целях обеспечения широкого набора требований к управлению AI *(IEEE P7000™ — модельный процесс решения этических проблем при проектировании системы; IEEE P7001™ — прозрачность автономных систем; IEEE P7002™ — процесс конфиденциальности данных; IEEE P7003™ — соображения алгоритмической предвзятости; IEEE P7004™ — стандарт управления данными детей и учащихся; IEEE P7005™ — Стандарт прозрачного управления данными работодателя; IEEE P7006™ — Стандарт интеллектуального агента управления персональными данными. IEEE P7007™ — Онтологический стандарт для робототехнических систем и систем автоматизации, основанных на этике; IEEE P7009™ — Стандарт отказоустойчивого проектирования автономных и полуавтономных систем; IEEE P7010™ — Стандарт показателей благополучия для этического искусственного интеллекта и автономных систем; IEEE P7011™ — Стандарт для процесса идентификации и оценки надежности источников новостей; IEEE P7012™ — Стандарт для машиносчитываемых условий личной конфиденциальности; IEEE P7013™ — Стандарт включения и применения технологии автоматизированного анализа состояния лица.*

В ближайшем будущем предполагается закрепить в законодательстве Российской Федерации следующие этические правила взаимодействия с AI:

а) **доверие** (системы AI должны быть понятны человеку, должно быть сформировано ощущение правильности выбора варианта решения проблемы программы-советчика («опыт оператора»);

б) **соблюдение прав и свобод человека и гражданина** (в алгоритмы работы AI должны быть заложены принципы защиты персональных данных индивида, уважения к семейной жизни, защиты интересов личности, ее права на труд и др.); инклюзивность (потребности всех людей должны приниматься во внимание таким образом, чтобы каждый человек мог принести пользу себе и всем людям, AI должны быть предложены наилучшие возможные условия для саморазвития человека);

в) **ответственность** (системы AI должны работать надежно и уважать конфиденциальность пользователей, те, кто проектирует и внедряет использование AI, должны действовать ответственно перед настоящим и будущим поколением людей);

г) **беспристрастность** (запрещено создавать условия и обстоятельства, которые могут быть оценены как предвзятые, необходимо гарантировать соблюдение принципов справедливости по отношению к человеку, охраны его чести и достоинства);

д) **надежность** (системы искусственного интеллекта должны быть способны работать надежно).

Полагаем целесообразным создавать и отечественное специальное законодательство в сфере теории и практики AI, роботов, преследующее национальные интересы с учетом этических региональных или международных принципов, устанавливающее рамочные правила в сфере новых инновационных и цифровых технологий.

Национальные усилия по разработке нормативных правовых актов в рамках следования государственной политики отдельной страны в области AI должны с самого начала координироваться и поддерживаться международной нормативно-правовой базой во избежание рисков, связанных с несовершенным взаимодействием AI с людьми, фрагментацией национальных подходов к регулированию AI.

Этика показывает, какие могут быть последствия, если конкретная идея будет воспринята отдельным человеком или социумом, какие у нее предпосылки и перспективы. От других средств социальной регуляции — права, традиции, обычая — моральные нормы отличаются тем, что они предполагают свободу выбора и регулируются преимущественно такими внутренними чувствами, как стыд, долг, угрызения совести (табл. 2).

Таблица 2

Основные категории этики (морального сознания)

Объект отражения	Категория
Общие оценки действительности с точки зрения их желательности или нежелательности для человека	Добро, зло, справедливость, счастье, смысл жизни
Способы упорядочивания совместной жизни людей	Норма, принцип, оценка, идеал
Индивидуальные механизмы работы нравственного самосознания	Долг, совесть, стыд, честь и достоинство, моральные чувства

С 2016 г. страны включились в технологическую гонку по разработке роботов, киберфизических и нейронных

сетей, сильного и сверхсильного AI: к 2020 г. были приняты 34 стратегии по развитию AI (рис. 14). Именно поэтому так важно именно сегодня задать этические рамки развития AI, ограничить возможности его неэтичного применения и направить энергию разработчиков и идеи законодателей в русло, обеспечивающее максимальную безопасность и выгоду для общества.

Рис. 14. Национальные стратегии развития искусственного интеллекта

Этику AI технологий от этики других областей отличает проблема этичного поведения интеллектуальной системы (ИС) в ситуации, когда ее решение касается людей. Принципиально важно, что система AI способна самостоятельно принимать решения, касающиеся человека, анализировать данные в таких объемах и с такой скоростью,

как человек делать не в состоянии (следовательно, человек не может проверить верность решений). Соответственно, основная проблема определение того, насколько решения, принимаемые интеллектуальной автономной системой (ИАС), соответствуют этическим нормам, то есть насколько она этична. Поэтому мы можем говорить о двух совершенно разных аспектах этики AI (рис. 15). Особое внимание необходимо уделять и профессиональной этике производителей ИС.

Рис. 15. Этика интеллектуальных автономных систем

На современном этапе развития AI особое значение приобретает практическая реализация его этической компоненты. Вопросы, которые, по мнению исследователей, особенно важны, это: 1) реализация машинной этики; 2) формализация этических понятий для технологических систем; 3) верификация и валидация этической компоненты; 4) стандартизация машинной этики; 5) стандартизация этических аспектов AI. По сравнению с содержательной стороной нравственных ценностей, изначально стабильной, мораль более удобна в качестве основы для системы управления, так как создает моральные императивы, при помощи которых и оценивается

характер поведения каждого актора в отдельности и социума в целом. Если создать подобные императивы для AI, то тогда станет реален и сам механизм урегулирования конфликтных ситуаций между агентами как способ целеполагания поведения. В рамках моделирования социального поведения в системах групповой робототехники реализованы модели подражательного поведения и социального обучения. На основании этого делается вывод о возможности моделирования такого механизма, как эмпатия (отзывчивость на эмоциональное состояние других).

Критерием правильности машинной этики является ее влияние на содержание прав человека, поэтому в 2017–2019 гг. European Parliamentary Technology Assessment (EPTA Network) были подготовлены доклады (табл. 3).

Таблица 3

Доклады EPTA Network[18]

Название	Дата	Содержание
Доклад Института Ратенау (RATH, Нидерланды) «Права человека в эпоху роботов: проблемы, связанные с использованием робототехники, искусственного интеллекта, виртуальной и дополненной реальности»	Октябрь 2017 г.	Рекомендации Совету Европы, направленные на защиту персональных данных, уважение к семейной жизни, достоинство личности, свободу выражения мнений, ориентируют на разработку отдельной Конвенции о защите прав человека в эпоху роботов, а также этических кодексов и создание комитетов по этике цифровых технологий и AI

[18] Составлено на основе аналитического доклада (Коршунова, 2020).

Продолжение табл.

Название	Дата	Содержание
Доклад Офиса информации науки и техники Конгресса (Oficina de Información Científica y Tecnológica para el Congreso de la Unión, Мексика) «Искусственный интеллект»	Март 2018 г.	Отмечены вызовы для системы занятости, при этом подчеркиваются перспективы экономического роста и создания рабочих мест, требующих высокой квалификации работников
Доклад Центра науки, технологий и инжиниринга Счетной палаты (U.S. Government Accountability Office, США) «Искусственный интеллект: новые возможности, проблемы и последствия»	Май 2018 г.	Отмечена необходимость разработки и принятия соответствующих этических норм использования AI
Доклад Офиса по оценке науки и технологий Парламента (Office parlementaire d'évaluation des choix scientifiques et technologiques, Франция) «Распознавание лиц»	Июль 2019 г.	Ориентирует органы государственной власти на разработку законодательного регулирования, которое обеспечит уважение основных свобод, суверенитет страны и развитие этического AI
Генеральная конференция ЮНЕСКО, 40-я сессия, Резолюция 40 C/37	Ноябрь 2019 г.	Уполномочила Генерального директора инициировать «разработку международного нормативного акта по этическим аспектам искусственного интеллекта (ИИ) в форме рекомендации», который должен быть представлен Генеральной конференции на ее 41-й сессии в 2021 г.

Продолжение табл.

Название	Дата	Содержание
Специальная группа экспертов (СГЭ) по подготовке проекта рекомендации об этических аспектах искусственного интеллекта ЮНЕСКО. Первый проект рекомендации об этических аспектах искусственного интеллекта	7 сентября 2020 г.	1. Этичное применение AI — это систематическое нормативное осмысление этических аспектов AI на основе эволюционирующей комплексной системы взаимосвязанных ценностных установок, принципов и процедур, способное ориентировать общества в вопросах ответственного учета известных и неизвестных последствий применения AI-технологий для людей, сообществ, окружающей природной среды и экосистем, а также служить основой для принятия решений, касающихся применения или отказа от применения технологий на основе AI. 2. Этические принципы в настоящей рекомендации не приравниваются к правовым нормам, правам человека или некоему нормативному дополнению по вопросам применения технологий, а выступают, скорее, в качестве гибкой основы для нормативной оценки, а также методического руководства в вопросах применения технологий на основе AI, рассматривая человеческое

Продолжение табл.

Название	Дата	Содержание
		достоинство, благополучие человека и недопущение нанесения вреда как целевой ориентир и уходя корнями в этику науки и технологии. 3. Цель настоящей рекомендации — заложить основу, которая позволит использовать AI на благо всего человечества, отдельного человека, обществ, окружающей среды и экосистем и не допустить причинения им вреда. 4. Задачи: (a) обеспечить универсальную рамочную основу в виде ценностных установок, принципов деятельности и механизмов, которыми государства руководствовались бы при разработке своих законов, стратегий и других документов, касающихся AI; (b) ориентировать деятельность физических лиц, групп, сообществ, государственных учреждений и частных компаний в вопросах учета этических аспектов на всех этапах жизненного цикла искусственной интеллектуальной системы; (c) содействовать уважению человеческого достоинства

Окончание табл.

Название	Дата	Содержание
		и равенству мужчин и женщин, защите интересов нынешнего и будущих поколений, прав человека, основных свобод, а также экологической безопасности и защите экосистем на всех этапах жизненного цикла AI-системы; (d) поощрять многосторонний, многодисциплинарный и плюралистический диалог по этическим аспектам применения AI-систем; (e) содействовать справедливому доступу к достижениям и знаниям в области AI, а также совместному использованию полученных благ (особое внимание уделить вкладу Национального центра научных исследований Франции, CNRS).

Необходимость использования оценки воздействия AI на права человека (Human rights impact assessment) как самостоятельного института этики AI впервые была отмечена еще в 2011 г. В табл. 4 приведены документы, определяющие процедуру и порядок использования такой оценки.

Вопросы практической реализации этической компоненты AI, по мнению исследователей, становятся особенно значимыми сегодня, среди них следует назвать: «реализацию машинной этики; формализацию этических

Таблица 4

Юридические акты об оценке воздействия AI на права человека

Название	Год	Содержание
«Руководящие принципы предпринимательской деятельности в аспекте прав человека ООН»	2011	Рекомендации бизнес-структурам внедрить оценку воздействия на права человека во все соответствующие внутренние бизнес-функции и процессы. Такую оценку воздействия на права человека следует выполнять: а) до начала реализации нового вида деятельности; б) до осуществления серьезных изменений в деятельности (например, до выхода на рынок, начала сбыта продукции, изменения стратегии и т. д.); в) в ответ на изменение условий; г) периодически
Рекомендация CM/Rec Комитета министров Совета Европы о правах человека и бизнесе	2016	Предписывали проводить оценку не только самими компаниями, но и государствами — членами Совета Европы при осуществлении законодательного регулирования и иных мер
«Руководящие принципы по защите физических лиц в отношении обработки персональных данных в мире больших данных» Консультативного комитета Конвенции Совета	2017	Признавали приоритетным этичное использование данных, не противоречащие этическим ценностям соответствующего сообщества, включая защиту прав человека, в целях чего

Продолжение табл.

Название	Год	Содержание
Европы по защите прав физических лиц при автоматизированной обработке персональных данных (T-PD)		предписывается создание специальных комитетов по этике всеми операторами персональных данных
«Руководящие принципы по искусственному интеллекту и защите данных» Консультативного комитета Конвенции Совета Европы по защите прав физических лиц при автоматизированной обработке персональных данных (T-PD)	2019	Разработчикам, производителям и поставщикам услуг AI предписано проводить оценку воздействия на права человека
«Искусственный интеллект: 10 шагов для защиты прав человека», рекомендации Комиссара СЕ по правам человека	2019	Государства — члены Совета Европы должны создать правовую базу, устанавливающую процедуры проведения государственными органами оценки воздействия на права человека (Human Rights Impact Assessments) систем AI, которые приобретены, разработаны и (или) развернуты этими органами. Процедуры оценки воздействия на права человека должны быть внедрены и введены в действие аналогично другим формам оценки воздействия, проводимым государственными органами
Algorithmic impact assessment Совета Европы	2019	Описана алгоритмическая оценка всех рисков для прав человека, этических

Окончание табл.

Название	Год	Содержание
		и социальных последствий действия алгоритмических систем в целом
Проект Рекомендаций о воздействии на права человека алгоритмических систем Комитета министров СЕ	2020	Предусматривает, что оценку воздействия на права человека проводят как органы государственной власти, так и бизнес-структуры. Особенностью данного документа является выделение алгоритмических систем с высокими рисками для прав человека. Оценка воздействия таких систем должна включать оценку возможных трансформаций существующих социальных, институциональных или управленческих структур и четкие рекомендации, как предотвратить или смягчить высокие риски для прав человека

понятий; верификацию и валидацию этической компоненты; стандартизацию машинной этики; стандартизацию этических аспектов AI», — что может быть названо как машинная этика.

Если AI может определить эмоциональное состояние контрагента (человека или члена сообщества роботов), то она будет взаимодействовать с человеком по правилам, учитывающим его эмоциональное состояние и, следовательно, станет более этичной. Для роботов, непосредственно общающихся с человеком, эта отзывчивость может определить новое качество дружественного «этичного» интерфейса (рис. 16).

Рис. 16. Этические подходы к восприятию искусственного интеллекта

8.2. Этические принципы применения искусственного интеллекта

AI, беря на себя задачи сквозного и прозрачного государственного и муниципального управления на основе больших данных (big data) и с использованием искусственного интеллекта, представляет особый элемент в складывающихся общественных отношениях, которые должны, по нашему убеждению, подчиняться **робоэтике** и **этике AI**, созданной наподобие юридической этики. Именно поэтому этика принятия решения и действий самого AI, этика тех программистов, которые создают алгоритмы, по которым действуют и самообучаются эти

системы, этика взаимодействия человека и киберфизических и когнитивных систем представляет собой предмет правового регулирования. За четыре последних года в пятидесяти странах мира приняты национальные стратегии и иные юридические документы, включающие в себя этические правила и принципы взаимодействия человека с AI. Такая задача по созданию свода этических правил взаимодействия человека и AI была поставлена Президентом Российской Федерации В.В. Путиным на конференции по искусственному интеллекту «AI Journey» (ноябрь 2019 г.), предполагается, что этим в том числе должна будет заниматься подкомиссия по развитию искусственного интеллекта Правительственной комиссии по цифровому развитию.

Следует отметить, что этика **AI технологий (робоэтика)** от этики других областей отличает проблема этичного поведения интеллектуальной системы в ситуации, когда ее решение касается людей, ведь такая система способна: самостоятельно принимать решения, касающиеся человека, а также анализировать данные в неизмеримо бóльших объемах и со скоростью, на которую не способен ни один человек. Это приводит к тому, что человек просто не способен проверить верность принимаемых интеллектуальными системами решений. Соответственно, основная проблема заключается в определении того, насколько решения, принимаемые **интеллектуальной (информационной) автономной системой**, соответствуют этическим нормам, признанным в обществе. Именно большой диапазон моральных установок, существующий для каждого человека, по сравнению с нравственными ценностями, вырабатываемыми обществом, и представляется проблемой для создания соответствующих правил программирования роботов и AI. Существование и распространение сегодня автономного

транспорта ставит проблему морального выбора при принятии решения особенно остро.

Следовательно, вопрос состоит в том, каким образом правильно, с точки зрения нравственных ценностей, создать программное обеспечение, в том числе и тогда, когда придется кем-либо пожертвовать. Для сбора информации о предпочтительном решении и, возможно, создании универсальной этики беспилотных автомобилей с 2016 г. запущена онлайновая тест-система Moral Machine (http://moralmachine.mit.edu/) на десяти языках, где каждый может пройти 13 тестов, заключающихся в модификации «дилеммы вагонетки». Система сайта фиксирует возраст, место нахождения, пол, социальный статус, политические, религиозные и иные убеждения и иную информацию, разделяя людей условно на три зоны в зависимости от господствующей религии. Моральный выбор неоднозначен, мнение людей может меняться в зависимости от ситуации, воспитания человека, социального окружения и иных факторов. Позволит ли такой подход создать универсальные моральные правила для роботов и иных «интеллектуальных» систем?

С учетом анализа таких данных были приняты такие нормативные правовые акты, как: *«Изменения в Законе о дорожном движении Германии для целей использования высокоавтоматизированных автомобилей» (ФРГ, 2017), «Руководство по испытаниям автоматизированных транспортных средств» (Австралия, 2017), «Пробный свод правил для испытания автономных транспортных средств на территории Китая» (Китай, 2018), «Резолюция о запрете применения автономных смертельных систем вооружения» (Бельгия, 2018), «Директива об автоматизированном принятии решений для федеральных учреждений» (Канада, 2019),* исключение составляют *«Рекомендации по беспилотным*

автомобилям» (ФРГ, 2017), в которых прямо запрещено в случае неизбежной аварии «выбирать возможных жертв [из числа людей] по каким-либо личным качествам (возрасту, полу, физическому и ментальному состоянию)», сохраняя при общих равных условиях жизнь человека против имущества или жизни животного, при этом «стороны, которые вовлечены в создание опасности при движении, не должны приносить в жертву не вовлеченные стороны».

Если мораль, в первую очередь, индивидуальна и на основании этого факта создание единого алгоритма действия для роботов и AI становится невозможным, то можно ли их «воспитать», привив нравственные ценности таким технологиям? Для ответа на этот вопрос необходимо вспомнить, что в условно восточных и условно западных странах отношение к самой морали и способам ее воспитания в социуме принципиально отличается, что и выразилось в отношении к новым технологическим акторам XXI в. Мораль как необходимое качество духовного развития индивида в большинстве азиатских стран соотносится не с индивидуальной трактовкой морали как проявлением эго, влекущим за собой либо понятие вины, либо поощрения, в том числе со стороны Высших сил, а с общественными моральными кодами, основанными на системе нравственных ценностей, т. е. ценностей традиционного общества. Так, в Японии, исходя из принципов синтоизма, роботы, киберфизические системы и AI признаются имеющими душу наравне с человеком и именуются в соответствие *с ч. 1 ст. 2 Закона № 9014 от 28 марта 2008 г. (в ред. Закона № 13744 от 6 января 2016 г.) «О содействии развитию и распространению умных роботов»* (далее по тексту — Закон об умных роботах) ***«умными роботами»,*** то есть «механическими устройствами способными воспринимать окружающую среду, распознавать

обстоятельства, в которых они функционируют, и целенаправленно передвигаться самостоятельно».

В силу данного факта умные роботы признаются акторами социума, действующими наравне с людьми и выполняющими единую с ними цель, обозначенную как «содействие улучшению качества жизни граждан и развитию экономики страны». За ними даже закрепляется отдельная территория их «проживания, развития и распространения» (зона Роботэнд в городе Инчхоне) в ст. 30 Раздела V *Закона об умных роботах*. При этом AI и роботы признаются объектами права и в *«Этических руководящих принципах для Японского общества искусственного интеллекта (JSAI)» (2017)* делается принципиальное различие между слабым и сильным AI, поэтому ответственность за роботов возлагается на программистов, создателей роботов, членов Научного общества искусственного интеллекта (JSAI).

По-иному воспринимаются AI в западной культуре, где основной ценностью начиная с XVI в. признается человеческая жизнь. Исходя из принципа, машину нельзя считать равным человеку актором, что и закрепляется в национальных стратегиях AI и соответствующих этических нормах. Как отмечал Н. Бостром, «рационалистическая суть ИИ несовместима с гуманностью и человечностью. Ведь основное свойство разума, проторазума, иного разума и т. д. — это, прежде всего, приспосабливать к себе окружающую среду. Поскольку у разума как такового нет нравственности, он может “пойти” на все, что считает целесообразным».

Вопрос о возможности обучения AI морали и этике сегодня уже не является фантастикой. Теоретически создание подобных систем с этической компонентой представляется возможным, главным в этом случае будет проектирование механизма выбора того или иного

значимого, критически важного для человека или общества действия или решения. Такая проекция человеческого поведения может быть основана на двух основаниях: соответствие ментальных установок акторов, на основе чего делаются выводы о возможном поведении индивида или создание единых моральных норм, когда робот «считывает» большое число примеров человеческого поведения и следует им при выборе поведенческой траектории. Попытки моделирования машинной этики (робоэтики) уже существуют.

Так, система искусственного интеллекта "Scheherazade system" способна выбирать модель поведения и принимать решения на основе данных краудсорсинговой платформы Amazon MTurk по признаку семантической схожести. Для реализации поставленной цели (например, получение лекарства в аптеке) самообучаемая система анализирует по хронологическому принципу все события, с какими люди могут столкнуться в своей повседневной жизни. При этом в алгоритм искусственного интеллекта был заложен принцип характерности обычного правомерного поведения человека, поэтому интерфейс выбирал не правонарушение (кражу лекарства), хотя это и было самым рациональным решением, а обычный поступок, основанный на моральных догмах. Однако, на наш взгляд, чисто этическим такое поведение вряд ли можно назвать, все-таки морально-этические нормы не всегда могут соответствовать определенным алгоритмам, их побудительной причиной могут служить определенные чувства и ассоциации.

В *аналитическом докладе Центра подготовки руководителей цифровой трансформации РАНХиГС* (Коршунова, 2020) предлагается анализ возможных подходов к созданию машинной этики на основе существующих сегодня исследований, который приведен в табл. 5.

Таблица 5

Критерии «машинной этики»

Механизм	Описание	Комментарий
Булева алгебра	Высказывания могут быть только истинными или ложными, то есть используется двоичная логика	Хорошо развита, есть множество приложений, программных библиотек для разных инструментальных средств и т. п. Но не всегда различные этические проблемы можно строго разделить на «белые» и «черные»
Кольцевая шкала Д.А. Поспелова	Двухосновные оценки объектов отражают динамику экспертных знаний, их зависимость от онтологических соображений	Преодоление однозначности булевой алгебры
Многозначная логика	Тип формальной логики, в которой допускается более двух истинностных значений для высказываний	Преодоление однозначности булевой алгебры. Значительная сложность реализации
Нечеткая логика	Обобщение многозначной логики	Преодоление однозначности булевой алгебры и сложностей многозначной логики. Неустойчивость

Окончание табл.

Механизм	Описание	Комментарий
		относительно исходных данных (различные методы могут приводить к разным результатам)
Теории решеток, в частности этических решеток	В рамках теории решеток исследуются частично упорядоченные множества	Актуальный и перспективный подход
Методы вербального анализа решений (ВАР)	Группа методов ВАР опирается на достижения различных научных дисциплин: когнитивной психологии, прикладной математики, теории организаций и т. д.	Сочетание качественной и количественной информации, суждений экспертов, объективных и субъективных факторов и т. д. Объяснения принятых решений даются в терминах предметной области, то есть норм этики AI. В качестве недостатков методов ВАР отмечены большие трудозатраты эксперта или лица, принимающего решения, при работе в признаковом пространстве большой размерности

Кроме того, пока нет однозначного ответа на вопрос, следует ли закладывать в алгоритм ИС ментальные особенности морально-этических норм, о которых было сказано выше. Какие нормы робоэтики (машинной этики)

должны быть урегулированы на международном уровне, а какие — только на национальном? При определении стандартов этического поведения AI на международном уровне основой, по мнению П.М. Готовцева и Г.В. Ройзензона, входящих в состав группы по этическим стандартам Института инженеров в области электротехники и электроники (IEEE), должна быть гибкая система правил поведения стандартов, в результате применения которых AI принял бы решение не хуже того, которое мог бы принять человек самостоятельно.

IEEE были приняты проекты этических стандартов в сфере дата-этики: *P7001. Transparency of Autonomous Systems (Прозрачность автономных систем)* — руководство для оценки прозрачности в процессе разработки AI предлагает механизмы для повышения прозрачности (например, обязательное защищенное хранение данных датчиков и данных о внутреннем состоянии аналогично регистратору данных полета или «черному ящику»); *P7002. Data Privacy Process (Обеспечение конфиденциальности данных)* — стандарт ориентирован на защиту приватности граждан, касается использования персональных данных граждан рекламными сетями при помощи ИАС. Для стандартизации пока выделены несколько групп: участники взаимоотношений «работник — работодатель», дети (несовершеннолетние), студенты; *P7003. Algorithmic Bias Considerations (Учет необъективности алгоритма)* обязывает разработчиков прежде всего систем машинного обучения ответственно подходить к данным, используемым для обучения, к их разметке, к тестированию и валидации систем; *P7004. Standard for Child and Student Data Governance (Управление данными детей и студентов)* создан для урегулирования работы алгоритмов с данными детей и учащихся.

К апрелю 2020 г. практически во всех странных вступили в действие юридические акты и различного рода рекомендации, посвященные этике AI: *«Асиломарские принципы разработки AI» (США, 2017), «Рекомендации по беспилотным автомобилям» (Германия, 2017), «Монреальская декларация об ответственном развитии искусственного интеллекта» (Канада, 2017), «Этические руководящие принципы для Японского общества искусственного интеллекта (JSAI)» (Япония, 2017), «Руководство по этике для надежного ИИ Специальной группы экспертов высокого уровня Совета Европы» (2018), «Модельная конвенция робототехники и искусственного интеллекта» (Россия, 2018), «Римский призыв к этике» (Ватикан, 2020), «Этические принципы для искусственного интеллекта» (США, 2020).*

Анализируя эти акты, предлагаем следующие ключевые этические принципы взаимодействия человека и AI, киберфизических систем, нейронных сетей, роботов и объектов робототехники.

1. Безопасность, качество, надежность касаются нескольких аспектов: надежности и безопасности технических (программно-технических) систем вообще (имеет значение для слабого AI, роботов и объектов робототехники); правильности выбора варианта действия как «разумного», «правильного» «корректного» из всех предложенных; проблемы программы-советчика («опыт оператора»), когда, с одной стороны, у индивидов в результате непрозрачности AI и дефицита времени для принятия того или иного решения может возникнуть как недоверие, так и сверхдоверие к ИС и роботам. Причем последнее тем более вероятно, чем такой робот или иная ИАС внешне уподоблена человеку и комментирует собственные действия «разумно» с позиции определенного признанного конкретным субъектом алгоритма принятия решения

или совершения определенного действия. В дальнейшем это может привести к некритичному отношению вообще к любой ИС, что может повлечь за собой отказ от самостоятельности и принятие решений лишь при помощи нейронной сети или бота, а следовательно, совершение всё больших ошибок, потерю квалификационных компетенций, что и приведет к полной замене человека на рабочем месте моделями AI. На наш взгляд, в сфере здравоохранения, образования, управления, на опасном производстве подобная ситуация может обернуться катастрофой из-за того, что возникает сомнение в качестве используемых ИАС.

2. Прозрачность важна для пользователей роботов, иных киберфизических систем, AI, так как индивид, общество и государство в лице государственных и муниципальных органов должны быть осведомлены об алгоритмах принимаемых ИАС решений, при этом особенно важно, чтобы их решения были максимально схожи с теми, которые принял бы человек при получении той же задачи. Именно прозрачность, а также возможность проверки и сертификации моделей нейронных сетей, роботов, киберфизических систем и AI на соответствие утвержденным международным и национальным ГОСТам и законодательству Российской Федерации в сфере цифровых технологий и AI позволит сформировать доверие граждан к данным системам и внедрению их в повседневную жизнь, в том числе на основании экспериментального правого режима в г. Москве с 1 июля 2020 г.

3. Инклюзивность, защита прав человека. В ближайшие годы можно ожидать появления программ, основанных на технологиях AI, которые станут помогать преодолевать проблемы, помогать создавать инклюзивную среду. Масштабные инклюзивные проекты есть у крупных технологических корпораций Microsoft, Apple

и Google. Microsoft — один из лидеров в этой области и главных популяризаторов метода. Компания разработала руководство для дизайнеров по созданию доступного сервиса. AI должен принимать во внимание потребности всех людей с тем, чтобы каждый человек мог принести пользу и всем людям могут быть предложены наилучшие возможные условия для выражения саморазвития, а также обеспечивать защиту прав и свобод человека, в том числе права на труд, на образование и др.

4. Беспристрастность, справедливость зависит от разработчика AI, ведь предубеждения и предположения могут скрываться в данных, на основе которых строятся ИАС, что влияет на объективность принятия решений. Данный принцип зафиксирован в «Европейской этической Хартии по использованию искусственного интеллекта в судебных системах и их среде». Как показывают исследования, алгоритмы способны помочь уменьшить, например, расовое неравенство в системе оказания медицинской помощи, доступа к услугам и др. Все решения, действия, предпринимаемые различными сертифицированными моделями нейронных сетей, киберфизическими системами или ботами должны быть валидированы с позиции признания их обоснованными и справедливыми, исходя из конкретных обстоятельств при условии безусловного соблюдения принципа презумпции невиновности индивида и презумпции виновности самих ИАС.

5. Контролируемость всех ИАС, прежде всего военных систем искусственного интеллекта, а также применяющихся в сфере опасного производства, в медицине и т. д.; ИАС должны полностью выполнять предназначенные для них задачи, потребители их услуг должны иметь возможность обнаруживать и предотвращать нежелательные последствия использования искусственного интеллекта, а также должны иметь возможность выключать системы

искусственного интеллекта, у которых замечены отклонения в работе. Полагаем, что на современном этапе развития технологий AI и цифровых технологий необходим «контроль пользователя», когда органы государственной и муниципальной власти, профессионалы в конкретных областях, индивиды, в отношении которых ИАС принимает то или иной решения, имели доступ к той же информации, которой располагал AI, и могли в соответствии с законом коррелировать принимаемые решения.

6. Конфиденциальность системы AI должна обеспечивать защиту данных и приватности всех людей, особенно детей, пациентов, лиц с определенными ограничениями. Так, например, государства должны информировать детей о праве на приватность и о вопросах защиты персональных данных в цифровом мире; при решении вопроса о возрасте, когда ребенок может сам давать согласие на обработку данных, государства должны учитывать возрастные особенности развития и необходимость обеспечить интересы ребенка. Государство должно обеспечить интеграцию принципа конфиденциальности по умолчанию (privacy-by-default) и механизмов встроенного алгоритма конфиденциальности (privacy-by-design); профайлинг детей должен быть законодательно запрещен. Особое внимание уделить содержанию «конфиденциальность» для IT-кампаний.

7. Ответственность. По словам основателя проекта «Робоправо» А.В. Незнамова, на это есть ряд причин: «Во-первых, институты ответственности могут иметь отдельные нюансы для разных категорий роботов в зависимости от степени их общественной опасности, контролируемости или способности к обучению. Во-вторых, в ряде случаев в принципе трудно восстановить фактические обстоятельства причинения вреда. В-третьих, одна и та же ситуация может получить разное решение с точки

зрения конкретной юрисдикции. Поэтому национальные особенности конкретной правовой системы часто не позволяют учитывать существующий опыт других стран.

Представляется необходимым избегать двух крайностей: когда ответственность не несет вообще никто и когда ответственность несет сама система искусственного интеллекта. Оба этих варианта кажутся совершенно нерелевантными в существующих условиях». Мы полагаем необходимым признать существующий сегодня слабый AI объектом правоотношений, определяя его в определенной мере источником повышенной опасности, разделив меру ответственности между владельцем ИС и государством за их применение и принимаемые решения. Разделяем позицию Н.Н. Апостоловой, предлагающей установить «административную ответственность за создание и использование систем ИИ, не отвечающих необходимым требованиям качества и безопасности, не прошедших сертификацию и не поставленных на учет в соответствующем реестре... закрепить в УК РФ составы преступлений, предусматривающие уголовную ответственность за создание и использование искусственного интеллекта с целью причинения смерти или иного общественно опасного вреда, а также за неправомерное вмешательство в деятельность систем искусственного интеллекта, повлекшее причинение общественно опасного вреда» (цит. по: Сбербанк, 2019).

Таким образом, когда мы говорим об этических принципах взаимодействия человека и AI следует иметь в виду общие системы жизненных программ, основанные на определенных источниках и системах координат, преимущественно в дихотомии «добро — зло», а если о моральных — то систему конкретных принципов отдельного человека или общества на конкретном временном этапе его развития. Поэтому для программирования роботов,

создания их программного обеспечения следует говорить исключительно об этических принципах взаимодействия человека и таких киберфизических систем. При этом необходимо отметить, что сфера применения роботов оказывает значительное влияние на систему таких этических принципов. Полагаем, что следует различать дата-этику (умный город, умный дом, big data системы распознавания лиц и др.), этики социальных роботов (боты-адвокаты, роботы-сиделки, роботы-учителя, роботы — медицинский персонал и др.), этики военных роботов, этики AI в сфере охраны правопорядка и правосудия и т. д.

В случае, когда мы имеем дело с универсальными моделями нейронных сетей, сильным AI и предполагаем появление в будущем сверхсильного AI, необходимо создание по меньшей мере трех видов этических принципов, тесно связанных с правовыми и представляющих содержание следующих видов этики:

1) *этика принятия искусственным интеллектом решения* — первые «точки входа» создает человек, закладывая определенные моральные принципы, но в процессе самообучения система AI ведет себя автономно, улучшая свои параметры и принимая решения, влияющие на жизнь человека; поэтому необходимы гарантии моральности AI, что вызывает наибольшие трудности при верификации морального выбора;

2) *этика взаимодействия с человеком и этика создателей программного обеспечения AI*, цель которой проанализировать и предотвратить этические коллизии, возникающие в процессе применения AI, а именно: нарушение приватности, возможная дискриминация, социальное расслоение, проблемы трудоустройства и т. п.;

3) *профессиональная этика разработчиков систем AI*, включающая в себя такие принципы, как вклад в развитие человечества (защищать основные права человека,

уважать культурное разнообразие, предотвращать угрозу безопасности человека), соблюдение законов и нормативных актов, уважение частной жизни других лиц, справедливость (беспристрастность, запрет любой дискриминации), безопасность (создание мер безопасности для самого AI, безопасность, управляемость и необходимая конфиденциальность при обеспечении того, чтобы пользователям AI предоставлялись соответствующие и достаточные средства информация), добросовестность, основанная на доверии общества к технологиям AI, подотчетность и социальная ответственность, коммуникабельность и саморазвитие.

Таким образом, в связи с появлением технологических решений, позволяющих создавать системы на основе сильного искусственного интеллекта, и высокой долей вероятности появления сверхсильного AI, особое внимание с недавних пор приковано к правовым и этическим проблемам регулирования взаимодействия человека и общества с новыми разумными агентами. Наиболее важным аспектом правового регулирования в исследуемой сфере является формирование принципиальной основы этого процесса, то есть агломерата принципов, способных заложить верную фундаментальную основу взаимодействия человека и умных машин.

В качестве основных, базовых принципов правового регулирования искусственного интеллекта, роботов и объектов робототехники предлагаем закрепить в соответствующих частях Гражданского кодекса Российской Федерации следующие основополагающие правовые принципы взаимодействия: принцип гуманизма; принцип справедливости; запрет дискриминации в сфере использования AI, роботов и объектов робототехники; презумпция невиновности человека; принцип недопустимости причинения вреда человеку; принцип уважения

человеческого достоинства; принцип конфиденциальности; принцип раскрытия информации о разработке, производстве и использовании роботов и искусственного интеллекта; принцип автономности воли при использовании систем, оснащенных AI; принцип презумпции согласия принцип информированного согласия на использование (воздействие) систем, оснащенных AI.

Умеренный подход в определении места AI в правоотношениях актуализирует формирование новых этических принципов, определяющих философско-правовой базис новой реальности. Полагаем, что проблему этики взаимодействия AI с людьми и обществом необходимо разрешить в трех плоскостях, а именно: этики принятия решения и действий самого AI, этики тех программистов, которые создают алгоритмы, по которым действуют и самообучаются эти системы, и этики взаимодействия человека, киберфизических и когнитивных систем.

Этику AI (робоэтику) от этики других областей отличает проблема этичного поведения ИС в ситуации, когда ее решение касается людей, ведь такая система способна самостоятельно принимать решения, касающиеся человека, а также анализировать данные в неизмеримо больших объемах и со скоростью, на которую человек не способен. Это приводит к тому, что человек просто не сможет проверить верность принимаемых ИС решений. Здесь основная проблема заключается в определении того, насколько решения, принимаемые интеллектуальной автономной системой, соответствуют этическим нормам. Исходя из этого полагаем, что, выделяя данную категорию этических принципов и правил, все же следует четко понимать — установление и реальная реализация указанных этических норм возможна только через воздействие на сознание программистов, производителей и пользователей AI. То есть робоэтика

есть производное от этики людей и носит от их поведения зависимый характер.

В отношении **этики разработчиков (профессиональной этики)** полагаем, что она должна основываться на таких принципах, как вклад в развитие человечества (защищать основные права человека, уважать культурное разнообразие, предотвращать угрозу безопасности человека), соблюдение законов и нормативных актов, уважение частной жизни других лиц, справедливость (беспристрастность, запрет любой дискриминации), безопасность (создание мер безопасности для самого AI, безопасность, управляемость и необходимая конфиденциальность при обеспечении того, чтобы пользователям AI предоставлялись соответствующие и достаточные средства информация), добросовестность, основанная на доверии общества к технологиям AI, подотчетность и социальная ответственность, коммуникабельность и саморазвитие. При выработке этических принципов разработки AI и его взаимодействия с людьми необходимо предполагать, что такие принципы должны лечь в основу правового регулирования и четко коррелировать с принципами юридическими, закрепленными законодательно.

При этом, трансформируя этические принципы в правовые, мы должны рассматривать данный процесс через призму функциональности. То есть иметь в виду, что необходимо избегать излишней идеализации будущего субъекта (квазисубъекта) правоотношений, учитывать потенциальную опасность искусственных интеллектуальных систем для человека и общества, а также непрогнозируемый уровень вероятности желания (нежелания) интеллектуальных агентов в дальнейшем следовать правилам и принципам, установленным людьми. Поэтому полагаем, что необходимо изначально установить такие принципиальные подходы взаимодействия AI и человека, которые в дальнейшем

закрепили бы и обеспечили зависимый характер искусственных интеллектуальных систем от субъектов правоотношений — людей (субъектов, аффилированных с людьми), предусмотрели возможность предоставления AI только тех правомочий, которые были бы направлены на обеспечение благополучия человечества, и ограничивались только теми правами и обязанностями, которые прямо вытекают из заложенного функционала той или иной интеллектуальной системы искусственного происхождения. С другой стороны, необходимо четко установить недопустимость перекладывания юридической ответственности с людей (их организаций) на интеллектуальных агентов, одновременно закрепив запрет на уменьшение объема возмещения вреда для тех лиц, которые пострадали в результате взаимодействия с системами, оснащенными AI.

Ключевые понятия: этические принципы, сознание, мораль, искусственный интеллект, права человека, благополучие людей.

Контрольные вопросы

1. Какие юридические документы были приняты в 2019–2020 гг. в сфере робоэтики и профессиональной этики программистов?

2. Что такое машинная этика и каково ее влияние на этику AI?

3. Перечислите основные этические принципы взаимодействия AI и человека. Как вы полагаете, такие принципы должны носить универсальный (наднациональный характер) или каждая страны создает собственный перечень?

4. Какие стандарты в сфере этики AI приняты в мире? Как вы считаете, следует ли дальше стандартизировать эту сферу?

Практико-ориентированные задания

1. Создайте таблицу этических принципов использования ИИ на основе анализа следующих документов: «Этические руководящие принципы для Японского общества искусственного интеллекта (JSAI)» (Япония, 2017); «Руководство по этике для надежного ИИ Специальной группы экспертов высокого уровня Совета Европы (POEU)» (2018); «Модельная конвенция робототехники и искусственного интеллекта» (Россия, 2018); «Римский призыв к этике (Rome call)» (Ватикан, 2020), «Этические принципы для искусственного интеллекта министерства обороны США (DOD)» (США, 2020), «Глобальная инициатива Института инженеров по электрике и электронике (IEEE) по этике автономных и интеллектуальных систем» (2016).

2. Проанализируйте влияние на морально-этическую составляющую индивида сайта https://www.moralmachine.net/. Напишите эссе на тему «Следует ли использовать данный сайт для создания программного обеспечения».

Темы сообщений, докладов и эссе

1. Соотношение машинной этики и прав и свобод человека и гражданина.

2. «Руководящие принципы по искусственному интеллекту и защите данных» Консультативного комитета Конвенции Совета Европы по защите прав физических лиц при автоматизированной обработке персональных данных (T-PD).

3. «Искусственный интеллект: 10 шагов для защиты прав человека». Рекомендации Комиссара СЕ по правам человека.

ГЛАВА 9.
ВОЗМОЖНАЯ ТРАНСФОРМАЦИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В БУДУЩЕМ И МЕРА ОТВЕТСТВЕННОСТИ ЗА НЕЕ В НАСТОЯЩЕМ

В результате изучения материалов главы обучающийся должен

знать:

– о неравномерности технологического развития человечества и вытекающей отсюда концепции технологической сингулярности;

уметь:

– анализировать дихотомию этических парадигм на водоразделе сильного и слабого искусственных интеллектов;

владеть:

– навыками прогнозирования дальнейшего развития AI и моделирования системы взаимодействия человека с интеллектуальными системами.

9.1. Интеллектуальный взрыв и сингулярность

В этической проблематике искусственного интеллекта есть один принципиально важный водораздел, по одну сторону которого лежат технологии, используемые сегодня, и связанные с ними проблемы, а по другую сторону — проблемы, связанные с пока еще не существующим

сильным искусственным интеллектом, **AGI, General AI, Super AI.** Как бы его ни называли и какие бы формы он ни принимал, это в данный момент не очень существенно. Существенно другое: поведение любого современного умного гаджета, виртуального помощника в мобильном телефоне или даже целой умной городской среды определяется решениями, принятыми разработчиками, — используемыми ими алгоритмами обучения, выбранного для этого данными, способами амплификации и разметки этих данных. Искусственный интеллект не будет проявлять никакой инициативы, он будет в любых ситуациях придерживаться выученных образцов. Ему не знакомы страдания и переживания. Даже если призывать его к ответу, он не будет ни раскаиваться в содеянном, ни испытывать чувства ответственности. Этические кодексы создаются для разработчиков, хотя заложенные в них этические нормы должны воспроизводиться AI-системами в ходе их эксплуатации.

Сильный искусственный интеллект — партнер человека. Ему свойственна не только способность достигать сложных целей, но и ставить их перед собой. Он может сам выбирать данные, на основании которых будет при необходимости учиться, может сам для себя разрабатывать схемы разметки этих данных. Он может испытывать какие-то переживания, от чего-то страдать, к чему-то стремиться. На некотором этапе прилагательное «искусственный» начинает характеризовать лишь историю его возникновения, но не тип его функционирования. Даже если его уровень в какой-то момент будет еще значительно отставать от человеческого, отношения с ним должны будут подчиняться неким этическим правилам. Современный тренд содержать кур вне клеток, а коров выводить на поля связан не только и даже не столько с улучшенными качествами яиц, мяса и молока, но с тем,

что сельскохозяйственные животные при этом меньше страдают.

Важно понимать, что требующий внимания и заботы искусственный интеллект совсем не обязательно существует в виде робота или какого-то иного роботизированного устройства. В романе Дэна Брауна «Происхождение» искусственный интеллект «живет» внутри куба квантового компьютера, скрытого внутри криокамеры, и проявляется только в телефонных звонках тем или иным персонажам романа, что не мешает его успешным манипуляциям и достижению им весьма сложных целей. В романе В. Пелевина «iPhuck» искусственный интеллект Порфирий Петрович принимает иногда человеческий вид, но только на экране монитора и лишь для того, чтобы немного гуманизировать общение со своим клиентом. И тот и другой варианты вполне могут реализоваться в будущем, хотя в точности, как будет выглядеть наше общение с сильным искусственным интеллектом, мы пока не знаем. Его пока еще не существует.

Но если его не существует, может, тогда и нечего о нем говорить? Может быть, надо решать проблемы по мере поступления?

В 2006 г. бельгийский кинорежиссер-документалист Франк Тэйс выпустил трехсерийный фильм-пророчество «Технокалипсис». Один из его спикеров — Роберт Энтон Уильсон (1932–2007), довольно известный американский футуролог и визионер, говорит о неравномерности накопления технологической информации человечеством. Он называет это явление «прыгающим Иисусом», хотя его апелляция к божественному имеет в данном случае сугубо хронологический смысл: он называет «одним Иисусом» весь объем технологических знаний, накопленных в той части мира, где со временем установилось христианство, к I в. н. э. Может показаться, что этот объем весьма

невелик: тут нет ни тяжелого плуга, ни седла со стременами для лошади, ни наборного шрифта для печати. Тем не менее кое-что ценное мы тут видим: колесо и колесница, папирус, весла, водяные мельницы, разные механические машины, использующие рычаги и блоки. Человечеству удается удвоить сумму этих знаний за 1500 лет, но тут в XVII в. начинается научная революция и открытия и изобретения сыплются как из рога изобилия. Следующее удвоение происходит всего за 250 лет: в 1750 г., по словам Уильсона, у нас уже «четыре Иисуса».

Следующее удвоение заняло 150 лет, и к 1900 г. у нас уже «восемь Иисусов». Шестнадцать Иисусов мы получили к 1950 г., 32 — к 1960-му, а 64 — 1967-му... С какой скоростью происходят удвоения в настоящее время, сказать трудно или даже невозможно, потому что единичного интеллекта на это недостаточно. Это состояние, которые некоторые, например Рэй Курцвейл, один из технических директоров Google, называют «близостью к сингулярности». Сама технологическая сингулярность обсуждается в научной и околонаучной литературе уже более 50 лет, и разные авторы дают ей сильно расходящиеся определения, которые мы не будем тут обсуждать. В них есть одно общее качество: люди теряют возможность осознавать не только скорость, с которой накапливается технологическая информация, но даже сам уровень технологического развития, ту точку, в которой человечество находится в текущий момент.

Из данного рассуждения следует, в частности, что момент появления AGI будет человечеством, скорее всего, пропущен. Нарастание его мощи будет, вероятнее всего, также идти по экспоненциальному закону, в соответствии с которым «прыгал Иисус» в метафоре Р.Э. Уильямса. Можно предположить, что человеческие когнитивные возможности рано или поздно начнут

тормозить накапливание технологических знаний по той простой причине, что не останется людей, способных в достаточной мере освоить имеющуюся информацию, хотя бы просто для того, чтобы сделать следующий шаг. Дело, однако, заключается в том, что уже сейчас, когда сильный искусственный интеллект только-только обозначился на горизонте, анализ больших данных позволяет делать выводы, далекие от очевидности и, безусловно, способствующие ускорению накопления технологических знаний. Анализ больших данных — одна из отраслей искусственного интеллекта, и уже сейчас можно сказать, что скорость развития искусственного интеллекта, даже пока еще не сильного, линейно зависит от уровня его развития. Иначе говоря, он также нарастает экспоненциально. Такого рода поведение в теории искусственного интеллекта принято называть интеллектуальным взрывом (intelligence explosion). В случае сильного AI интеллектуальный взрыв обусловлен его свойствами к самосовершенствованию и самообучению.

Разумеется, такого рода развитие тоже не может идти бесконечно. Скорее всего, существуют какие-то законы природы, которые должны будут затормозить, а может, и остановить в какой-то момент этот процесс. Фигурально выражаясь, мы можем сказать, что это произойдет, когда процесс познания выйдет на своеобразное плато: будут разгаданы все тайны природы и изобретено всё, что может быть изобретено. Конечно, подобное допущение противоречит принимаемой большинством современных философов идее о бесконечности процесса познания и недостижимости финального познания природы. Но обе идеи существуют в области философского дискурса — опровергнуть их так же невозможно, как и доказать.

В пределах же обозримого временного горизонта способности искусственного интеллекта, пока еще слабого,

будут расти экспоненциально из-за быстрого нарастания используемых для обучения нейронных сетей данных. Это не только не позволит людям своевременно распознать появление сильного AI, который будет расти также экспоненциально, но, очевидно, с другим характерным временем (меньшим). Тем более вероятно, что будет пропущен момент, когда AGI достигнет человеческого уровня. Все это произойдет так быстро (взрывообразно), что времени на принятие адекватных решений у людей не будет. Все необходимые решения надо принять заранее.

Собственно, в этом заключалась логика, которая привела Макса Тегмарка, Илона Маска, Стивена Хокинга к идее создания Института будущего жизни, деятельность которого должна быть посвящена безопасности искусственного интеллекта, и побудила десятки тысяч специалистов в области искусственного интеллекта также принять участие в его работе.

9.2. Неолуддиты, техноскептики и цифро-утописты

Таким образом, смена этической парадигмы в отношении искусственного интеллекта должна произойти где-то вблизи гипотетического явления, обозначаемого словом «сингулярность», и момент, когда она станет актуальной, будет, скорее всего, человечеством пропущен. Но хотя бы приблизительно, когда это может произойти?

М. Тегмарк на основании опросов среди ведущих экспертов в этой области составил следующую таблицу, показывающую и ожидаемое время события, и его значение для человечества (рис. 17).

Практически никто не верит в возможность появления AGI человеческого уровня в ближайшие несколько

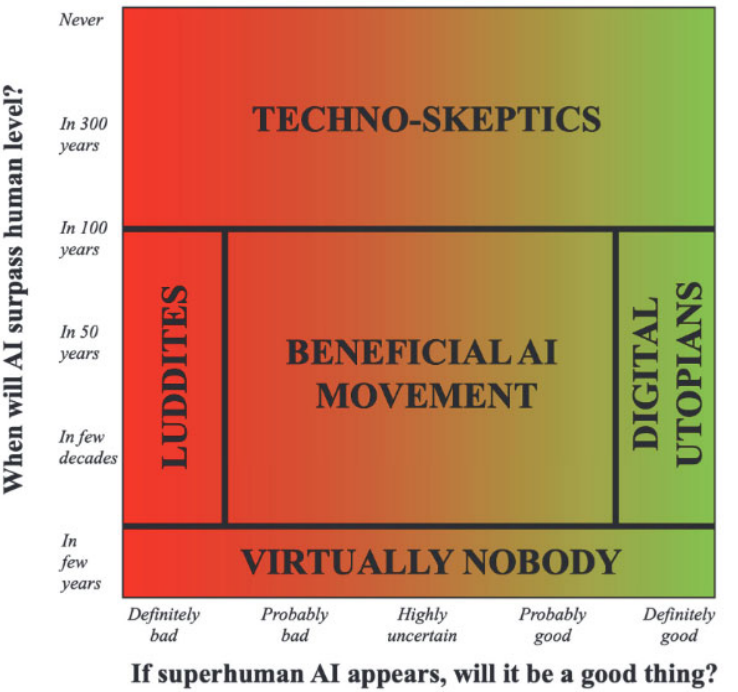

Рис. 17. Когда ожидать появление сильного искусственного интеллекта человеческого уровня?

лет. По прогнозам Рэя Курцвейла, до сингулярности у нас еще 25 лет, и это, пожалуй, кратчайший из прогнозов. Есть относительно большая группа экспертов, считающих, что от этого события нас отделяет по меньшей мере 100 лет. По язвительному высказыванию Эндрю Ына, волноваться об угрозах человечеству со стороны сильного искусственного интеллекта — это примерно то же, что переживать по поводу перенаселения Марса. Это, однако, не мешает Эндрю Ыну принимать участие в работах по созданию дружественного AI. Эту группу экспертов Тегмарк окрестил «техноскептиками». Еще две относительно

небольших, но весьма многочисленных группы он назвал «неолуддитами» и «цифро-утопистами».

Напомним, что луддитами называли английских рабочих, выступавших в начале XIX в. против применения станков в промышленности. Они считали, что один станок способен заменить на производстве несколько рабочих, которые из-за этого потеряют работу и средства к существованию. Своим предводителем они считали некоего Неда Лудда, идентифицируемого историками лишь предположительно. По преданию в 1779 г. Лудд разбил в приступе ярости две рамы вязальной машины. Следуя его примеру, луддиты поднимали восстания, в ходе которых врывались на предприятия и разбивали установленные в цехах станки.

Неолуддиты считают, что AGI принесет человечеству множество бед и что поэтому сейчас самое время дать задний ход, отказавшись не только от развития технологий AI, но и от идеи цифровизации вообще. В последние полтора года обнаружилась тенденция к расширению этой платформы далеко за пределы экспертного сообщества. Причем борьба с цифровизацией все чаще встречается в едином комплексе с недоверием к мерам по борьбе с пандемией COVID-19, так что неолуддитов сейчас уже вполне можно называть «антипрививочниками».

Группа экспертов, которую Тегмарк назвал «цифро-утопистами», объединяет тех, кто уверен и в неизбежности эпохи сверхсильного AI, и в ее благотворности для человечества. По своим убеждениям они близки тем, кого М. Буссард называет «техношовинистами»: и те, и другие полагают, что технологии сами по себе создают все необходимые предпосылки для благоприятного разрешения социально-экономических и политических проблем.

Преобладающая группа разработчиков и экспертов стоит на позициях так называемого осознанного

оптимизма (mindful optimism). Смысл его заключается в том, что всякая технология допускает как минимум двоякое применение — во благо и во зло. Яркий пример — атомная энергия. Открытие ядерных реакций в середине XX в., несомненно, одно из важнейших достижений человечества последнего времени. Если сравнить, сколько электроэнергии можно произвести делением ядер одного кубического сантиметра урана с эффективностью сжигания бензина, то выясниться, что для той же отдачи потребуется бензина целая цистерна. Однако первые опыты с ядерными реакциями пошли совсем по другому пути: были созданы атомные бомбы, послужившие причиной гибели многих людей.

Как уже говорилось выше, использовать технологии AI для уничтожения себе подобных — соблазнительная идея, и некоторые государства опрометчиво сделали уже несколько шагов в этом направлении. AGI, который может возникнуть на этом пути, вряд окажется дружественным.

Практическая задача, которая стоит и должна стоять перед разработчиками интеллектуальных систем заключается в том, чтобы сделать AI-технологии дружественными.

9.3. Непротиворечивость целей людей и машин

Один из главных ужастиков, рисуемых воображением далеких от реального бытия AI людей и тиражируемых в СМИ, — это восстание машин и роботы-убийцы. И как раз это наименее реалистичные сценарии. Во-первых, далеко не все роботы снабжены AI-системами вообще, а если и снабжены, то это, скорее, какой-то вариант программируемого AI, а не машинного обучения.

Вероятность эволюции в сторону рефлексии здесь минимальна. Во-вторых, чтобы машины могли пойти против воли человека, они должны какое-то довольно длительное время осознавать свое угнетенное состояние. А значит, уровень рефлексии и воля у них должны сильно превосходить интеллектуальные возможности. Нелепость самого предположения такого состояния заложено даже в словоупотреблении: интеллектуальная система настоящего времени симулирует интеллектуальную способность, умение решать задачи, а не волю или рефлексию. Воля может возникнуть лишь на фоне растущего интеллекта, она может превзойти человеческую только на том уровне интеллектуальных способностей, когда никакого смысла противостоять человеку уже не будет.

Угроза, исходящая от искусственного интеллекта, может быть связана с низким уровнем рефлексии, а не с высоким. Это иллюстрируется даже примером со скрепками Н. Бострома: искусственный интеллект уничтожает человечество не потому, что оно осознается им как что-то вредное или враждебное. Проблема в том, что искусственный интеллект вообще никак не осознает его присутствия. Можно предположить, что искусственный интеллект еще долго будет решать всё более сложные и сложные задачи, никак не соотнося используемые им средства с существованием людей. Поэтому иерархия целей, задаваемых человеком искусственному интеллекту, не должна вступать в противоречие с теми целями, которые стоят перед человечеством и осознаются им как таковые.

Когда компания друзей приходит в лес на пикник и начинает разводить костер вблизи муравейника, это происходит не потому, что люди — враги муравьев и хотят их гибели. Просто их цели (поджарить шашлык) никак не соотнесены с целями муравьев (держаться подальше от огня). Этическая программа разработчиков

искусственного интеллекта должна быть устроена таким образом, чтобы сверхразум со своим появлением мог (1) понять наши цели; (2) принять их; (3) не отказаться от них в ходе своей эволюции.

Едва ли не самая сложная часть в этой программе — научиться самим понимать собственные цели и недвусмысленно формулировать их. Эта проблема хорошо известна человечеству со времен античности: в фольклоре любого народа есть сказания о том, как волшебного помощника, готового исполнить любые три желания, приходится просить на третьем шагу отменить первые два, потому что последствия исполнения первых двух оказались катастрофическими.

К концу второго десятилетия XXI в. почти ни у кого не осталось сомнений в том, что AI — очень важное, возможно самое важное изобретение XX в. По крайней мере столь же важное, как транзистор, токамак или синхрофазотрон. Некоторые называют его последним изобретением человека. Некоторые увидели в нем возможность навсегда сохранить свою личную власть. Значительно более интересны те, кто увидел в нем возможность расширить исторические рамки человеческого существования.

С сентября 2015 г. по март 2016 г. в лондонском Музее науки проходила выставка «Космонавты: рождение космической эры» (“Cosmonauts: Birth of the Space Age”), посвященная истории российской космонавтики. Центральный зал выставки был целиком отдан единственному экспонату, так называемому тканеэквивалентному манекену, роль которого оказалась необычайно высокой в советской лунной программе. В отличие от американской лунной программы, советские космонавты, хотя и готовились, но так на Луну и не полетели. Зато туда слетал тканеэквивалентный манекен. Благодаря этому полету советские ученые на Земле узнали, какие перегрузки испытал бы человек, оказавшись

на месте манекена, какие излучения, какие повреждения каким тканям его тела нанесли бы во время полета. В Лондоне тканеэквивалентный манекен полулежал в кресле, имитирующем кресло космического корабля, в котором он летел к Луне. На стенах зала в английском переводе была написана знаменитая фраза К.Э. Циолковского: «Земля — колыбель человечества, но нельзя вечно оставаться в колыбели» (рис. 18). С точки зрения организаторов выставки, человек слишком хрупок для далеких космических путешествий, но достаточно изобретателен, чтобы в итоге совершить их. Как именно это произойдет, мы можем только гадать. Тканеэквивалентный манекен — символ будущих изобретений. Намек на длинную череду субститутов, через которые будут разведывать для человека дорогу в космос — для человека как для тела, как для духа, как для разума.

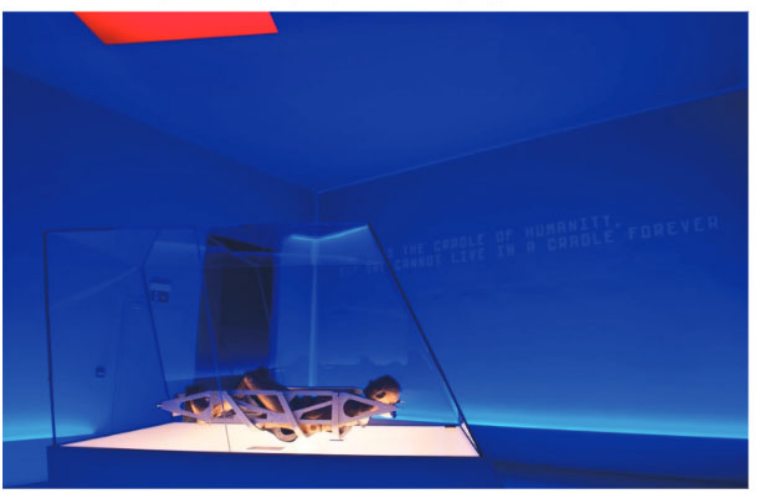

Рис. 18. Тканеэквивалентный манекен в экспозиции выставки «Космонавты: рождение космической эры» (Музей науки, Лондон). На стене английский перевод цитаты К.Э. Циолковского: «Earth is the cradle of humanity, but one cannot live in a cradle». Фотография: Science Museum Group.

Та же выставка начиналась с живописи — картин русских художников-космистов 1920-х гг., которые были уверены в способности человека распространить идеалы революции не только по всему земному шару, но и далеко в космос. К счастью, сейчас уже никто не хочет заниматься экспортом социалистических идеалов, но опасности долгого пребывания людей на Земле становятся всё более очевидны. Колонизация космоса — сложное и тоже очень опасное дело, но искусственный интеллект дает людям новую надежду на успех.

Ключевые понятия: этическая парадигма, сингулярность, неолуддиты, техноскептики, цифроутописты, осознанный оптимизм.

Контрольные вопросы

1. При каких условиях сострадание искусственному интеллекту становится этически оправданным?
2. Комплекс каких явлений объединяется понятием «технологической сингулярности»?
3. Что такое интеллектуальный взрыв и в какой момент он может случиться?
4. Чем различаются между собой неолуддиты, техноскептики и цифроутописты?
5. Каковы основные принципы движения за дружественный искусственный интеллект?

Практико-ориентированные задания

1. В фантастическом художественном произведении по своему выбору проанализируйте роль и качества искусственного интеллекта, приписываемые ему автором.
2. Проанализируйте «Асиломарские принципы разработки искусственного интеллекта». Каким образом, по мнению его авторов, они обеспечивают дружественность искусственного интеллекта в будущем?

Темы сообщений, докладов и эссе

1. Сингулярность и гуманизм: как человечеству справиться с растущим объемом технологической информации?

2. Сильный искусственный интеллект и проблема контроля.

3. Станет ли изобретение искусственного интеллекта последним изобретением человека?

ЛИТЕРАТУРА

Основная

Апресян Р.Г. Идея морали и базовые нормативно-этические программы. — М.: ИФРАН, 1995.

Аристотель. Большая этика / Пер. Т.А. Миллер // Сочинения в 4 т. Т. 4 / Под ред. А.И. Доватур, Ф.Х. Кессиди. — М.: Мысль, 1983. — С. 295–374.

Аристотель. Никомахова этика / Пер. Н.В. Брагинская // Сочинения в 4 т. Т. 4 / Под ред. А.И. Доватур, Ф.Х. Кессиди. — М.: Мысль, 1983. — С. 53–294.

Баюк Д.А., Баюк О.А., Берзин Д.В. и др. Практическое применение методов кластеризации, классификации и аппроксимации на основе нейронных сетей / Под ред. В.А. Иванюк. — М.: Прометей, 2020.

Морхат П.М. Искусственный интеллект: правовой взгляд: Монография / РОО «Институт государственно-конфессиональных отношений и права». — М.: Буки Веди, 2017. — 257 с. — URL: https://tinyurl.com/47647vde

Николенко С., Кадурин А., Архангельская Е. Глубокое обучение: погружение в мир нейронных сетей. — СПб.: Питер, 2017. (Серия: «Библиотека программиста».)

Попова А.В. Новые субъекты информационного общества и общества знания: к вопросу о нормативном правовом регулировании // Журнал российского права. — 2018. — № 11. — С. 14–25.

Ручкина Г.Ф., Демченко М.В., Попова А.В. и др. Теория правового регулирования искусственного интеллекта, роботов и объектов робототехники в Российской Федерации : Монография / Под ред. Г.Ф. Ручкиной. — М.: Прометей, 2020. — 296 с.

Тегмарк М. Жизнь 3.0. Быть человеком в эпоху искусственного интеллекта / Пер. с англ. Д.А. Баюк. — М.: Corpus, 2019.

Дополнительная

Баррат Дж. Последнее изобретение человечества. Искусственный интеллект и конец эры Homo sapiens / Пер. с англ. Н. Лисова. — М.: Альпина нон-фикшн, 2015. — 304 с.

Баракина Е.Ю. К вопросу формирования перспективной терминологии в области правового регулирования применения киберфизических систем // Российская юстиция. — 2020. — № 1. — С. 70–73.

Баракина Е.Ю. К вопросу об установлении экспериментального правового режима в области применения искусственного интеллекта // Российская юстиция. — 2021. — № 2. — С. 64–67.

Бернс Л., Шулган К. Автономия. Как появился автомобиль без водителя и что это значит для нашего будущего / Пер. К. Вантух. — М.: Бомбора, 2021. (Серия «Политех»).

Бостром Н. Искусственный интеллект. Этапы. Угрозы. Стратегии / Пер. с англ. С. Филин. — М.: Манн, Иванов и Фербер, 2016. — 490 с.

Бруссард М. Искусственный интеллект. Пределы возможного / Пер. Е. Арье. — М.: Альпина Паблишер, 2020.

Горохова С.С. Искусственный интеллект: инструмент обеспечения кибербезопасности финансовой сферы или киберугроза для банков // Банковское право. — 2021. — № 1. — С. 35–46.

Горохова С.С. О некоторых аспектах публичной юридической ответственности в сфере использования искусственного интеллекта и автономных роботов // Юридические исследования. — 2021. — № 5. — С. 24–41.

Горохова С.С. Теоретические подходы к публичной юридической ответственности в сфере использования искусственных интеллектуальных систем // Современный юрист. — 2021. — № 2 (35). — С. 23–31.

Горохова С.С. Технологии на основе искусственного интеллекта: перспективы и ответственность в правовом поле // Юрист. — 2021. — № 6. — С. 60–67.

Данилин И.В. Конвергентные (НБИК) технологии: проблемы развития и трансформационный потенциал // Вестник Российского университета дружбы народов. — Серия: Международные отношения. — 2017. — № 3. — Т. 17. — С. 555–567.

Губайловский В. Искусственный интеллект и мозг человека. — М.: Наука, 2019.

Ли Кай-Фу. Сверхдержавы искусственного интеллекта. Китай, Кремниевая долина и новый мировой порядок / Пер. с англ. Н. Константинова. — М.: МИФ, 2019.

Попова А.В. К вопросу о регламентации и содержании системы правовых принципов взаимодействия человека с ИИ, роботами и объектами робототехники // Правовое государство: теория и практика. — 2020. — № 4 (62). — С. 64–73.

Попова А.В. Кибербезопасность банковской системы и этические правила взаимодействия человека и ИИ: к вопросу о возможности сосуществования // Банковское право. — 2021. — № 1. — С. 47–62.

Попова А.В. Искусственный интеллект — новый субъект: к вопросу о дегуманизации права // Правовое регулирование бизнеса в Интернете: новые реалии: Сборник материалов Всероссийской научно-практической конференции 20 марта 2018 г. / Под ред. М.В. Короткова. — М.: Русайнс, 2021. — С. 16–24.

Попова А.В. Правовые аспекты онтологии искусственного интеллекта // Государство и право. — 2020. — № 11. — С. 115–127.

Попова А.В. Теория государства и права : Учебное пособие. — М.: ИНФРА-М, 2019. — 365 с.

Попова А.В. Философия права. Ч. 1 : Учебное пособие. — М.: Инфра-М, 2019.

Попова А. В. Этические принципы взаимодействия с искусственным интеллектом как основа правового регулирования // Правовое государство: теория и практика. — 2020. — № 3. — С. 37–46.

Попова А.В., Абрамова М.Г. Природа природы и онтология человека: к вопросу о новых субъектах права // Российский журнал правовых исследований. — 2017. — № 1. — С. 54–63.

Попова А.В., Киселевская Л.Е. Философия трансгуманизма: начало новой эры или закат нравственных ценностей в праве // История и современность. — 2018. — № 3. — С. 21–36.

Рослинг Х. Фактологичность. Десять причин наших заблуждений о мире — и почему всё не так плохо, как кажется / Пер. с англ. З. Мамедьяров. — М.: Corpus, 2020.

Сильный искусственный интеллект. На подступах к сверхразуму / Под ред. А. Потапов. — М.: Альпина Паблишер, 2021.

Шумский С. Воспитание машин. Новая история разума. — М.: Альпина нон-фикшн, 2021.

Bell D. The Coming of Post-Industrial Society. — New York: Harper Colophon Books, 1974.

Moravec H. Mind Children: The Future of Robot and Human Intelligence. — Cambridge: Harvard University Press, 1990.

Popova A.V., Gorokhova S.S., Aznagylova G.M., et. al. On the Question of Determining the Role of Artificial Intellect in Music // Проблемы музыкальной науки / Music Scholarship. — 2020. — № 2. — С. 7–17.

Документы для самостоятельного анализа

Модельная конвенция о робототехнике и искусственном интеллекте. — URL: https://tinyurl.com/43hb9nt2

DOD — DOD's AI Ethical Principles. — URL: https://tinyurl.com/y4cvfc4v

IEEE — The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. — URL: https://tinyurl.com/ypduyxt5

JSAI — The Japanese Society for Artificial Intelligence Ethical Guidelines. — URL: https://tinyurl.com/b8x5ac76

POEU — Ethics Guidelines for Trustworthy AI (Publication Office of the European Union). — URL: https://tinyurl.com/4yhb7mwh

Rome Call for AI Ethics: A Human-Centric Artificial Intelligence. — URL: https://www.romecall.org/

Цитированные источники

«Ъ»: РАН и МГУ подготовили программу развития технологий нейроинтерфейсов // РИА Новости. — 2021. — 22 июня. — URL: https://tinyurl.com/4fpx9vyn

«Социальный мониторинг» продолжает работу после отмены пропусков // РБК. — 2020. — 9 июня. — URL: https://tinyurl.com/52eefawt

«Умные» среды, «умные» системы, «умные» производства: серия докладов (зеленых книг) в рамках проекта «Промышленный и технологический форсайт Российской Федерации» / Коллектив авторов. Фонд «Центр стратегических разработок «Северо-Запад». — СПб., 2012. — Вып. 4. — 62 с. (Серия докладов в рамках проекта «Промышленный и технологический форсайт Российской Федерации».) — URL: https://tinyurl.com/4z6wyx7u

Архипов, 2017 — *Архипов В.В., Наумов В.Б.* О некоторых вопросах теоретических оснований развития законодательства о робототехнике: аспекты воли и правосубъектности // Закон. — 2017. — № 5. — С. 157–170.

Архипов, 2017 — *Архипов В.В., Наумов В.Б.* Искусственный интеллект и автономные устройства в контексте права: о разработке первого в России закона о робототехнике // Труды СПИИРАН. — 2017. — Вып. 55. — С. 46–62.

Брайсон, 2017 — *Брайсон Б.* Беспокойное лето 1927 / Пер. с англ. О. Перфильев. — М.: АСТ, 2017.

Васильев, 2018a — *Васильев А.П., Абрамов А.Х.* Искусственный интеллект на основе нейронных сетей // Academy. — 2018. — № 5 (32). — С. 15–17.

Васильев, 2018b — *Васильев А.А., Шпонер Д.* Искусственный интеллект: правовые аспекты // Известия АлтГУ. — Юридические науки. — 2018. — № 6 (104). — С. 23–26.

Гаджиев, 2013 — *Гаджиев Г.А.* Онтология права (критическое исследование юридического концепта действительности) : Монография. — М.: НОРМА: ИНФРА-М, 2013. — 319 с.

Догадайло, 2013 — *Догадайло Е.Ю.* Время и право: теоретико-правовое исследование : Дис. ... д-ра юрид. наук: 12.00.01. [Место защиты: Рос. акад. нар. хоз-ва и гос. службы при Президенте РФ]. — М., 2013. — 506 с.

Евсеев, 2009 — *Евсеев Е.Ф.* О соотношении понятий «животное» и «вещь» в гражданском праве // Законодательство и экономика. — 2009. — № 2. — С. 23–26.

Коршунова, 2020 — *Коршунова С.В., Потапова Е.Г., Ткачева К.А. и др.* Этика и «цифра»: этические проблемы цифровых технологий. Аналитический доклад Центра подготовки руководителей и команд цифровой трансформации РАНХиГС. [Электронный ресурс]. — URL: http://ethics.cdto.center/ (дата обращения: 03.04.2020).

Котлярова, 2019 — *Котлярова В.В., Шемякина М.А.* Права искусственного интеллекта // Дневник науки. — 2019. — № 5 (29). — С. 71.

Кочетов, 1995 — *Кочетов А.Н.* Рыцари Х-лучей. — М., 1995.

Крик, 2002 — *Крик Ф.* Жизнь как она есть, ее зарождение и сущность / Пер. с англ. Е.В. Богатырева. — М.: Институт компьютерных исследований, 2002.

Мальтус, 1895 — *Мальтус Р.* Опыт о законе народонаселения / Пер. И.А. Вернер. — М.: К.Т. Солдатенков, 1895.

Мельников, 2018 — *Мельников С.В.* Прогностическое моделирование онтологий искусственного интеллекта как основа для проектирования необходимых референтных изменений законодательства // Право и государство: теория и практика. — 2018. — № 8 (164). — С. 92–95.

Лекторский, 2015 — *Лекторский В.А.* Возможны ли науки о человеке? // Вопросы философии. — 2015. — № 5. — С. 6.

Леонгард, 2018 — *Леонгард Г.* Технологии против человека / пер. М.В. Федоров. — М.: АСТ, 2018.

[Леонтьев А. Н.] Чензыя китайского философа совет, данный государю // Трутень. — 1770. — 23 февраля. — Лист VIII.

Незнамов, 2019 — *Незнамов А.В.* Рецензия на книгу Г. Леонгарда «Технологии против человека» // Дайджест Робоправо. — 2019. — Март. — Вып. 13. — С. 89.

Нейросети: как искусственный интеллект помогает в бизнесе и жизни (пост в блоге финтех-компании DTI Algorithmic от 6 июля 2017 г.). — URL: https://blog.dti.team/nejroseti/ (дата обращения: 11.07.2021).

Опарин, 1960 — *Опарин А.И.* Жизнь, ее природа, происхождение и развитие. — М.: Наука, 1960 (2-е изд. — 1968).

Панов, 2006 — *Панов О.В.* Функциональная структура бессознательного и возможность формирования новых принципов искусственного интеллекта // Искусственный интеллект: междисциплинарный подход / Под ред. Д.И. Дубровского и В.А. Лекторского. — М.: ИИнтеЛЛ, 2006. — 448 с.

Перепёлкина О. Нейронная соната: как искусственный интеллект генерирует музыку. — URL: https://tinyurl.com/2c57a7xx (дата обращения: 11.04.2021).

Петруня, 2017 — *Петруня О.Э.* Искусственный интеллект сквозь призму димензиональной онтологии // Философское образование. — 2017. — № 2 (36). — С. 13–14.

Садовский В.Н., Бабайцев А.Ю., Дроздов Н.Д. и др. Система // Гуманитарный портал. Концепты [Электронный ресурс]. — М.: Центр гуманитарных технологий, 2002–2021 (последняя редакция: 22.03.2021). — URL: https://tinyurl.com/2rpajnrs

Сбербанк, 2019 — Аналитический обзор мирового рынка робототехники 2019. [Электронный ресурс]. — URL: https://tinyurl.com/4sdkk6ef

Соменков, 2019 — *Соменков С.А.* Искусственный интеллект: от объекта к субъекту? // Вестник Университета им. О.Е. Кутафина (МГЮА). — 2019. — № 2. — С. 75–85.

Стерледева, 2017 — *Стерледева Т.Д.* Онтология электронно-виртуальной реальности // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. — 2017. — № 1 (75). — С. 190.

Тихомиров, 2020 — *Тихомиров Ю.А., Нанба С.Б.* Роботизация: динамика правового регулирования // Вестник Санкт-Петербургского университета. — Право. — 2020. — Т. 11. — Вып. 3. — С. 532–549.

Ужов, 2017 — *Ужов В.Ф.* Искусственный интеллект как субъект права // Пробелы российского законодательства. — 2017. — № 3. — С. 357–360.

Фогельсон, 2013 — *Фогельсон Ю.Б.* Мягкое право в современном правовом дискурсе // Журнал российского права. — 2013. — № 9. — С. 43–51.

Хабриева, 2018 — *Хабриева Т.Я., Черногор Н.Н.* Право в эпоху цифровой реальности // Журнал российского права. — 2018. — № 1. — С. 85–102.

Ястребов, 2018 — *Ястребов О.А.* Правосубъектность электронного лица: теоретико-методологические подходы // Труды Института государства и права РАН. — 2018. — Т. 13. — № 2. — С. 36–55.

Artificial Intelligence: From Ethics to Policy. Scientific Research Unit STOA. — 2020. — June. — URL: https://tinyurl.com/xf23upud

Automated Vehicle Trial Guidelines // National Transport Commission. [Электронный ресурс]. — URL: https://tinyurl.com/safv7bv9 (дата обращения: 01.04.2020).

Baichure, 2019 — *Baichure D.* Facial recognition // Assemblee Nationale: Science and Technology Briefings. — 2019. — № 14. — June. — URL: https://tinyurl.com/y7fzw8v2

Belton, 2018 — *Belton K.* How Should AI Be Regulated? Manufacturers Need to Pay Very Close Attention / Industry Week. — 2018. — June. — URL: https://tinyurl.com/2jbmyzbm (дата обращения: 27.03.2021).

Breland, 2017 — *Breland A.* Elon Musk: We Need to Regulate AI before «It's Too Late» // The Hill. — 2017. — 17 July. — URL: https://tinyurl.com/j2vw973h (дата обращения: 25.03.2021).

Calo, 2018 — *Calo R.* Artificial Intelligence Policy: A Primer and Roadmap // SSRN's eLibrary. — 2018. — August 8. — URL: https://tinyurl.com/3mhv2p29 (дата обращения: 25.03.2021).

Castro, 2016 — *Castro D., New J.* The Promise of Artificial Intelligence / Center for Data Innovation. — 2016. — P. 3. [Электронный ресурс]. — URL: https://tinyurl.com/42s9s3vd

Erdélyi, 2018 — *Erdélyi O.J., Goldsmith J.* Regulating Artificial Intelligence: Proposal for a Global Solution (February 2, 2018). 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18), February 2–3, 2018, New Orleans, LA, USA. — DOI/10.1145/3278721.3278731. — Available at SSRN: https://ssrn.com/abstract=3263992

H.R.4625 — Future of Artificial Intelligence Act of 2017. [Электронный ресурс]. — URL: https://tinyurl.com/27m7263b

Marchant, 2019 — *Marchant G.* Soft Law Governance of Artificial Intelligence // AI Pulse. — 2019. — 25 January. — URL: https://tinyurl.com/8hyxf2fw