

## Пределы принятия юридически значимых решений с использованием искусственного интеллекта\*

В. В. Архипов<sup>1</sup>, В. Б. Наумов<sup>2</sup>, К. М. Смирнова<sup>3</sup>

<sup>1</sup> Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

<sup>2</sup> Институт государства и права Российской академии наук, Российская Федерация, 119019, Москва, ул. Знаменка, 10

<sup>3</sup> ООО «Дентонс Юроп», Российская Федерация, 191011, Санкт-Петербург, Невский пр., 32–34

**Для цитирования:** Архипов, Владислав В., Виктор Б. Наумов, Ксения М. Смирнова. 2021. «Пределы принятия юридически значимых решений с использованием искусственного интеллекта». *Вестник Санкт-Петербургского университета. Право* 4: 882–906. <https://doi.org/10.21638/spbu14.2021.405>

Искусственный интеллект — одна из «сквозных цифровых технологий», предопределяющих современный ландшафт развития общества. Ключевая проблема ее применения с точки зрения права — пределы принятия юридически значимых решений с использованием искусственного интеллекта. Хотя на поверхности данная проблематика может показаться достаточно новой, в действительности она рассматривается в юридической науке и междисциплинарных исследованиях уже не один десяток лет. Указанные пределы могут быть разделены на парадигмальные, аксиологические и прагматические. Парадигмальные пределы связаны с фундаментальными представлениями о значении субъектности права и предопределяют принцип человеческого участия в решениях, принимаемых с помощью искусственного интеллекта. Аксиологические пределы предполагают последовательную реализацию определенных формализованных ценностей применительно к новым общественным отношениям. Наиболее разработанная система аксиологических пределов вытекает на данный момент из европейской концепции прав человека. Прагматические пределы обусловлены особенностями как самого права, так и технологической архитектуры искусственного интеллекта в социальном контексте. К числу прагматических пределов в данном смысле относятся проблемы «открытой текстуры» юридического языка и «смысловых пределов» права. Также в качестве прагматического предела может быть рассмотрен принцип недискриминации субъектов права при машинном обучении. Анализ современного регуляторного ландшафта на уровне практик регулирования и саморегулирования, сосредоточенный на репрезентативном международном опыте, позволяет верифицировать гипотезу о парадигмальных, аксиологических и прагматических пределах и предложить использовать их при дальнейшем развитии правового регулирования отношений в области принятия решений с использованием систем искусственного интеллекта.

**Ключевые слова:** искусственный интеллект, право, алгоритмическое принятие решений, юридически значимые решения, пределы права, цифровая экономика.

---

\* Статья подготовлена в рамках тематического плана исследований Института права цифровой среды, Национальный исследовательский университет «Высшая школа экономики».

© Санкт-Петербургский государственный университет, 2021

## 1. Введение

Правовая проблематика применения технологий искусственного интеллекта (далее — ИИ) представляет собой одну из наиболее актуальных областей. Во многом это связано с тем, что в мире за последние годы принято множество инициатив в данной области, и Россия не является исключением. Действует Национальная программа «Цифровая экономика Российской Федерации»<sup>1</sup>, в рамках которой технологии ИИ признаются одними из «сквозных цифровых технологий». Принята Национальная стратегия развития искусственного интеллекта на период до 2030 г., утв. Указом Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации»<sup>2</sup>. Правительством РФ утверждена Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.<sup>3</sup> Реализуются новые инициативы, меняется законодательство.

На поверхности вопрос об использовании технологий ИИ может показаться новым. В то же время правовая проблематика использования технологий ИИ, особенно ее центральная часть — пределы принятия юридически значимых решений с использованием ИИ<sup>4</sup>, не нова для науки. Первые исследования в области права и ИИ относятся к 1970-м и даже к 1950-м годам. Различия между экспертными системами, развивавшимися тогда преимущественно в экспериментальном режиме, и современными нейронными сетями, основанными на технологиях машинного обучения, составляющими часть актуальной повестки дня, могут показаться принципиальными с технологической точки зрения, однако с точки зрения права они не во всем столь принципиальны. Главное — это сам принцип автоматизации принятия юридически значимого решения. Разумеется, на новых технологических этапах появляются и новые проблемы, например дискриминации при использовании машинного обучения.

Принятие юридически значимых решений с использованием систем ИИ теоретически и практически возможно в разных, не всегда связанных между собой областях деятельности и в контексте различных отраслей права — от частноправовых отношений до государственного управления и правоприменительной деятельности. Однако на наиболее высоком уровне теоретического обобщения между всеми возможными контекстами прослеживается общий признак: в результате примене-

<sup>1</sup> Утверждена протоколом заседания президиума Совета при Президенте РФ по стратегическому развитию и национальным проектам от 04.06.201 № 7. Здесь и далее все ссылки на российские и международные нормативно-правовые акты приводятся по СПС «КонсультантПлюс». Дата обращения 7 июля, 2021. <http://consultant.ru>.

<sup>2</sup> В п. «а» ч. 5 Национальной стратегии указано: «Искусственный интеллект — 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».

<sup>3</sup> Распоряжение Правительства РФ от 19.08.2020 № 2129-р.

<sup>4</sup> Иногда она называется проблемой делегирования принятия решений ИИ — данное описание может показаться спорным, однако его допустимо рассматривать как легитимированное в социально-гуманитарных науках через отдельные концепции философии техники, см., напр.: (Latour 1994).

ния такой технологии происходит изменение правового статуса какого-либо субъекта права. Это позволяет предложить гипотезу, согласно которой общие условия и ограничения (т.е. пределы) принятия юридически значимых решений в таких случаях окажутся инвариантными и будут определяться как фундаментальными, так и прикладными представлениями о праве. Именно эта гипотеза и определяет содержание настоящего исследования, намечающего связь между теоретическими представлениями о пределах принятия решений с использованием ИИ и актуальным практическим контекстом с особым акцентом на международный опыт, который нельзя игнорировать. При этом за рамками статьи остаются вопросы ответственности за принятие юридически значимых решений с использованием ИИ, которые могут составить предмет самостоятельного исследования.

## 2. Основное исследование

### 2.1. Предмет исследования: основные понятия и термины

Несмотря на то что словосочетание «искусственный интеллект» прочно вошло в современный язык и повсеместно используется уже и в юридической литературе, оно может применяться в разных, пусть и родственных, смыслах, а потому нуждается в конкретизации. Некоторые авторы относят появление данного словосочетания как такового к 1956 г. Правда, на тот момент оно использовалось в смысле обозначения дисциплины в сфере компьютерных наук (Hunter 2020, 1213). Впоследствии одно из основных его значений стало определяться не столько областью знания, сколько различными технологиями, подразумевающими автоматизированные действия компьютера. В самом широком смысле словосочетание «искусственный интеллект» используется в большей степени как призванный придать драматичность факту использования компьютерных технологий риторический прием, за которым скрывается обычный алгоритм или компьютерная программа<sup>5</sup>, а потому в строгом научном смысле при таком употреблении он избыточен. В то же время при определенном подходе применение рассматриваемой фразы может иметь смысл.

В качестве недостатка словосочетания «искусственный интеллект» Р. Сасскинд, один из ключевых исследователей проблематики права и ИИ, отмечал, что оно «используется слишком многими способами и часто применяется в качестве не более чем грубого маркетингового оружия или как часть привлекающего внимание заголовка или твита» (Susskind 2019, 264). Сасскинд предлагает рассматривать два возможных подхода к определению ИИ в более или менее строгом смысле слова: архитектурный и функциональный. Определение ИИ с точки зрения архитектурного подхода предполагает отсылку к используемой технологии<sup>6</sup>. С функциональной точки зрения ИИ определяется фактом использования программных систем для

---

<sup>5</sup> Например, в игровой индустрии уже десятилетия назад закрепилось использование словосочетания «искусственный интеллект» для обозначения компьютерного оппонента, независимо то того, насколько сложная технология в действительности применяется.

<sup>6</sup> Во время первой волны развития ИИ такие технологии были основаны на принципе построения эксплицитных программных алгоритмов, а во время второй — на нейронных сетях и машинном обучении.

выполнения тех задач, которые традиционно требовали применения человеческого мышления независимо от технологии (Susskind 2019, 264–265). Мы соглашаемся с ученым в том, что именно функциональный подход к определению ИИ наиболее релевантен юридическим исследованиям, однако подчеркнем, что он не отменяет необходимость учитывать архитектурные особенности отдельных технологий ИИ, поскольку они могут существенным образом влиять на правовую проблематику.

Одной из таких технологий являются *нейронные сети*, предполагающие *машинное обучение*. Следует согласиться с теми авторами, которые связывают всеобщий интерес к проблематике ИИ именно с данными технологиями (Hunter 2020, 1214), — это подтверждают и отдельные исследования в области правовой информатики, отмеченные в следующем разделе настоящей работы. Полагаем нецелесообразным для целей юридического исследования заострять внимание на архитектурном техническом определении нейронных сетей в деталях — для целей статьи важен функциональный подход, определяющий юридически значимые принципы данной технологии. Принципиально важно в ней следующее: в отличие от экспертных систем первой волны (согласно условной периодизации Р. Сасскинда), нейронные сети не имеют заранее определенного алгоритма в строгом смысле слова, а «обучаются» на основе больших данных, отобранных для обработки такой системой. При этом нейронные сети могут оперировать любыми данными, включая правовые, для того чтобы установить зависимости между ними (Hunter 2020, 1215).

Д. Хантер отмечает: «В своей основе, глубокое [машинное] обучение представляет собой статистический метод для классификации паттернов, основанный на больших количествах образцов данных, задействующий нейронные сети с множеством уровней» (Hunter 2020, 1215). Иными словами, нейронная сеть по умолчанию не только не имеет заранее определенного алгоритма, понимаемого буквально<sup>7</sup>, но и не действует таким образом, который мог бы быть соотнесен с логикой принятия какого-либо решения в строгом смысле слова. Нейронные сети на основе машинного обучения, на первый взгляд, даже отдаленно не имитируют процесс логической аргументации<sup>8</sup>. Таким образом, принцип, лежащий в основе функционирования нейронной сети, может быть описан как прогнозирование на основе *sui generis* статистического анализа. Это представляется важным для понимания того, как именно нейронная сеть может «принять решение» — по принципу предсказания о том, каким соответствующее решение должно быть, на основе уже имеющихся примеров. Разумеется, технически не исключено сочетание технологий первой и второй волн развития ИИ. Однако если представить гипотетическое технологическое решение, основанное исключительно на нейронной сети с машин-

<sup>7</sup> Такой алгоритм мог бы, например, воспроизводить логику юридической аргументации, как это рассматривалось при юридическом анализе особенностей возможного применения экспертных систем в период «первой волны» ИИ.

<sup>8</sup> Возможное теоретическое возражение данному утверждению может следовать из той идеи, что нейронные сети изначально разрабатывались на основе теоретической модели функционирования человеческого мозга (что и предопределило их название). Сами люди, собственно говоря, также основывают свои решения в том числе на имеющихся данных, относящихся к прошлому (Susskind 2019, 290). Если предположить, что человеческое сознание первоначально выдвигает интуитивную гипотезу, основанную на опыте, и лишь потом легитимирует ее, притом что она может быть аргументирована *postfactum*, то такой процесс в самом общем виде будет мало отличаться от принципа функционирования нейронных сетей, которые, собственно, и моделируют функционирование человеческого мозга. Однако рассмотрение данной проблемы уже выходит за рамки статьи.

ным обучением, то предоставляемые им выводы могут быть уподоблены изучению иностранного языка не дедуктивным образом в школе, а путем подражания иностранцам и имитации их речи при личном погружении в языковую среду (Susskind 2019, 269–270).

В свою очередь, термин «юридически значимое решение» используется в настоящем исследовании в инструментальном смысле разновидности юридического факта, а именно юридического факта-действия. Юридический факт может быть определен как «жизненное обстоятельство, с которым норма права (общая или индивидуальная) связывает возникновение, изменение или прекращение правового отношения» (Поляков, Тимошина 2005, 399). Такая разновидность юридического факта, как действие, имеет волевую природу (Поляков, Тимошина 2005, 400). Однако для контекста исследования, связанного с ИИ, принципиально важен не только волевой, но и интеллектуальный аспект возникновения таких юридических фактов. Одно из значений слова «решение» — «обдуманное намерение сделать что-либо, заключение, вывод из чего-либо»<sup>9</sup>. Данный термин представляется удачным, поскольку для выявления условий и ограничений принятия решений, принимаемых с использованием систем ИИ, принципиален тот способ, с помощью которого решение будет достигнуто. Кроме того, с прагматической точки зрения применение данного термина в указанном смысле уже может считаться общепринятым, в том числе на уровне международных исследований данной проблематики (La Diega 2018 и др.), в контексте которых, как правило, используется устоявшийся термин «automatic decision-making». Поскольку с юридическими фактами связываются возникновение, изменение или прекращение правового отношения, а правовое отношение выражается в правах и обязанностях субъектов права, естественным признаком юридической значимости какого-либо решения выступает и изменение правового статуса субъекта права как сочетания субъективных прав и обязанностей.

## ***2.2. Исследования в области условий и ограничений принятия юридически значимых решений с использованием искусственного интеллекта***

### ***2.2.1. Общий контекст исследований в области искусственного интеллекта и права***

Как архитектурный, так и функциональный подход к пониманию ИИ в терминологии Р. Сасскинда позволяет отметить, что проблематика ИИ и права, несмотря на кажущуюся новизну, известна достаточно давно. Разумеется, актуальные проблемы принятия юридически значимых решений с использованием нейронных сетей и машинного обучения в контексте современных ценностных представлений о правах человека относятся к последним годам. Однако часть проблем, обусловленных самим фактором автоматизации (т. е. ИИ в широком смысле и с точки зрения функционального эффекта для юриспруденции), известна уже не менее нескольких десятков лет. Остановимся на наиболее заметных примерах и актуальных обобщающих исследованиях.

<sup>9</sup> «Решение». *Большой толковый словарь русского языка. Gramota.ru.* 2014. Дата обращения 19 октября, 2021. <http://gramota.ru/slovari/dic/?word=решение&all=x>.

До повсеместной практической актуализации технологий ИИ в последние годы предмет исследований, в центре которых находятся ИИ и право, мог рассматриваться скорее как узкоспециализированный и потому недостаточно известный широкому научному сообществу, однако первые работы, прямо связанные с данной проблематикой, некоторые зарубежные исследователи относят как минимум к 1971 г. (Hunter 2020, 1213), когда в юридической литературе были поставлены вопросы применения логики высказываний в юридической технике с использованием компьютерных технологий и проанализированы проблемы, связанные с экспериментом в области создания юридической экспертной системы Taxman, разработанной в Стэнфорде в 1971–1973 гг. (McCarty 1977).

В более широком смысле Д. Хантер связывает начало исследований в данной области с анализом возможностей применения формальных логических систем в области юридической техники, хотя и без применения экспертных систем (Hunter 2020, 1213; см. также: Allen 1957). Если смотреть еще шире, то появление данного направления может быть также соотнесено с формированием на рубеже 1950–60-х годов правовой информатики, предполагающей среди прочего анализ проблем применения кибернетических методов в юриспруденции, в том числе в трудах не только в американских, но и европейских ученых (Грачева 2019, 59).

Предложенная Е. В. Трофимовым и О. Г. Мецкером модель тематической периодизации публикаций в данной области (с 1970-х годов по настоящее время), связанных с направлением «право и методы искусственного интеллекта», включает следующие типы систем, которые в целом могут быть соотнесены с последовательными этапами технологического развития (приводятся в авторской терминологии):

- правовые экспертные системы на правилах и кейсах;
- гибридные правовые экспертные системы;
- абстрактные правовые аргументы, шаблоны «классной доски» и сервис-ориентированная архитектура;
- системы поддержки принятия правовых решений и мультизадачность;
- правовые информационные системы с интеграцией аргументационной функции;
- правовые информационные системы с интеграцией предсказательной функции;
- правовые информационные системы с интеграцией функции управления документами, контентом и знаниями;
- правовые информационные системы контрольного назначения.

Современный этап авторы соотносят с анализом методов машинного обучения и больших данных в предметной области права (Трофимов, Мецкер 2020, 155–160).

Отдельно отметим, что проблемы правовой информатики исследовались и в СССР — направление, мало освещенное в зарубежной литературе. Данные исследования относятся к началу 1960-х годов, продолжались в следующие десятилетия, а в 1990-е годы в значительной степени были сосредоточены на вопросах обработки правовой информации, в том числе в контексте развития справочных правовых систем (Грачева 2019, 60–63). Некоторые авторы проводят тематическую периодизацию отечественной правовой кибернетики, отмечая даже первые попытки разработки подходов к алгоритмизации юридического процесса и доказы-

вания, но далее определяя в качестве заметных направлений криминалистические автоматизированные информационные системы, юридический информационный язык и информационный поиск, информационно-поисковые системы (Трофимов, Мецкер 2020, 153–155).

Заметные и актуальные в контексте настоящего исследования работы, имеющие ярко выраженную источниковедческую направленность и содержащие значительное количество полезных библиографических ссылок, включают работу известного ученого в данной области Тревор Бенч-Капона «История ИИ и права в 50 работах: 25 лет международной конференции по ИИ и праву» (Bench-Caron et al. 2012), а также недавние статьи отечественных авторов (Грачева 2019; Трофимов, Мецкер 2020).

### 2.2.2. Классификация пределов принятия юридически значимых решений с использованием искусственного интеллекта

Как будет показано ниже, существует несколько наиболее типичных групп условий и ограничений, с которым в науке и практике связываются *возможность*, *ограниченная возможность* или *невозможность* принятия юридически значимых решений с использованием ИИ. В настоящей работе используется следующая инструментальная классификация пределов:

- *парадигмальные пределы* — пределы, определяемые господствующими фундаментальными представлениями и природе права, изменение которых невозможно без изменения общей парадигмы современной социальной реальности; с точки зрения области знаний данные пределы связаны с теоретико- и философско-правовыми исследованиями;
- *аксиологические пределы* — пределы, определяемые господствующими и формально закрепленными ценностями, используемыми в нормотворчестве и юридической аргументации; в данной работе в силу ограничений аксиологические пределы будут рассмотрены исключительно на примере европейской и российской моделей прав человека;
- *прагматические пределы* — пределы, не связанные с представлением о природе права или правовых ценностях как таковых, но обусловленные ограничениями или архитектурными особенностями либо правового материала как он может обрабатываться ИИ, либо самого ИИ.

Данная классификация пределов несколько условна, поскольку каждый из видов пределов может существенно пересекаться в зависимости от принятой модели объяснения. Рассмотрим указанные виды пределов последовательно, отмечая наиболее релевантные работы соответствующих авторов и актуальные на момент исследования нормативные и экспертные документы.

#### 2.2.2.1. Парадигмальные пределы

В терминологии настоящей статьи парадигмальные пределы определяются фундаментальными представлениями о праве. В актуальной форме в юридической литературе постановка вопроса в начальной форме обнаруживается еще в трудах

Р. Сасскинда 1980-х годов. Так, в 1986 г. он ставил вопрос о том, что экспертная система (вариант ИИ на тот момент — без нейронных сетей и машинного обучения), по-видимому, не может создаваться абстрактно, а должна исходить из определенной теоретической модели права, которая разрабатывается собственно в теории и философии права. Автор допускал пользу от анализа накопленного методологического потенциала теоретико-правовой науки, в частности выраженного в трудах таких теоретиков и философов права, как Г. Кельзен, Г. Л. А. Харт, Е. В. Булыгин, К. Альчуррон, Дж. Раз и др. Причина заключается в том, что «экспертные системы должны соответствовать определенной теоретико-правовой концепции потому, что все экспертные системы, используемые в области права, с необходимостью подразумевают предположения относительно природы права и юридического мышления. Если более конкретно, то все экспертные системы должны включать теорию структуры и разделения аконов, теорию правовых норм, теорию дескриптивной юридической науки, теорию юридического мышления, теорию логики и права и теорию правовых систем, а также элементы семантической теории, социологии и психологии права (теории, которые сами по себе должны базироваться на более основополагающих философских основаниях)» (Susskind 1986, 183).

Помимо догматических концепций, Р. Сасскинд уже тогда отмечал элементы семантической теории, социологии и психологии права как необходимые предпосылки для построения соответствующих экспертных систем. Появление технологий нейронных сетей, основанных на машинном обучении, предполагающем «тренировку» и формирование оснований для решений на основе статистических прогнозов, а не использования жестко запрограммированных алгоритмов, на наш взгляд, не влияет на актуальность данного подхода, поскольку с функциональной точки зрения нам важно не то, как именно технически указанные концепции заложены в результаты проектирования данных систем с технологической точки зрения, а то, каковы границы их возможного применения, исходя из фундаментальных теоретико-правовых и философских представлений. Ответ на вопрос о таких границах может изменяться в зависимости от типа правопонимания, однако здесь следует учитывать, что даже эксклюзивный юридический позитивизм в интерпретации, например, Е. В. Булыгина исключает ту же нравственную оценку законов<sup>10</sup> не в принципе из социальной реальности как таковой, а из строгого предмета юридической науки (Булыгин 2016, 13), который, по словам Г. Кельзена, должен представлять собой «чистое учение о праве». Даже если согласиться с такой позицией, анализ правовых проблем ИИ по своей логике должен быть в любом случае междисциплинарным, а значит, учет иных, более широких концепций природы права представляется необходимым в данном контексте.

Подробный анализ исторических и современных концепций, относящихся к различным типам правопонимания, в настоящей работе будет излишним, поэтому заострим внимание на следующем тезисе: из любой антропоцентричной концепции права с необходимостью следует принцип человеческого участия (*human in the loop*) в области применения ИИ для принятия юридически значимых решений, и это обусловлено тем, что в центре парадигмальных представлений о праве нахо-

---

<sup>10</sup> А также по аналогии оценку и иных областей, например области человеческой психики, внимание к которой характерно для психологического типа правопонимания и скандинавского правового реализма, если отделять его от данного общего направления.



дится субъект и, соответственно, понятие субъектности. В интегральном правопонимании «1) право конструирует, конституирует и регулирует отношения между людьми. Никакое явление, не имеющее отношение к человеку и его поведению, не может называться правом; 2) регулировать можно только то, что способно к такому регулированию, т. е. обладает способностью к личностному и деятельному существованию — *разумностью, свободой воли* (курсив наш. — В. А., В. Н., К. С.). Разумность репрезентирует себя через знаково-символические системы. Это означает, что право возможно только как коммуникативная система, определяющая отношения между субъектами. Право не существует между объектами — например, предметами материального мира, но право существует между субъектами по поводу объектов материального мира» (Поляков 2016, 304–305).

Схожий подход раскрывается в недавней работе Дж. Дэвиса «Искусственная мудрость? Потенциальные пределы ИИ в праве» (Davis 2019). Исходя из представленной им метафоры, ИИ оперирует в «пространстве научного мира», занимает позицию «внешнего наблюдателя» и, как следствие, может оказаться способным лишь распознавать и реализовывать средства, но не выбирать цели (Davis 2019, 53–54). По этим же причинам ИИ может оказаться неспособен распознавать социально значимый смысл<sup>11</sup> и выносить моральные (собственно, если смотреть шире, ценностные или смысловые) суждения (Davis 2019, 71). Автор соотносит данные ограничения ИИ с традицией, предполагающей, что выносящий юридическое суждение сам должен подчиняться праву, т. е. быть полноценным субъектом права (Davis 2019, 62–63). Эта идея согласуется и с современным теоретико-правовым дискурсом, в рамках которого необходимым качеством субъектности является возможность «признавать таким же субъектом любого другого» (Максимов 2016, 54). Соответственно, ИИ не имеет правосубъектности, а значит, юридически значимые решения невозможны без участия человека.

Контекст парадигмальных пределов такого рода допустимо расширить, если включить в предмет обсуждения и общеправовые сомнения в возможности принципиальной замены человека (например, выносящего юридически значимые решения) машиной. При таком подходе объем релевантных публикаций будет чрезвычайно широк. Остановимся лишь на одной репрезентативной работе. Еще в 1974 г. Х. Дрейфус высказал сомнения в том, что ИИ может заменить человека по причине двух методологических презумпций, характерных для области ИИ: эпистемологической, согласно которой «все разумное поведение может быть имитировано устройством, единственным способом обработки информации которого является обработка с точки зрения отвлеченного, лишённого тела и объективного наблюдателя», и онтологической, согласно которой, «основываясь на логическом атомизме, все необходимое для разумного поведения может быть в принципе понято как ограниченный набор независимых элементов» (Dreyfus 1974, 23). Компьютеры хорошо работают с абстракциями, но они лишены тела во всей его философской и органической полноте, не имеют нервной системы. Как следствие, могут наблюдаться существенные ограничения в тех случаях, когда требуется распознавание паттернов или решение проблем (Dreyfus 1974, 23). Несмотря на убедительность

---

<sup>11</sup> Здесь автор проводит аналогию с мысленным экспериментом Дж. Сёрла «Китайская комната» (Davis 2019, 71–72).

аргументов автора, характер его аргументации мог бы измениться, если бы его рассуждения были применены к современным нейронным сетям<sup>12</sup>.

Общее заключение данного параграфа: невозможность как правосубъектности ИИ, так и принятия юридически значимых решений с использованием лишь ИИ не соответствует современной правовой парадигме, но не только потому, что это может противоречить определенной ценностной модели (см. следующий параграф статьи), но и потому, что без субъектности нельзя представить современное понимание права. В то же время многие ключевые правовые концепции, например юридическая ответственность, с необходимостью предполагают такую субъектность, а потому данный предел может быть рассмотрен не только как ценностный, но и как парадигмальный. Подчеркнем: это не означает ни того, что ИИ в принципе нельзя использовать при принятии юридически значимых решений, ни того, что смена парадигмы, в которой субъектность играет ключевую роль, теоретически невозможна. Полагаем, примерно об этом пишет Р. Сасскинд, допуская изменение общественного восприятия гипотетического возникновения полностью «компьютерных» судей (Susskind 2019, 292). В контексте исследования данный парадигмальный предел тем не менее является интерпретацией антропоцентричной направленности большинства известных практике подходов к регулированию ИИ, раскрываемых ниже.

#### 2.2.2.2. Аксиологические пределы

Наиболее подробный и актуальный пример выявления аксиологических пределов принятия решений с использованием ИИ на данный момент отражен в комплексном исследовании *feasibility study*<sup>13</sup> Комитета Совета Европы *ad hoc* по искусственному интеллекту САНАИ (далее — Исследование САНАИ)<sup>14</sup>. Ценность этого документа определяется целым рядом обстоятельств: он отражает официальную, пусть и предварительную, и не нормативную позицию стран ЕС по правовым аспектам ИИ, особенно в контексте принятия юридически значимых решений; он содержит подробные отсылки к различным документам и исследованиям, которые легли в основу отраженных в нем заключений; наконец, он может быть вполне определенно интерпретирован как пример подхода, последовательно устанавливающего пределы и ограничения использования ИИ, исходя с необходимостью из применения в том числе дедуктивного метода, предполагающего в качестве одной из исходных посылок Европейскую конвенцию о защите прав человека

<sup>12</sup> Технологии развиваются. До того как компьютер Deep Blue компании IBM одержал верх над чемпионом мира по шахматам, исследователи ИИ сомневались в этом или вовсе полагали такое невозможным (Susskind 2019, 268; Dreyfus 1974, 22 и др.). Так, в 1974 г. Х. Дрейфус сопоставлял способности компьютеров в области шахмат с навыками «любителя класса С» и полагал, что для данной области характерны «постоянные провалы, следующие за оптимистичными прогнозами» (Dreyfus 1974, 22). Более показателен контраст между изначальным скепсисом относительно возможности искусственного интеллекта распознавать паттерны (Dreyfus 1974, 24–27) и развитием способностей нейронных сетей в данной области за последнее время (Yun, Hуен, Lu 2018).

<sup>13</sup> Одним из контекстуальных аналогов термина «*feasibility study*» в русском языке является «технично-экономическое обоснование», однако данный перевод представлялся бы неточным для рассматриваемого документа.

<sup>14</sup> “*Feasibility Study. Ad Hoc Committee on Artificial Intelligence (CAHAI)*”. *Council of Europe*. 2020. Дата обращения 7 июля, 2021. <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>.

и основных свобод от 04.11.1950 (далее — ЕКПЧ). В рамках настоящего исследования европейская модель прав человека рассматривается не как самостоятельное парадигмальное условие существования права самого по себе, а прежде всего как нормативно закрепленная система ценностей, в связи с чем вытекающие из Исследования САНАИ условия и ограничения применения ИИ могут рассматриваться именно в качестве аксиологических.

В разд. 7 Исследования САНАИ предлагается ряд предложений, которые достаточно новы, но не предполагают радикального изменения основ правового регулирования; напротив, они основаны на расширении европейской концепции прав человека к отношениям в области ИИ без изменения фундаментальной природы и понимания данных прав. Большинство новых предложений обусловлены юридически значимыми технологическими особенностями ИИ и связанными с ними новыми правовыми рисками. Так, одно из предложений заключается в закреплении права физических лиц получать информацию о взаимодействии с ИИ, иметь альтернативу в виде взаимодействия не с ИИ, а с другим человеком, права на пересмотр решения, принятого посредством ИИ. С одной стороны, право на информирование о применении ИИ и на отказ от взаимодействия с ИИ — это *prima facie* новые права. С другой стороны, в Исследовании САНАИ они напрямую обосновываются правом на человеческое достоинство и правом на физическую и психическую неприкосновенность в смысле ЕКПЧ.

Иные предложения, связанные с принятием юридически значимых решений, также обосновываются особенностями применения известных концепций прав человека в условиях функционирования ИИ. Ряд «специальных» прав рассматривается в Исследовании САНАИ как новая интерпретация свободы человека и автономии личности применительно к областям использования ИИ. К числу таких прав относятся, в частности, право на эффективное оспаривание решений, принятых на основе ИИ или принятых самой системой ИИ, и возможность требовать пересмотра такого решения человеком (по модели *opt-out*), право свободно принимать решение об исключении себя из манипуляций при помощи ИИ, профилирования и прогнозирования при обработке как персональных данных, так и данных, не являющихся таковыми, право на выбор взаимодействия с человеком или с ИИ<sup>15</sup>. Среди прочего такое понимание расширяет смысл общего правила о невозможности принятия решений на основании исключительно автоматизированной обработки персональных данных, предусмотренного Общим регламентом о защите данных<sup>16</sup>.

Основное заключение данного параграфа: пределы принятия юридически значимых решений с использованием ИИ могут определяться на аксиологических основаниях, примером которых служит европейская модель прав человека. В данном случае важно то, что в реальной, а не гипотетической правовой системе<sup>17</sup> возмож-

<sup>15</sup> В данном случае не общее право, а право, связанное с принятием решений на основе автоматизированной обработки персональных данных.

<sup>16</sup> “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)”. *EUR-Lex*. 2016. Дата обращения 7 июля, 2021. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

<sup>17</sup> Предполагаются ее иерархичность и наличие базовых ценностей, закрепленных на уровне основополагающих актов.

ность использовать системы ИИ определенным образом будет определяться не только парадигмальными представлениями вроде представлений о правосубъектности и не только прагматическими ограничениями архитектуры ИИ либо контекста их функционирования, но и такого рода ценностями. Перспективным будет следующий анализ аксиологических пределов, характерных, например, для российской правовой системы (которая в части ценностей на данный момент в отдельных аспектах отличается от европейской) или для мусульманского права.

### 2.2.2.3. Прагматические пределы

При анализе различных пределов принятия юридически значимых решений с использованием ИИ нельзя обойти стороной и пределы, которые можно назвать прагматическими, вытекающими из ограничений технологии, существующих на данный момент, или из контекста их применения, но не связанных напрямую ни с фундаментальной природой права, ни с базовыми ценностями определенной правовой системы.

Объем прагматических пределов будет ситуативно определяться уровнем развития технологии и особенностями ее архитектуры. Так, рассматривая экспертные системы, еще в 1987 г. Р. Сасскинд выделял пять практических ограничений, которые необходимо учитывать, исходя из уровня развития технологий: 1) область (отрасль) права для применения экспертной системы должна быть относительно автономна, а число источников — ограничено по количеству и разумно определено; 2) проблемы в такой области должны требовать экспертных знаний, а не простого исследования; 3) предпочтительным является глубокое погружение в узкую область, чем поверхностное — в широкую; 4) в избираемой области должно быть по возможности наибольшее согласие среди экспертов; 5) экспертные знания в избираемой области не должны быть слишком сильно связаны с аргументами из здравого смысла (Susskind 1987, 227–229). В ином случае, по мнению автора, ожидать эффективности от экспертной системы сложно. Данный подход в полной мере распространялся именно на системы, программируемые в строгом смысле этого слова, но, видимо, должен быть пересмотрен применительно к системам, основанным на нейронных сетях и машинном обучении. В то же время есть как общие пределы и для программируемых экспертных систем, и для ИИ, основанного на нейронных сетях, так и пределы, характерные именно для тех технологий ИИ, в которых задействовано машинное обучение.

К числу общих прагматических пределов, свойственных всем возможным системам ИИ, прямо или косвенно подразумевающим анализ правовых текстов, относится известная в теории и философии права проблема «открытой текстуры» языка, выявленная Г. Л. А. Хартом (Оглезнев 2016). Данная проблема, согласно одной из интерпретаций предполагающая принципиальную неопределенность пограничных (или относящихся к «полутени») значений слов, используемых в правовых текстах, которая упоминается и сейчас (Hunter 2020, 1214), отмечалась и на заре экспериментального развития юридических экспертных систем, и в период исследования потенциала нейронных сетей в области юриспруденции (Bench-Capon, Visser 2002). С этой проблемой также может быть связана и проблема так называемых семантических, или смысловых, пределов права как концепции, в основе кото-

рой лежит тезис о необходимости различать симулякры современного медиапространства и реальные объекты, что требует, в свою очередь, способности субъекта к восприятию социально значимых смыслов (Архипов 2019; Архипов 2020). Последняя проблема сущностно пересекается с ограничением, связанным со здравым смыслом, отмеченным ранее у Р. Сасскинда. Обе проблемы пересекаются и подразумевают не проясненный до конца вопрос о том, как должны быть спроектированы системы ИИ, чтобы разрешить или обойти указанные проблемы применительно к своему функционированию.

Наконец, термин «прагматические пределы» применительно к рассматриваемой области представляется удачным для актуального контекста ограничений, следующих из особенностей технологии, и применения современных архитектурных решений ИИ, основанных на нейронных сетях и машинном обучении, предполагающем использование больших, в том числе персональных, данных (La Diega 2018 и др.). Значительная часть подобных исследований ориентирована на исследование автоматического принятия решений (automatic decision-making) и проблемы дискриминации. Один из наиболее строгих подходов к определению такого рода пределов, собственно, и представлен в работе Гвидо Ното Ла Диега «Против дегуманизации принятия решения. Алгоритмические решения на перекрестке интеллектуальной собственности, защиты данных и свободы информации» (La Diega 2018). Автор, ориентируясь уже на современный этап развития ИИ, выдвигает ряд аргументов против замены принимающих решение людей алгоритмами (в широком смысле). Среди прочего им отмечены: 1) сложность и неопределенность процесса юридического толкования с теоретической точки зрения, не позволяющие ИИ толковать право так же, как это сделал бы человек<sup>18</sup>, а также практическая ориентированность современных практик толкования на телеологический подход; 2) заметная пропасть между человеческим обучением и машинным обучением, не позволяющим ИИ решать когнитивные задачи, требующие холистического мышления и интуиции (La Diega 2018, 5–18). Далее автор подробно раскрывает иные аргументы прагматического характера, включая отсутствие оснований «доверять» алгоритмам с точки зрения недостатков машинного обучения и возможной дискриминации, недостаточной прозрачности алгоритмов, имеющей многочисленные импликация, и сохранение оснований «доверять» людям (по разным фундаментальным причинам)<sup>19</sup>. Отдельное внимание Ла Диега уделяет тому, что ограничения права интеллектуальной собственности и защиты персональных данных по европейской модели существенным образом влияют как на прозрачность, так и на эффективность принятия решений с помощью ИИ (La Diega 2018, 11–21). Вместе с тем некоторые авторы допускают, что прозрачность решений, принимаемых с помощью ИИ, при правильном проектировании может быть выше, чем прозрачность неавтоматизированных решений за счет возможности обеспечить более последовательное выявление случаев дискриминации (Kleinberg et al. 2018).

---

<sup>18</sup> В данном случае автор упоминает концепции Г. Л. А. Харта и Р. Дворкина.

<sup>19</sup> В частности, имеются в виду: стремление людей подражать большинству других людей, что обеспечивает стабильность и предсказуемость социальных практик; психологический страх наступления санкций; стремление к рационализации поступков (пусть и лицемерное); эмпатия (La Diega 2018, 10–11).

### **2.3. Современное состояние прикладных подходов к разрешению проблемы принятия юридически значимых решений с использованием искусственного интеллекта**

Исследование мирового опыта регулирования общественных отношений, связанных с проблемой использования ИИ при принятии юридически значимых решений, позволяет сделать вывод о том, что данная область является крайне актуальной в мировом масштабе уже и с практической точки зрения. Хотя непосредственно императивное правовое регулирование в данной сфере пока еще слабо развито как в целом, так и применительно к отдельным отраслям, стремление различных стран к разработке соответствующего регулирования в форме нормативных правовых актов либо документов рекомендательного характера обнаруживается достаточно явно. При этом рекомендательные документы на сегодня составляют большинство правовых текстов по рассматриваемой проблеме.

Сложность формирования правового регулирования в данной сфере обусловлена неопределенностью относительно объема рисков, связанных с широким использованием ИИ в системах принятия решений. При этом предварительное, проактивное регулирование показывает свою неэффективность (Reed 2018), если такие риски еще не изучены в удовлетворительной степени. Намечается общая тенденция развития регулирования, нацеленного на снижение уже известных и очевидных рисков, в сочетании с постоянной аналитической работой по подготовке изменений в системе правового регулирования.

На наличие нормативных правовых документов в сфере делегирования решений, функций и обязанностей системам ИИ влияет сам подход к возможности и пределам использования автоматизированных решений. На уровне европейских государств прослеживается сдержанный подход, связанный с большим вниманием к рискам для граждан, угрозам для права на человеческое достоинство (Jones 2017, 230), который также обусловлен наличием развитого регулирования в сфере защиты прав субъектов персональных данных<sup>20</sup>. В США наблюдается обратный подход: действующее регулирование подходит к вопросам использования систем ИИ и иных автоматизированных систем принятия решений более открыто, хотя в литературе все же обсуждается необходимость сохранения окончательного решения за человеком в отдельных сферах (например, при использовании военных роботов (Citron, Pasquale 2014), применении ИИ при отправлении уголовного правосудия). В Канаде усматривается промежуточный подход: на момент проведения исследования нет полного отрицания возможности делегирования решений ИИ, но прослеживается более сдержанное отношение по сравнению с США.

Как правило, уполномоченные органы в той или иной сфере разрабатывают рекомендации, программные документы и документы стратегического характера. Нормативные правовые акты пока еще не так распространены, однако встречаются в ряде зарубежных государств (США, Канаде и др.). Принятое регулирование нацелено, во-первых, на легитимацию делегирования решений системам ИИ, а во-вторых, на минимизацию рисков причинения вреда пользователям таких систем.

---

<sup>20</sup> В соответствии с ним принятие автоматизированных решений в отношении субъекта персональных данных по общему правилу запрещено.

На федеральном уровне в США действуют правила, которые распространяются на автоматическое принятие решений в отдельных сферах, а также есть отдельные нормативные правовые акты на уровне конкретных штатов. Например, на федеральном уровне в сфере кредитования действует Закон о беспристрастной кредитной отчетности, устанавливающий требование об обязательном предоставлении потребителю пояснений в том случае, когда в его отношении было вынесено негативное решение на основании отчета о потребителе, полученного от бюро кредитных историй, при этом такой отчет может быть получен и с использованием систем ИИ<sup>21</sup>. Также на момент проведения исследования на федеральном уровне рассматривается Законопроект о прозрачности алгоритмов от 2019–2020 гг.<sup>22</sup> Законопроект предусматривает обязанность отдельных категорий предпринимателей и компаний, использующих системы автоматического принятия решений, проводить проверки таких систем на точность, справедливость, отсутствие предубеждения и дискриминации, ущемления прав на частную жизнь и безопасность, а также проверки уровня защищенности данных. Несмотря на наличие указанного законопроекта, регулирование данной темы на федеральном уровне на сегодня преимущественно отсутствует (Чае 2020). Что касается уровня штатов, то в качестве примера можно привести запрет использования систем распознавания лиц (в силу технологической природы такие системы предполагают использование ИИ) и «иных видов биометрического наблюдения» в Калифорнии<sup>23</sup>.

В Канаде разработана Монреальская декларация об ответственном развитии искусственного интеллекта, которая раскрывает в том числе вопросы использования ИИ при принятии решений<sup>24</sup>. Кроме того, в Канаде существует национальный стандарт создания и использования систем автоматического принятия решений с учетом подходов из области этики<sup>25</sup>. Еще один важный документ в сфере принятия решений ИИ — Директива по автоматизированному принятию решений<sup>26</sup>, которая касается использования ИИ для принятия решений в сфере государственного управления Канады. Директива направлена на снижение возможного негативного влияния решений, принятых ИИ, на граждан Канады и федеральные органы. Кроме того, она призвана обеспечить соответствие таких решений требованиям эффективности, точности, последовательности и легкости для интерпретации.

На уровне Европейского союза приняты различные программы, концепции и разъяснения, касающиеся правового регулирования сферы ИИ в целом, включая аспекты делегирования. В частности, в Открытой концепции по искусственному ин-

<sup>21</sup> “Fair Credit Reporting Act. § 615(a)”. *Federal Trade Commission*. 2018. Дата обращения 7 июля, 2021. [https://www.ftc.gov/system/files/documents/statutes/fair-credit-reporting-act/545a\\_fair-credit-reporting-act-0918.pdf](https://www.ftc.gov/system/files/documents/statutes/fair-credit-reporting-act/545a_fair-credit-reporting-act-0918.pdf).

<sup>22</sup> “Algorithmic Accountability Act”. *Congress.gov*. 2019–2020. Дата обращения 1 сентября, 2021. <https://www.congress.gov/bill/116th-congress/house-bill/2231/text>.

<sup>23</sup> “Assembly Bill No. 1215”. *California Legislative Information*. 2019. Дата обращения 7 июля, 2021. [http://leginfo.ca.gov/faces/billNavClient.xhtml?bill\\_id=201920200AB1215](http://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=201920200AB1215).

<sup>24</sup> “The Montréal Declaration for a responsible development of artificial intelligence”. *Montreal Declaration Responsible AI*. 2017. Дата обращения 7 июля, 2021. <https://www.montrealdeclaration-responsibleai.com/the-declaration>.

<sup>25</sup> “CAN/CIOSC 101:2019. Ethical design and use of automated decision systems”. *The Standards Council of Canada*. 2019. Дата обращения 7 июля, 2021. <https://www.scc.ca/en/standardsdb/standards/29998>.

<sup>26</sup> “Directive on Automated Decision-Making”. *Government of Canada*. 2021. Дата обращения 7 июля, 2021. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>.

теллекту, подготовленной Европейской комиссией, указывается на необходимость исследовать способность существующего правового регулирования противостоять рискам и угрозам, возможность адаптации существующих или необходимость разработки новых моделей и механизмов регулирования<sup>27</sup>. Среди ключевых аспектов, нуждающихся в исследовании, выделяются вопросы обработки персональных данных, защиты приватности, недискриминации, безопасности и ответственности. Отмечается, что разрабатываемое правовое регулирование ИИ должно быть направлено в том числе на повышение уровня прозрачности с целью соблюдения требований законодательства и предоставления более доступной возможности выявить и доказать нарушение правовых норм, определить ответственное лицо, а также на обеспечение информационной безопасности.

В других юрисдикциях регулирование, непосредственно касающееся принятия решений системами ИИ (или иными автоматизированными системами — в зависимости от того, как определять понятие ИИ), также развивается, в том числе в условиях, когда делегирование решений таким системам либо фактически осуществляется, либо его планируется осуществлять. Часто данное направление реализуется за счет инициатив по саморегулированию, предполагающих формирование отраслевых принципов. Примером могут послужить Япония<sup>28</sup> или Новая Зеландия<sup>29</sup>. В обоих случаях подразумевается акцент на человекоцентричном подходе. В то же время в июле 2020 г. в Новой Зеландии опубликована и правительственная Хартия алгоритмов<sup>30</sup>, которая включена в систему инструментов регулирования и саморегулирования и демонстрирует подход к тому, как правительственные учреждения должны использовать алгоритмы. Хартия вводит матрицу рисков, в зависимости от которых соблюдение провозглашенных принципов является возможным, желательным или обязательным. Термин «алгоритм» намеренно используется в Хартии в широком смысле слова, включая в себя и нейронные сети. Среди закрепленных принципов — документирование алгоритмов на ясном английском языке, доступность информации о данных и процессах, управление предвзятостью, обеспечение защиты неприкосновенности частной жизни, реализации этических принципов и прав человека, человеческий контроль за функционированием алгоритмов, в том числе в процессе принятия решений.

Анализ мирового опыта в части определения областей деятельности, в которых возможно и фактически осуществляется или планируется делегирование полномочий, функций, обязанностей системам ИИ, можно выделить следующие:

- государственные услуги и управление;
- правоохранительная деятельность: распознавание лиц системами видеонаблюдения (Ding 2018), автоматическое назначение штрафов за нарушение

<sup>27</sup> “White Paper on Artificial Intelligence. A European approach to excellence and trust”. *European Commission*. 2020. Дата обращения 7 июля, 2021. [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf).

<sup>28</sup> “Social Principles of Human-Centric AI”. *Japan Cabinet Office*. 2019. Дата обращения 7 июля, 2021. <https://www8.cao.go.jp/cstp/english/humancentricai.pdf>.

<sup>29</sup> “Trustworthy AI in Aotearoa The AI Principles”. *Artificial Intelligence Forum of New Zealand*. 2020. Дата обращения 7 июля, 2021. <https://aiforum.org.nz/wp-content/uploads/2020/03/Trustworthy-AI-in-Aotearoa-March-2020.pdf>.

<sup>30</sup> “Algorithm Charter for Aotearoa New Zealand”. *New Zealand Government*. 2020. Дата обращения 7 июля, 2021. [https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020\\_Final-English-1.pdf](https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf).



- правил дорожного движения<sup>31</sup>, определение случаев, в которых может потребоваться вмешательство полиции<sup>32</sup>, обработка данных с целью предупреждения, расследования, розыска или следствия в уголовных делах<sup>33</sup>;
- судебная деятельность: поддержка решений судьи, в том числе для обеспечения единообразия судебной практики<sup>34</sup>, скоринговые системы, оценивающие вероятность совершения рецидива (Stevenson, Doleac 2019)<sup>35</sup>, рассмотрение определенных категорий споров (Niiler 2019);
  - таможенная деятельность (Kavoua, 2020);
  - сфера здравоохранения (Greenfield 2019);
  - автоматизированный транспорт: в числе прочего беспилотный городской транспорт<sup>36</sup>, использование дронов для проверки промышленных трубопроводов<sup>37</sup>, автомобили;
  - сфера экологии (устойчивого развития)<sup>38</sup>;
  - банковская деятельность<sup>39</sup>: кредитный скоринг<sup>40</sup>, предложение персонализированных услуг<sup>41</sup>;

<sup>31</sup> Отдельный пример использования таких систем в КНР, демонстрирующий возможные неточности, см., напр.: (Shen 2018).

<sup>32</sup> См., напр.: “The Law Society. Algorithm use in the criminal justice system report System”. *The Law Society*. 2019. Дата обращения 7 июля, 2021. <https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report>.

<sup>33</sup> В рамках европейской модели данная область непосредственно соотносится с общими правилами относительно обработки персональных данных, включающими ограничения на принятие решений с использованием исключительно автоматизированной обработки персональных данных. Согласно ст. 11 (1) соответствующей директивы такая обработка в указанной области все же может быть разрешена Союзом или государством — членом ЕС, но минимальным обязательным условием является обеспечение реализации права на человеческое вмешательство со стороны контролера данных (“Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ 2016 L 119/89 (Law Enforcement Directive)”. *EUR-Lex*. 2016. Дата обращения 7 июля, 2021. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0680>).

<sup>34</sup> Перспективы и текущее состояние в этой области может продемонстрировать, в частности, следующий обзорный материал: (Guodong, Meng 2019).

<sup>35</sup> Один из наиболее известных примеров использования подобных технологий в ходе деятельности по отправлению правосудия представлен также в следующем деле, которое неоднократно анализировалось и в научных публикациях: *State v. Loomis*. 881 N. W.2d 749 (Wis. 2016) (“Wisconsin Supreme Court requires warning before use of algorithmic risk assessments in sentencing”. *Harvard Law Review*. 2017. Дата обращения 7 июля, 2021. <https://harvardlawreview.org/2017/03/state-v-loomis>).

<sup>36</sup> См., напр.: *Fabulous project*. 2020. Дата обращения 7 июля, 2021. <https://fabulos.eu>.

<sup>37</sup> См., напр.: *Aeroarms project*. 2020. Дата обращения 7 июля, 2021. <https://aeroarms-project.eu>.

<sup>38</sup> См., напр.: “Communication from the Commission to the European Parliament, the European Council, the Council, the European economic and social Committee and the Committee of the regions The European Green Deal”. *EUR-Lex*. 2019. Дата обращения 7 июля, 2021. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2019%3A640%3AFIN>.

<sup>39</sup> См., напр.: “Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore’s Financial Sector”. Дата обращения 7 июля 2021. <https://www.mas.gov.sg/~media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf>.

<sup>40</sup> См., напр.: “Financial Stability Board. Artificial intelligence and machine learning in financial services Market developments and financial stability implications”. *Financial Stability Board*. 2017. Дата обращения 7 июля, 2021. <https://www.fsb.org/wp-content/uploads/P011117.pdf>.

<sup>41</sup> См., напр.: “Algorithm Assessment Report”. *New Zealand Government*. 2018. Дата обращения 7 июля 2021. <https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf>.

- подбор персонала (Pasquale 2015);
- рынок недвижимости (скоринг благонадежности потенциальных арендаторов) (Selbst 2019).

Дополнительно стоит обратить внимание на использование технологий распознавания лиц, которая активно применяется правоохранительными органами разных стран. В США и отдельных штатах происходит отказ правоохранительных систем от использования систем ИИ для распознавания лиц. На федеральном уровне в конце июля 2020 г. в Конгресс США был внесен законопроект о моратории на использование технологий распознавания лиц и биометрических технологий<sup>42</sup>. На уровне штатов подобные акты приняты в Иллинойсе, Техасе, Калифорнии, Вашингтоне, Колорадо и Арканзасе<sup>43</sup>. Запрет направлен на усовершенствование технологий распознавания лиц, поскольку сейчас встречаются случаи ошибочного распознавания людей, что приводит к дискриминации.

Еще одной сферой, в которой делегирование решений и функций системам ИИ воспринимается чувствительным образом, является военная сфера. Обсуждается установление запрета на использование военных роботов, которые могут принимать полностью автоматизированные решения относительно жизни и смерти людей (см., напр.: Citron, Pasquale 2014).

Среди ключевых требований к системам ИИ, которым могут быть делегированы какие-либо полномочия, функции, обязанности и которые в контексте исследования соотносимы с выявленными парадигмальными, аксиологическими и прагматическими пределами, назовем следующие:

- отслеживаемость, объяснимость, прозрачность и проверяемость<sup>44</sup> применительно к отдельным решениям и процессу полностью, включая описания сбора, систематизации данных, логики алгоритмов<sup>45</sup>; процесс должен быть логирован и документирован;
- подотчетность и ответственность за действия таких алгоритмов<sup>46</sup>;
- эффективность решений, принимаемых алгоритмами (Ziewitz 2016);

<sup>42</sup> “The Facial Recognition and Biometric Technology Moratorium Act of 2020”. *Congress.gov*. 2020. Дата обращения 7 июля, 2021. <https://www.congress.gov/bill/116th-congress/senate-bill/4084>.

<sup>43</sup> См. пример обзора: “Facial recognition regulation in the USA: an efficient legal patchwork?”. *Mathias Avocats*. 2020. Дата обращения 7 июля, 2021. <https://www.avocats-mathias.com/donnees-personnelles/facial-recognition-usa>.

<sup>44</sup> См., напр.: “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions. Building Trust in Human-Centric Artificial Intelligence”. *European Commission*. 2019. Дата обращения 7 июля, 2021. <https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-168-F1-EN-MAIN-PART-1.PDF>; “H. R. 2231 — Algorithmic Accountability Act of 2019”. *Congress.gov*. 2019. Дата обращения 7 июля, 2021. <https://www.congress.gov/bill/116th-congress/house-bill/2231>; “Directive on Automated Decision-Making”. *Government of Canada*. 2021. Дата обращения 7 июля, 2021. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>.

<sup>45</sup> См., напр.: “Algorithm Charter for Aotearoa New Zealand”. *New Zealand Government*. 2020. Дата обращения 7 июля, 2021. [https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020\\_Final-English-1.pdf](https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf).

<sup>46</sup> См., напр.: “Draft AI R&D GUIDELINES for International Discussions”. *Ministry of Internal Affairs and Communications of Japan*. 2020. Дата обращения 7 июля, 2021. [https://www.soumu.go.jp/main\\_content/000507517.pdf](https://www.soumu.go.jp/main_content/000507517.pdf).

- отсутствие необъективности, дискриминации, неравенства, несправедливых решений, порождаемых алгоритмами, предвзятости алгоритмов (Eubanks 2018); данное требование может выполняться посредством использования наиболее полных массивов данных для обучения ИИ и проведения аудитов;
- соответствие систем ИИ не только нормам права, но и этическим принципам;
- возможность контроля со стороны человека (Grüll 2020);
- устойчивость и безопасность (Glaser, Funaiole 2019).

Чем меньше человеческого контроля предполагает конкретный случай делегирования полномочий системам ИИ, тем более жесткие требования должны предъявляться к тестированию и системному управлению технологиями ИИ. Проверка выполнения требований к системам ИИ может выполняться посредством проведения аудитов.

На уровне Европейского союза в наиболее актуальной и систематизированной форме содержание данных принципов, наряду с иными следствиями, вытекающими из европейской модели прав человека применительно к области ИИ, отражено в рассмотренном выше Исследовании САНАИ. Оно может рассматриваться как выражение накопленного методологического потенциала, позволяющего выявить пределы принятия решений с использованием систем ИИ

### 3. Выводы

Проблема принятия юридически значимых решений с использованием ИИ может показаться новой с точки зрения современного информационного пространства и актуального этапа внедрения информационных технологий. Однако такое представление обманчиво. В самом общем смысле исследования подобных пределов продолжают уже более 50 лет. Что же касается внедрения, то и здесь ситуация неоднозначна: да, на момент написания настоящей статьи делегирование правоприменительной (в широком смысле) функции ИИ еще не стало общим правилом, однако в ряде случаев<sup>47</sup> уже является реальностью. Кроме того, не следует игнорировать и широкое применение ИИ в частной сфере, особенно в рамках изолированных экосистем различных интернет-платформ, где ИИ де-факто выполняет правоприменительные функции в отношении пользователей.

Накопленный методологический потенциал исследования пределов принятия решений с использованием ИИ позволяет выделить три общих группы таких пределов: парадигмальные (связанные с природой права), аксиологические (связанные с формализованными ценностями) и прагматические (связанные с границами эффективности, определяемыми возможностями технологий ИИ или контекстом их применения). Следствием парадигмальных пределов, предполагающих субъектность ответственных за решения агентов, является концепция «human in the loop», предполагающая обязательное участие человека как минимум на каком-либо из этапов принятия решений с помощью ИИ. Аксиологические пределы в наиболь-

---

<sup>47</sup> Например, в части отдельных штрафов за нарушение правил дорожного движения либо примеров, связанных с использованием технологий распознавания лиц, основанных на ИИ.

шей степени проявляются в тех правовых системах, для которых характерна формализация ценностей, предполагающая их последовательное применение в рамках интерпретации иерархических правовых норм, примером чему может послужить Исследование САНАИ, отражающее следствия реализации европейской модели прав человека в области ИИ. Наконец, прагматические пределы связаны с особенностями юридического языка и толкования права («открытая текстура» и «смысловые пределы») и опасениями по поводу практических следствий технологических особенностей технологий ИИ, таких как риски дискриминации при машинном обучении нейронных сетей на основе больших данных.

Как показывает анализ современного регуляторного ландшафта, понимаемого в широком смысле и включающего в том числе опыт разработки принципов саморегулирования и примеры фактического использования технологий ИИ при принятии юридически значимых решений, представления о пределах такого рода характерны для современной общественной дискуссии по данной теме на мировом уровне. Представляется, что они должны как минимум учитываться при дальнейшем развитии подходов к правовому регулированию отношений, связанных с принятием юридически значимых решений посредством ИИ.

## Библиография

- Архипов, Владислав В. 2019. «Действие правовых норм в цифровом медиaprостранстве и семантические пределы права». *Правоведение* 63 (1): 8–27.
- Архипов, Владислав В. 2020. «Философско-правовые и этические основы разработки и применения систем ИИ и РТ: аспекты правосубъектности и социально значимых смыслов». *Правовые и этические аспекты, связанные с разработкой и применением систем искусственного интеллекта и робототехники: история, современное состояние и перспективы развития*, 10–71. СПб.: НП-Принт.
- Булыгин, Евгений В. 2016. «Мое видение рациональности права». *Булыгин, Евгений В. Избранные работы по теории и философии права*, 11–17. СПб.: Алеф-Пресс.
- Грачева, Анастасия В. 2019. «От правовой информатики до legal tech: история развития в России и за рубежом». *Закон* 5: 56–65.
- Максимов, Сергей И. 2016. «Концепция правовой реальности». *Постклассическая онтология права*, под ред. Ильи Л. Честнова, 23–59. СПб.: Алетейя.
- Оглезнев, Виталий В. 2016. «“Открытая текстура” юридического языка». *Вестник Томского государственного университета. Философия. Социология. Политология* 2 (34): 237–244.
- Поляков, Андрей В. 2016. «Права человека и суверенитет государства». *Постклассическая онтология права*, под ред. Ильи Л. Честнова, 295–324. СПб.: Алетейя.
- Поляков, Андрей В., Елена В. Тимошина. 2005. *Общая теория права*. СПб.: Издательский Дом Санкт-Петербургского государственного университета, Издательство юридического факультета Санкт-Петербургского государственного университета.
- Трофимов, Егор В., Олег Г. Мецкер. 2020. «Использование компьютерных методов и систем в изучении права, интеллектуальном анализе и моделировании правовой деятельности: систематический обзор». *Труды Института системного программирования Российской академии наук* 32 (3): 147–170.
- Allen, Layman E. 1957. “Symbolic logic: A razor-edged tool for drafting and interpreting legal documents”. *The Yale Law Journal* 66 (6): 833–879.
- Bench-Capon Trevor, Michał Araszkievicz, Kevin Ashley, Katie Atkinson, Floris Bex, Filipe Borges, Daniele Bourcier, Paul Bourguine, Jack G. Conrad, Enrico Francesconi, Thomas F. Gordon, Guido Governatori, Jochen L. Leidner, David D. Lewis, Ronald P. Loui, L. Thorne McCarty, Henry Prakken, Frank Schilder, Erich Schweighofer, Paul Thompson, Alex Tyrrell, Bart Verheij, Douglas N. Walton, Adam Z. Wyner.

2012. "A history of AI and Law in 50 papers: 25 years of the international conference on AI and law". *Artificial Intelligence and Law* 20 (3): 215–319. <https://doi.org/10.1007/s10506-012-9131-x>.
- Bench-Capon, Trevor J.M., Pepijn R. S. Visser. 2002. "Open texture and ontologies in legal information systems". *Database and expert systems applications. 8<sup>th</sup> International conference, DEXA '97. Proceedings*. Toulouse. <https://doi.org/10.1109/DEXA.1997.617268>.
- Chae, Yoon. 2020. "U. S. AI regulation guide: Legislative overview and practical considerations". *The Journal of Robotics, Artificial Intelligence & Law* 3 (1): 17–40.
- Citron, Danielle K., Frank A. Pasquale. 2014. "The scored society: Due process for automated predictions". *Washington Law Review* 89: 1–33.
- Davis, Joshua P. 2019. "Artificial wisdom? A potential limit on AI and law (and elsewhere)". *Oklahoma Law Review* 72 (1): 51–89.
- Ding, Jeffrey. 2018. *Deciphering China's AI dream: The context, components, capabilities, and consequences of China's strategy to lead the world in AI*. Oxford: Centre for the Governance of AI, Future of Humanity Institute, University of Oxford.
- Dreyfus, Hubert L. 1974. "Artificial intelligence". *The annals of the American Academy of political and social science* 412: 21–33.
- Eubanks, Virginia. 2018. *Automating inequality: How high-tech tools profile, police and punish the poor*. London: St. Martin's Press.
- Glaser, Bonnie S., Matthew P. Funaiolo, eds. 2019. *Perspectives on Taiwan: Insights from the 2018 Taiwan — U. S. Policy Program. Report of the CSIS China Power Project*. Lanham; Boulder; New York; London: Rowman and Littlefield.
- Greenfield, Daniel, Sean Wilson. 2019. "Artificial intelligence in medicine: Applications, implications and limitations". *Science in the News*. Дата обращения 7 июля, 2021. <https://sitn.hms.harvard.edu/flash/2019/artificial-intelligence-in-medicine-applications-implications-and-limitations>.
- Grüll, Philipp. 2020. "Deutschland fordert schärfere KI-Regulierung auf EU-Ebene". *Euractiv*. Дата обращения 7 июля, 2021. <https://www.euractiv.de/section/digitale-agenda/news/deutschland-fordert-schaerfere-ki-regulierung-auf-eu-ebene>.
- Guodong, Du, Yu Meng. 2019. "Big Data, AI and China's justice: Here's what's happening". *China Justice Observer*. Дата обращения 7 июля, 2021. <https://www.chinajusticeobserver.com/a/big-data-ai-and-chinas-justice-heres-whats-happening>.
- Hunter, Dan. 2020. "The death of the legal profession and the future of law". *University of New South Wales Law Journal* 43 (4): 1199–1225.
- Jones, Meg Leta. 2017. "The right to a human in the loop: Political constructions of computer automation and personhood". *Social Studies of Science* 47 (2): 216–239.
- Kavoya, Job. 2020. "Machine learning for intelligence-driven Customs management". *African Tax and Customs Review* 1 (3): 50–58.
- Kleinberg, Jon, Ludwig Jens, Sendhil Mullainathan, Cass R. Sunstein. 2018. "Discrimination in the age of algorithms". *Journal of Legal Analysis* 10: 113–174.
- La Diega, Guido Noto. 2018. "Against the dehumanisation of decision-making. Algorithmic decisions at the crossroads of intellectual property, data protection, and freedom of information". *Journal of Intellectual Property, Information Technology and E-Commerce Law* 9 (1). Дата обращения 23 октября 2021. <https://www.jipitec.eu/issues/jipitec-9-1-2018/4677>.
- Latour, Bruno. 1994. "On technical mediation — philosophy, sociology, genealogy". *Common Knowledge* 3 (2): 29–64.
- McCarty, Thorne L. 1977. "Reflections on 'taxman': An experiment in artificial intelligence and legal reasoning". *Harvard Law Review* 90 (5): 837–893.
- Niiler, Eric. 2019. "Can AI be a fair judge in court? Estonia thinks so". *Wired*. Дата обращения 7 июля, 2021. <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so>.
- Pasquale, Frank A. 2015. *The black box society: The secret algorithms that control money and information*. Harvard: Harvard University Press.
- Reed, Chris. 2018. "How should we regulate artificial intelligence?" *Philosophical transactions. Series A, Mathematical, physical and engineering sciences* 376 (2128): 1–12. <https://dx.doi.org/10.1098%2Ffrsta.2017.0360>.

- Selbst, Andrew D. 2019. "A new HUD rule would effectively encourage discrimination by algorithm". *SLATE*. Дата обращения 7 июля, 2021. <https://slate.com/technology/2019/08/hud-disparate-impact-discrimination-algorithm.html>.
- Shen, Xinmei. 2018. "Facial recognition camera catches top businesswoman 'jaywalking' because her face was on a bus". *South China Morning Post*. Дата обращения 7 июля, 2021. <https://www.scmp.com/abacus/culture/article/3028995/facial-recognition-camera-catches-top-businesswoman-jaywalking>.
- Stevenson, Megan T., Jennifer L. Doleac. 2019. "Algorithmic risk assessment in the hands of humans". *Social Science Research Network*. Дата обращения 1 сентября, 2021. <https://dx.doi.org/10.2139/ssrn.3489440>.
- Susskind, Richard E. 1986. "Expert systems in law: A jurisprudential approach to artificial intelligence and legal reasoning". *The Modern Law Review* 49: 168–194.
- Susskind, Richard E. 1987. "Some preliminary considerations concerning expert systems in law". *Northern Kentucky Law Review* 14 (2): 211–235.
- Susskind, Richard E. 2019. *Online courts and the future of justice*. Oxford: Oxford University Press.
- Yun, Kyongsik, Alexander Huyen, Thomas Lu. 2018. "Deep neural networks for pattern recognition". *arXiv*. Дата обращения 1 сентября, 2021. <https://arxiv.org/ftp/arxiv/papers/1809/1809.09645.pdf>.
- Ziewitz, Malte. 2016. "Special issue introduction: Governing algorithms: Myth, mess and methods". *Science, Technology & Human Values* 41 (1): 3–16.

Статья поступила в редакцию 3 апреля 2021 г.;  
рекомендована в печать 2 сентября 2021 г.

#### Контактная информация:

Архипов Владислав Владимирович — д-р юрид. наук; [v.arhipov@spbu.ru](mailto:v.arhipov@spbu.ru)  
 Наумов Виктор Борисович — д-р юрид. наук; [pau@russianlaw.net](mailto:pau@russianlaw.net)  
 Смирнова Ксения Михайловна — магистр юриспруденции; [smirnova.ks.law@gmail.com](mailto:smirnova.ks.law@gmail.com)

### The limits of automatic decision-making based on artificial intelligence in cases that have legal significance\*

V. V. Arkhipov<sup>1</sup>, V. B. Naumov<sup>2</sup>, K. M. Smirnova<sup>3</sup>

<sup>1</sup> St. Petersburg State University,  
7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

<sup>2</sup> Institute of State and Law of the Russian Academy of Sciences,  
10, Znamenka ul., Moscow, 119019, Russian Federation

<sup>3</sup> LLC "Dentons Europe",  
32–34, Nevsky pr., St. Petersburg, 191011, Russian Federation

**For citation:** Arkhipov, Vladislav V., Victor B. Naumov, Kseniia M. Smirnova. 2021. "The limits of automatic decision-making based on artificial intelligence in cases that have legal significance". *Vestnik of Saint Petersburg University. Law* 4: 882–906. <https://doi.org/10.21638/spbu14.2021.405> (In Russian)

Artificial intelligence technologies represent one of the "pervasive digital technologies" that predetermine the current landscape of society's technological development. The key problem of applying artificial intelligence technologies from the point of view of law is the limits of making legally significant decisions based on artificial intelligence. Although on the surface this issue may seem quite new, in fact it has been investigated in legal science and interdisciplinary research for decades. These limits can be divided into paradigmatic, axiological and pragmatic

\* This article was prepared as part of the thematic research plan of the Institute of the Law of Digital Environment, National Research University Higher School of Economics.

ones. Paradigmatic limits are associated with fundamental notions about the meaning of agency in law, and they predetermine the principle of human participation in decisions made with the help of artificial intelligence. Axiological limits presuppose a consistent realization of certain formalized values with respect to new social relations. The most developed system of axiological limits follows from the European concept of human rights. Pragmatic limits are conditioned by peculiarities of both law itself and the technological architecture of artificial intelligence in a social context. Pragmatic limits in this sense include the problem of the “open texture” of legal language and the “semantic limits” of law. The principle of non-discrimination of legal subjects in machine learning can also be considered as a pragmatic limit. The analysis of the modern regulatory landscape at the level of regulatory practices and self-regulation, focusing on representative international experience, makes it possible to verify the hypothesis of paradigmatic, axiological and pragmatic limits and to suggest their use in the further development of legal regulation of decision-making relations using artificial intelligence systems.

*Keywords:* artificial intelligence, law, automatic decision-making, legally significant decisions, limits of law, digital economy.

## References

- Allen, Layman E. 1957. “Symbolic logic: A razor-edged tool for drafting and interpreting legal documents”. *The Yale Law Journal* 66 (6): 833–879.
- Arkhipov, Vladislav V. 2019. “The effect of legal norms in the digital media space and the semantic limits of law”. *Pravovedenie* 63 (1): 8–27. (In Russian)
- Arkhipov, Vladislav V. 2020. “Philosophical legal and ethical foundations for the development and application of AI and robotic systems: aspects of legal personality and socially significant meanings”. *Pravovye i eticheskie aspekty, svyazannyye s razrabotkoi i primeneniem sistem iskusstvennogo intellekta i robototekhniki: istoriia, sovremennoe sostoianie i perspektivy razvitiia*, 10–71. St. Petersburg, NP-Print Publ. (In Russian)
- Bench-Capon, Trevor J.M., Pepijn R.S. Visser. 2002. “Open texture and ontologies in legal information systems”. *Database and Expert Systems Applications. 8<sup>th</sup> International Conference, DEXA '97. Proceedings*. Toulouse. <https://doi.org/10.1109/DEXA.1997.617268>.
- Bench-Capon, Trevor J.M., Michał Araszkievicz, Kevin Ashley, Katie Atkinson, Floris Bex, Filipe Borges, Daniele Bourcier, Paul Bourguine, Jack G. Conrad, Enrico Francesconi, Thomas F. Gordon, Guido Governatori, Jochen L. Leidner, David D. Lewis, Ronald P. Loui, L. Thorne McCarty, Henry Prakken, Frank Schilder, Erich Schweighofer, Paul Thompson, Alex Tyrrell, Bart Verheij, Douglas N. Walton, Adam Z. Wyner. 2012. “A history of AI and Law in 50 papers: 25 years of the international conference on AI and law”. *Artificial Intelligence and Law* 20 (3): 215–319. <https://doi.org/10.1007/s10506-012-9131-x>.
- Bulygin, Eugenio V. 2016. “My vision of the rationality of law”. *Bulygin, Eugenio V. Izbrannyye raboty po teorii i filosofii prava*, 11–17. St. Petersburg, Alef-Press Publ. (In Russian)
- Citron, Danielle K., Frank A. Pasquale. 2014. “The scored society: Due process for automated predictions”. *Washington Law Review* 89: 1–33.
- Chae, Yoon. 2020. “U.S. AI regulation guide: Legislative overview and practical considerations”. *The Journal of Robotics, Artificial Intelligence & Law* 3 (1): 17–40.
- Davis, Joshua P. 2019. “Artificial wisdom? A potential limit on AI and law (and elsewhere)”. *Oklahoma Law Review* 72 (1): 51–89.
- Ding, Jeffrey. 2018. *Deciphering China's AI dream: The context, components, capabilities, and consequences of China's strategy to lead the world in AI*. Oxford, Centre for the Governance of AI, Future of Humanity Institute, University of Oxford.
- Dreyfus, Hubert L. 1974. “Artificial intelligence”. *The annals of the American Academy of political and social science* 412: 21–33.
- Eubanks, Virginia. 2018. *Automating inequality: How high-tech tools profile, police and punish the poor*. London, St. Martin's Press.

- Glaser, Bonnie S., Matthew P. Funaiolo, eds. 2019. *Perspectives on Taiwan: Insights from the 2018 Taiwan — U.S. Policy Program. Report of the CSIS China Power Project*. Lanham, Boulder, New York, London, Rowman and Littlefield.
- Gracheva, Anastasiya V. 2019. "From legal informatics to legal tech: history of development in Russia and abroad". *Zakon* 5: 56–65. (In Russian)
- Greenfield, Daniel, Sean Wilson. 2019. "Artificial intelligence in medicine: Applications, implications and limitations". *Science in the News*. Accessed 7 July, 2021. <https://sitn.hms.harvard.edu/flash/2019/artificial-intelligence-in-medicine-applications-implications-and-limitations>.
- Grüll, Philipp. 2020. "Deutschland fordert schärfere KI-Regulierung auf EU-Ebene". *Euractiv*. Accessed 7 July, 2021. <https://www.euractiv.de/section/digitale-agenda/news/deutschland-fordert-schaerfere-ki-regulierung-auf-eu-ebene>.
- Guodong, Du, Yu Meng. 2019. "Big Data, AI and China's justice: Here's what's happening". *China Justice Observer*. Accessed 7 July, 2021. <https://www.chinajusticeobserver.com/a/big-data-ai-and-chinas-justice-heres-whats-happening>.
- Hunter, Dan. 2020. "The death of the legal profession and the future of law". *University of New South Wales Law Journal* 43 (4): 1199–1225.
- Jones, Meg Leta. 2017. "The right to a human in the loop: Political constructions of computer automation and personhood". *Social Studies of Science* 47 (2): 216–239.
- Kavoya, Job. 2020. "Machine learning for intelligence-driven Customs management". *African Tax and Customs Review* 1 (3): 50–58.
- Kleinberg, Jon, Ludwig Jens, Sendhil Mullainathan, Cass R. Sunstein. 2018. "Discrimination in the age of algorithms". *Journal of Legal Analysis* 10: 113–174.
- La Diega, Guido Noto. 2018. "Against the dehumanisation of decision-making. Algorithmic decisions at the crossroads of intellectual property, data protection, and freedom of information". *Journal of Intellectual Property, Information Technology and E-Commerce Law* 9 (1). Accessed 23 October 2021. <https://www.jipitec.eu/issues/jipitec-9-1-2018/4677>.
- Latour, Bruno. 1994. "On technical mediation — philosophy, sociology, genealogy". *Common Knowledge* 3 (2): 29–64.
- Maksimov, Sergey I. 2016. "The concept of legal reality". *Postklassicheskaja ontologija prava*, ed. by Ilya L. Chestnov, 23–59. St. Petersburg, Aleteia Publ. (In Russian)
- McCarty, Thorne L. 1977. "Reflections on 'taxman': An experiment in artificial intelligence and legal reasoning". *Harvard Law Review* 90 (5): 837–893.
- Niiler, Eric. 2019. "Can AI be a fair judge in court? Estonia thinks so". *Wired*. Accessed 7 July, 2021. <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so>.
- Ogleznev, Vitaly V. 2016. "Open texture" of legal language". *Vestnik Tomskogo gosudarstvennogo universiteta. Filosofija. Sotsiologija. Politologija* 2 (34): 237–244. (In Russian)
- Pasquale, Frank A. 2015. *The black box society: The secret algorithms that control money and information*. Harvard, Harvard University Press.
- Polyakov, Andrey V. 2016. "Human rights and state sovereignty". *Postklassicheskaja ontologija prava*, ed. by Ilya L. Chestnov, 295–324. St. Petersburg, Aleteia Publ. (In Russian)
- Polyakov, Andrey V., Elena V. Timoshina. 2005. *General theory of law*. St. Petersburg, Sankt-Peterburgskii gosudarstvennyi universitet Publ., Iuridicheskii fakul'tet Sankt-Peterburgskogo gosudarstvennogo universiteta Publ. (In Russian)
- Reed, Chris. 2018. "How should we regulate artificial intelligence?" *Philosophical transactions. Series A, Mathematical, physical and engineering sciences* 376 (2128): 1–12. <https://dx.doi.org/10.1098%2Frsta.2017.0360>.
- Selbst, Andrew D. 2019. "A new HUD rule would effectively encourage discrimination by algorithm". *SLATE*. Accessed 7 July, 2021. <https://slate.com/technology/2019/08/hud-disparate-impact-discrimination-algorithm.html>.
- Shen, Xinmei. 2018. "Facial recognition camera catches top businesswoman 'jaywalking' because her face was on a bus". *South China Morning Post*. Accessed 7 July, 2021. <https://www.scmp.com/abacus/culture/article/3028995/facial-recognition-camera-catches-top-businesswoman-jaywalking>.
- Stevenson, Megan T., Jennifer L. Doleac. 2019. "Algorithmic risk assessment in the hands of humans". *Social Science Research Network*. Accessed September 1, 2021. <https://dx.doi.org/10.2139/ssrn.3489440>.



- Susskind, Richard E. 1986. "Expert systems in law: A jurisprudential approach to artificial intelligence and legal reasoning". *The Modern Law Review* 49: 168–194.
- Susskind, Richard E. 1987. "Some preliminary considerations concerning expert systems in law". *Northern Kentucky Law Review* 14 (2): 211–235.
- Susskind, Richard E. 2019. *Online courts and the future of justice*. Oxford, Oxford University Press.
- Trofimov, Egor V., Oleg G. Metsker. 2020. "Application of computer techniques and systems in the study of law, intellectual analysis and modeling of legal activity: A systematic review". *Trudy instituta sistemnogo programmirovaniia Rossiiskoi akademii nauk* 32 (3): 147–170. (In Russian)
- Yun, Kyongsik, Alexander Huyen, Thomas Lu. 2018. "Deep neural networks for pattern recognition". *arXiv*. Accessed September 1, 2021. <https://arxiv.org/ftp/arxiv/papers/1809/1809.09645.pdf>.
- Ziewitz, Malte. 2016. "Special issue introduction: Governing algorithms: Myth, mess and methods". *Science, Technology, & Human Values* 41 (1): 3–16.

Received: April 3, 2021  
Accepted: September 2, 2021

#### Authors' information:

Vladislav V. Arkhipov — Dr. Sci. in Law; v.arkhipov@spbu.ru  
Victor B. Naumov — Dr. Sci. in Law; nau@russianlaw.net  
Kseniia M. Smirnova — M. Jur.; smirnova.ks.law@gmail.com