

Научная статья

УДК 34.096

EDN YOZDTD

DOI 10.17150/2411-6262.2023.14(4).1643-1656



Е.Н. Афанасьева¹ , Д.Д. Фурман² 

¹ Томский государственный университет систем управления и радиоэлектроники, г. Томск, Российская Федерация

² Новосибирский государственный университет, г. Новосибирск, Российская Федерация

Автор, ответственный за переписку: Е.Н. Афанасьева, afeka@inbox.ru

ПРАВОВЫЕ АСПЕКТЫ ОБРАБОТКИ ИСКУССТВЕННЫМИ ИНТЕЛЛЕКТУАЛЬНЫМИ СИСТЕМАМИ ОБЪЕКТОВ АВТОРСКОГО ПРАВА В ПРОЦЕССЕ МАШИННОГО ОБУЧЕНИЯ

АННОТАЦИЯ. В настоящей статье анализируются правовые аспекты обработки объектов авторского права искусственными интеллектуальными системами, в том числе, легальность обработки произведений в процессе автоматизированного анализа текста и данных в цифровой форме — TDM (text and data mining) и машинного обучения. Кроме того, в статье рассматриваются основания для квалификации такой обработки в качестве нарушения авторских прав. Выявляются препятствия для полноценной защиты авторских прав на произведения в составе Больших данных (Big Data), которые выступают предметом обработки искусственными интеллектуальными системами. Авторами предлагается разработка системы публично-правовых принципов обработки произведений искусственными интеллектуальными системами, в частности: принцип ограниченного целевого назначения, в соответствии с которым обработка произведений должна осуществляться исключительно в целях, установленных оператором искусственной интеллектуальной системы; принцип ограниченного хранения, предполагающий хранение персональных данных (в доступной для идентификации субъектов этих данных форме) не дольше, чем этого требует достижение поставленных целей обработки; принцип прозрачной отчетности — ведение отчетности относительно количества и качества обрабатываемых наборов данных, доступной для любого лица, вне зависимости от заинтересованности.

КЛЮЧЕВЫЕ СЛОВА. Машинное обучение, text and data mining, авторское право, большие данные, искусственный интеллект.

ИНФОРМАЦИЯ О СТАТЬЕ. Дата поступления 23 сентября 2023 г.; дата принятия к печати 07 декабря 2023 г.; дата онлайн-размещения 29 декабря 2023 г.

Original article

Е.Н. Afanasyeva¹ , D.D. Furman² 

¹ Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation

² Novosibirsk State University, Novosibirsk, Russian Federation

Corresponding author: E.N. Afanasyeva, afeka@inbox.ru

LEGAL ISSUES OF COPYRIGHT OBJECTS PROCESSING BY AI SYSTEMS IN THE PROCESS OF MACHINE LEARNING

ABSTRACT. The research paper analyses the legal aspects of copyright objects processing by AI systems, including the legality of automated analysis of text and data in digital form — TDM (text and data mining) and machine learning. The research paper examines: the grounds for qualifying such processing as copyright infringement and certain obstacles to the full protection of copyright (including Big Data, being a subject of processing by AI systems). The paper proposes the adoption of certain public law principles of copyright objects processing by artificial intelligent systems,

© Афанасьева Е.Н., Фурман Д.Д., 2023

in particular: the principle of limited purpose, according to which the processing of works should be carried out exclusively for the purposes established by the operator of the artificial intellectual system; the principle of limited storage, which involves storing personal data (in a form accessible to identify the subjects of this data) no longer than required in-order to achieve the stated processing purposes; the principle of transparent reporting — reporting on the quantity and quality of processed data sets, accessible to any person, regardless of interest.

KEYWORDS. Machine learning, text and data mining, copyright, big data, artificial intelligence.

ARTICLE INFO. Received September 23, 2023; accepted December 07, 2023; available online December 29, 2023.

Развитие технологий искусственных интеллектуальных систем (далее — ИИС) и введение в имущественный оборот результатов, созданных с использованием ИИС и идентичных результатам творческого труда (произведения), порождает в юридической литературе вопросы о правовом режиме такого рода результатов и правосубъектности ИИС особенно в контексте авторско-правовых отношений. В существенно меньшей мере уделяется внимание самому процессу машинного обучения, его правовой квалификации и законных пределов легальности осуществления. Проблема правового регулирования машинного обучения и сопутствующих (смежных) операций будет рассмотрены далее.

В рамках настоящего исследования термин «искусственная интеллектуальная система» будет использоваться в трактовке, содержащейся в Руководящих этических принципах надежного искусственного интеллекта: «Искусственные интеллектуальные системы — это программные (и, возможно, также аппаратные) системы, разработанные людьми, которые, учитывая сложную цель, действуют в физическом или цифровом измерении, воспринимая свою среду посредством сбора данных, интерпретируя собранные структурированные или неструктурированные данные, рассуждая о знаниях или обрабатывая информацию, полученную из этих данных, и принимая решение о наилучших действиях для достижения данной цели¹.

Однако важно учитывать, что ИИС, как готовый к применению инструмент, не возникает с момента написания программного кода, создания алгоритма, а требует прохождения предварительного этапа машинного обучения. Термин «машинное обучение» является неоднозначным и в юридической и технической литературе трактуется различно. Связано это с множеством способов машинного обучения, каждый из которых обладает спецификой. Технические специалисты вовсе используют термины «ИИС» и «машинное обучение» как синонимичные или тождественные. Как отмечает Семин П.О., термин «машинное обучение» связывают с набором особых алгоритмов, способных решать определенный класс задач. В процессе машинного обучения структурными элементами выступают наборы данных (Datasets), сам алгоритм обучения (программное обеспечение), а также ЭВМ (компьютер), а результатом процесса машинного обучения выступает «еще один алгоритм, который получает какие-то данные и выдает какой-то ответ, опираясь на то, чему он научился» [1, с. 25].

То есть результатом машинного обучения является ИИС, функционирование которой направлено на решение задач заранее определенной категории. В случае отношений, связанных с предметом исследования, задачей такого рода ИИС является создание результатов, идентичных результатам творческого труда (произведениям), по запросу Пользователей ИИС. На основании вышеизложенного представляется возможным определить машинное обучение как совокупность

¹ Ethics Guidelines for Trustworthy AI // Futurium. European Commission. 2019. URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651. (дата обращения: 30.0.2023).

алгоритмов, направленных на создание ИИС в форме иного алгоритма, выполняющего по запросу Пользователя ИИС задачи определенной категории на базе обработанных такой ИИС наборов данных.

Для наиболее корректного, высококачественного и быстрого осуществления процесса машинного обучения в качестве набора данных нередко используются Большие данные (Big Data) для их обработки ИИС. Единого подхода к определению термина «Большие данные» в юридической литературе на данный момент не сформировалось, однако принято выделять следующие признаки Больших данных, которые свидетельствует о том, что определенные массивы данных можно идентифицировать в качестве Больших: Объем, Скорость и Разнообразие [2, с. 70]. Иногда также выделяют различные вариации иных признаков: «внутренний ценный смысл» [3, с. 156] (полезность / применимость), достоверность [4, с. 72] (точность определенного набора данных) и др. Отдельные исследователи выделяют и больше признаков, однако вышеуказанные являются ключевыми. На основании перечисленных признаков в целях исключения неоднозначности термина в рамках настоящего исследования предлагается условно использовать термин «Большие данные» в следующей трактовке: Большие данные — это единый комплекс огромных массивов разнообразной динамически собираемой информации (данных), выраженной в доступной для обработки техническими инструментами форме. Такие данные в силу своей взаимосвязанности с ИИС могут быть и, вероятно, будут отсортированы, объединены и сгруппированы по определенному критерию или общему признаку. В исследуемой сфере данные сортируются по критерию отношения к объектам авторского права или определенному его виду (литературные произведения или иные текста, аудио- и видеозаписи и т.д.).

Из-за своих полезных свойств Большие данные становятся одним из ключевых ресурсов (если не самым ключевым), выступающим в качестве объекта машинного обучения ИИС, которая в процессе и результате машинного обучения становится мощнее в своем технологическом функционале. Есть прямая корреляция между эффективностью отдельно взятой ИИС и количеством обработанных ею данных. С недавних пор стало доступно колоссальное количество изображений и аудиоданных. С 2015 г. генерируется 90 % мировых данных. В том же году цифровая вселенная, т. е. резервуар данных, созданных и скопированных, насчитывал менее 10 зеттабайт [5, с. 20].

Как уже указывалось ранее, осуществление машинного обучения требует набора вводных данных, на основе которых ИИС в дальнейшем функционирует. Однако в случае использования Больших данных в машинном обучении требуется проведение отдельного этапа, именуемого Text and Data Mining (пер. англ. — «добыча / извлечение текста и данных»; далее — TDM). Согласно положениям п. 2 ст. 2 Директивы Европейского парламента и Совета ЕС от 17.04.2019 г. № 2019/790 «Об авторском праве и смежных правах на Едином цифровом рынке и о внесении изменений в Директивы № 96/9/ЕС и 2001/29/ЕС» (далее — Директива № 2019/790), под TDM понимается любой автоматизированный аналитический метод нацеленный на анализ текста и данных в цифровой форме для получения информации, включающей модели, тенденции и корреляции, но при этом не ограниченной только ими². По своему содержанию данный процесс служит инструментом формирования набора данных (Dataset) определенного вида. Поэтому в теории TDM необязательно используется в целях машинного обучения, целью

² Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC // Official Journal of the European Union. URL: <https://eur-lex.europa.eu/eli/dir/2019/790/oj>. (дата обращения: 30.09.2023).

его осуществления могут выступать аналитические репорты, формирование баз данных и т.п. Тем не менее, на практике тяжело представить его реализацию без использования искусственных интеллектуальных систем, тем более, когда речь идет об обработке Больших данных.

На основании вышеизложенного можно выявить, что в базис «обучения» и функционирования ИИС входит обработка данных (информации) определенной категории. В рамках предмета исследуемой проблемы наборы данных (Datasets), как правило, содержат объекты авторского права (произведения). Выражается указанная обработка в следующем:

- TDM, как метод анализа текстов и данных, направленный на формирование набора данных, извлекаемых из Больших данных, для машинного обучения;
- машинное обучение, как совокупность алгоритмов, направленных на создание иного алгоритма (ИИС) на основе набора данных;
- создание результатов, идентичных результатам творческого труда, по итогам машинного обучения.

Важно отметить, что вышеуказанные организационно-технические операции являются элементами функционирования каждой конкретной ИИС, поэтому предстают смежными в отношении друг к другу и частично совпадают. Их последовательное и совокупное выполнение предлагается именовать как Обработку данных искусственной интеллектуальной системой, которая по содержанию в большей степени совпадает с Обработкой Больших данных в случаях, если предметом Обработки являются Большие данные (их содержимое), различие фактически выражается в объеме данных и технической методологии их добычи. Под Обработкой данных ИИС предлагается понимать любые операции с данными (наборами данных) в любой объективно выраженной форме, совершаемые с использованием ИИС и направленные на извлечение полезных свойств из этих данных. Например, формирование какой-либо статистики и выводов по ней или подготовка и структуризация материалов для машинного обучения. Исходя из вышеуказанного, Обработкой произведений ИИС далее по тексту будет пониматься такая Обработка данных ИИС, предметом которой выступают произведения.

Как уже упоминалось, содержимым наборов данных и Больших данных в определенных случаях являются охраняемые авторским правом произведения. Статьей 1259 Гражданского кодекса Российской Федерации (далее — ГК РФ) установлено, что объектами авторских прав являются произведения науки, литературы и искусства независимо от способа его выражения. Исходя из содержания ст. 1229 ГК РФ, правообладатель имеет юридическую монополию на распоряжение объектом авторского права любым незапрещенным законом способом. Никто, кроме правообладателя, не вправе использовать объект авторского права, пока правообладатель не даст свое разрешение, оформленное соответствующим законом способом. Для правовой оценки легальности Обработки данных ИИС в авторско-правовом аспекте следует подробнее проанализировать каждый из трех вышеуказанных элементов Обработки.

Чтобы изучить, какие юридические проблемы могут возникнуть в процессе TDM, важно отметить, как он работает. TDM включает в себя последовательность определенных операций, которые необходимо выполнить для экономически эффективного извлечения большого объема текста и / или данных. В связи с этим существуют три общих (однако не обязательных) этапа, которые могут включать:

- доступ к исходным материалам, подлежащим анализу, таким как данные, собранные индивидуально или систематизированные в базе данных;
- копирование значительного объема материала и / или извлечение данных, которые также могут включать предварительную обработку материала путем пре-

образования его в машиночитаемый формат и загрузку предварительно обработанного контента на платформу;

– извлечение данных и их повторное объединение для получения новых знаний и закономерностей в конечном результате [6].

Как можно заметить, TDM включает в себя копирование произведений на некий материальный носитель, с которого будет производиться извлечение данных и их систематизаций. Данное действие представляется возможным квалифицировать как воспроизведение произведения в цифровой и / или машиночитаемой форме. Воспроизведение традиционно относят к использованию произведения, которое является прерогативой правообладателя, на основании п. 1 ст. 1270 ГК РФ и ст. 2 Директивы № 2001/29/ЕС Европейского парламента и Совета Европейского Союза «О гармонизации некоторых аспектов авторских и смежных прав в информационном обществе» (далее — Директива № 2001/29/ЕС).

Согласно п. 1 ст. 5 Директивы № 2001/29/ЕС на временные акты воспроизведения, которые являются краткосрочными или побочными, неотъемлемой и существенной частью технологического процесса и единственной целью которых является обеспечение возможности передачи по сети между третьими сторонами с помощью посредника, или законное использование произведения или другого подлежащего созданию объекта, которые не имеют самостоятельного экономического значения, ограничение права на воспроизведение не распространяется³.

Положения, аналогичные вышеуказанным, содержатся и в российском законодательстве. Согласно подп. 1 п. 2 ст. 1270 ГК РФ использованием в том числе считается воспроизведение произведения, вне зависимости от того, совершается ли оно в целях извлечения прибыли или нет. Под воспроизведением понимается изготовление одного и более экземпляра произведения или его части в любой материальной форме, в том числе в форме звуко- или видеозаписи, изготовление в трех измерениях одного и более экземпляра двухмерного произведения и в двух измерениях одного и более экземпляра трехмерного произведения. При этом запись произведения на электронном носителе, в том числе запись в память ЭВМ, также считается воспроизведением.

Однако важно отметить то, что данный подпункт закрепляет аналогичное вышеуказанному в Директиве № 2001/29/ЕС исключение из вышеуказанного правила: воспроизведением не считается краткосрочная запись произведения, которая носит временный или случайный характер и составляет неотъемлемую и существенную часть технологического процесса, имеющего единственной целью правомерное использование произведения либо осуществляемую информационным посредником между третьими лицами передачу произведения в информационно-телекоммуникационной сети, при условии, что такая запись не имеет самостоятельного экономического значения.

М.А. Кольздорф в своем исследовании свободного использования произведений при обработке Больших данных проанализировала выводы судебной практики Суда ЕС касательно применения положений Директивы № 2001/29/ЕС и пришла к следующим выводам. Действие по воспроизведению квалифицируется временным тогда, когда его продолжительность не выходит за рамки необходимой для нормального функционирования соответствующего технологического процесса, в котором должно быть предусмотрено автоматическое (без вмешательства человека) прекращение такого воспроизведения, как только его функция по обеспечению реализации релевантного процесса выполнена. Термин «побочный /

³ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society // Wipo. URL: <https://www.wipo.int/wipolex/en/legislation/details/1453> (дата обращения: 30.09.2023).

второстепенный» касается копий, которые имеют второстепенный характер по отношению к технологическому процессу, к основному действию по использованию произведения. Действия по воспроизведению должны быть необходимы в том смысле, что без них технологический процесс не будет протекать правильно и эффективно, а воспроизведение должно осуществляться полностью в рамках технологического процесса. Правомерным считается использование, не ограниченное законом или согласованное правообладателем прямо или косвенно. Суд ЕС интерпретирует этот термин с точки зрения той цели, с которой копия была сделана (целевое использование). В свою очередь, наличие экономического значения подтверждается в том случае, если копии создают отдельную дополнительную прибыль, выходящую за рамки полученной от правомерного использования произведения. К тому же, экономическое значение возникает в случае изменения произведения в процессе воспроизведения [7, с. 149–150].

На основе вышеизложенных норм возможно провести определить пределы исключительных прав на произведение, права на его воспроизведение в процессе TDM, а также в процессе машинного обучения и создания результатов, идентичных результатам творческого труда.

Как уже указывалось ранее, первым этапом TDM является доступ к исходному материалу (данные или базы данных) как предмету TDM. Данное действие является неотъемлемой и существенной частью технологического процесса TDM, без осуществления которых TDM в принципе не представляется возможным произвести.

Копирование и извлечение данных уже непосредственно являются актом воспроизведения, следовательно, необходимо проанализировать кратковременность и побочность (случайность) сессий TDM в целях разрешения вопроса легальности такого воспроизведения. Исключение в связи с кратковременностью и побочностью не могут однозначно распространяться на все случаи использования произведений посредством технологий TDM. Это связано с тем, что воспроизведение при анализе массивов данных на определенных этапах может не быть временным или побочным (случайным), несмотря на то что оно будет обусловлено исключительно технологическими особенностями. Поэтому затруднительно распространить действие данного исключения на все случаи использования TDM [8, с. 44]. Кроме того, как указывалось ранее, копирование не может быть квалифицировано как временное в случае необходимости вмешательства человека ради инициации процесса удаления копий данных (произведений).

Помимо копирования данных на этом этапе производится их преобразование в машиночитаемую форму. Однако, как справедливо указывает Д.А. Грачева, перевод текста в машиночитаемую форму, который производится перед TDM, не является переработкой произведения. Перевод в машиночитаемую форму заключается в оцифровке с возможностью выделять отдельные слова и буквы из текста без изменения содержания произведения [9, с. 44–52]. К тому же, переработка должна носить творческий характер, что исключает квалификацию TDM как переработку произведения.

Так как правомерность использования определяется целеполаганием воспроизведения, то можно предположить, что если цель TDM заключается в извлечении информации с целью получения новых закономерностей и знаний, то такое использование может быть правомерным, поскольку информация как таковая не охраняется авторским правом. Однако стоит учитывать, что каждый этап TDM может обладать самостоятельной экономической ценностью в зависимости от конкретных обстоятельств [7, с. 150]. То есть извлечение данных и их повторное объединение должно производиться в пределах заранее обозначенной правомерной цели. Стоит также учитывать, что в контексте осуществления TDM все об-

рабатываемые произведения принимают форму информации (данных), то есть не представляют ценности по критерию творческого вклада, что позволяет исключить возникновение самостоятельного экономического значения TDM. Статьей 4 Директивы № 2019/790 были расширены и конкретизированы некоторые положения относительно воспроизведения и извлечения объектов авторского права в течение TDM. Например, были установлены императивные требования о свободном воспроизведении произведений в процессе TDM в целях научных исследований.

Тем не менее, такое воспроизведение и извлечение возможно только в том случае, если к произведению или иному охраняемому объекту имеется законный доступ со стороны пользователя, в том числе, когда оно было доведено до всеобщего сведения онлайн, и, если правообладатели не сохранили за собой надлежащим образом права на воспроизведение и извлечение для целей интеллектуального анализа текстов и данных. Д.А. Мотовилова отмечает, что «на первый взгляд ст. 4 Директивы № 2019/790 разрешает использование произведений и иных охраняемых объектов для целей TDM, фактически правообладатели могут запретить подобное использование охраняемого контента» [10, с. 108].

Таким образом, правовое регулирование воспроизведения в ЕС позволяет при определенных условиях использовать произведения в рамках TDM, если правообладателем намеренно не было закреплено исключительное право на такого рода использование произведения. В свою очередь, законодательство РФ на данный момент не предусматривает специальных норм для TDM, поэтому каждое использование произведений в процессе TDM должно соответствовать положениям ст. 1270 ГК РФ. Можно сделать вывод, что TDM в некоторых случаях может быть квалифицирован в качестве нарушения авторских прав при нарушении условий правомерного воспроизведения, указанных в подп. 1 п. 2 ст. 1270 РФ.

В свою очередь, как уже указывалось ранее, машинное обучение представляет собой схожий, в некоторых элементах смежный с TDM процесс, в результате которого создается ИИС. Однако ситуация в таком случае несколько иная из-за цели машинного обучения. Результат (модель) машинного обучения подразумевает сохранение в памяти ИИС обработанных произведений в целях реализации функционала ИИС⁴. Из этого следует, что машинное обучение (отдельно от TDM) представляется возможным квалифицировать как воспроизведение в машиночитаемой форме. Можно с уверенностью сказать, что, как и в случае с TDM, машинное обучение воспроизводит произведения в рамках технического процесса и без этого процесс был бы неосуществим. Тем не менее, в отличие от TDM, машинное обучение включает долгосрочное хранение в памяти данных (наборов данных), которые служили материалом для обучения модели ИИС, что выходит за пределы правомерного воспроизведения произведения. Таким образом, машинное обучение нарушает авторские права правообладателей произведений, содержащихся в памяти модели ИИС.

С критерием экономической самостоятельности воспроизведения в рамках машинного обучения все не так просто. Выявить нарушение авторских прав по данному критерию можно в связи со следующим этапом Обработки произведений ИИС — создание результатов, идентичных результатам творческого труда. Дело в том, что произведения, содержащиеся в памяти ИИС, не имеют самостоятельного экономического значения по идентичным доводам, как и в случае с TDM. Однако, чтобы уследить здесь правонарушение, следует учитывать, для чего именно было осуществлено машинное обучение и будет ли коммерциализирован результат машинного обучения.

⁴ Как работает память искусственного интеллекта // AnyQuestion. URL: <https://ru.anyquestion.info/a/kak-rabotaet-pamyat-iskusstvennogo-intellekta>. (дата обращения: 30.09.2023).

В качестве результатов машинного обучения можно выделить как саму готовую обученную модель ИИС, так и результаты, созданные с ее помощью. «Владелец» ИИС теоретически может продать права на нее, предоставить в возмездное пользование, монетизировать результаты, создаваемые ею. Указанные возможности использования ИИС образуют самостоятельное экономическое значение для обработанных в рамках машинного обучения произведений и для созданных на их основе результатов, идентичных результатам творческого труда. Исходя из сущности машинного обучения, чем больше алгоритм машинного обучения обрабатывает наборы данных, тем точнее (ценнее) будет конечный результат и тем более увеличивается самостоятельная экономическая значимость этого результата и обработанных произведений, так как более качественный итоговый продукт расширит пределы потенциально получаемой «Владельцем» ИИС прибыли.

Несмотря на уже выявленные основания для квалификации машинного обучения как нарушения авторских прав, следует отметить иную проблему правомерности машинного обучения. Использование ИИС для генерации различных результатов, идентичных результатам творческого труда, например, изображений, позволяет создавать такие результаты с буквально нечеловеческой скоростью, что наполняет рынок колоссальным количеством произведений соответствующей категории. Очевидно, что авторы (правообладатели) не способны конкурировать с ИИС в плане скорости, не прибегая к использованию этих самых ИИС. Тем самым нарушаются имущественные интересы авторов произведений.

Усугубляется характер такого нарушения тем, что без предварительной Обработки (использования) произведений было бы крайне затруднительно осуществить машинное обучение, что закономерно вызывает недовольство у авторов и порождает прецеденты исков к платформам, позволяющим использовать ИИС. Например, в настоящий момент подан коллективный иск на Stable Diffusion за использование набора данных с множеством объектов авторского права, включенных в данный набор без согласия авторов⁵. К тому же, с помощью ИИС при соответствующем обучении есть возможность воссоздавать уникальные (или трудновоспроизводимые) элементы в результатах, идентичных результатам творческого труда, например, авторский стиль рисования или художественный стиль письма⁶. Само по себе это не является нарушением, так как на основании пункта 5 статьи 1259 авторские права не распространяются на методы, способы решения различного рода задач, к которым и можно отнести авторский стиль произведения. Однако ключевое в данном случае — обесценивание и снижение рентабельности произведений на соответствующем рынке вследствие потенциального количества произведенных в определенном авторском стиле результатов, идентичных результатам творческого труда. На этом основании можно обозначить угрозу имущественных интересов всей группы авторов как вида субъектов гражданских правоотношений, что позволяет определить публично-правовую природу данной проблемы.

Своей спецификой обладает проблема легальности Обработки ИИС Больших данных. В случае квалификации указанной Обработки в качестве нарушения авторских прав по указанным в настоящем исследовании основаниям представляется затруднительным идентифицировать каждого автора у каждого произведения в силу того, что в Больших данных содержится колоссальный массив данных, в том

⁵ AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit. // Vox Media. 2023. URL: <https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart>. (дата обращения: 30.09.2023).

⁶ Генеративные нейросети и этика: появилась модель, копирующая стиль конкретного художника // DTF.Ru. URL: <https://dtf.ru/life/1436360-generativnye-neyroseti-i-etika-poyavilas-model-kopiruyushchaya-stil-konkretnogo-hudozhnika>. (дата обращения: 30.09.2023).

числе произведения, у каждой из которых есть автор. Различного рода технические методы поиска и идентификации авторов будет противоречить международным и национальным принципам и условиями обработки персональных данных, в частности, принцип минимизации данных, который налагает на контролеров обязанность собирать исключительно данные, отвечающие поставленной цели обработки, и хранить их только в течение срока, необходимого для реализации цели обработки, и принцип ограниченного целевого назначения, который устанавливает подчиненность целей обработки данных ее легитимным основаниям [11, с. 177–178].

М.В. Тюнин при рассмотрении вышеуказанной проблемы перечисляет потенциальный перечень решений, которые могли бы в некоторой мере легализировать машинное обучение:

- установление в качестве обязательного условия получения согласия каждого автора на обработку его произведений ИИС;
- введение специального правового режима свободного использования объектов авторского права в процессе машинного обучения в некоммерческих целях или для научных исследований;
- применение принудительной лицензии к обрабатываемым объектам авторского права с компенсационной выплатой правообладателю [12, с. 39; 13].

Первое решение реализовать будет крайне сложно, так как Обработка произведений происходит в немалых масштабах, особенно если речь идет об Обработке Больших данных. Данное решение возможно будет реализовать, если изначально при загрузке на соответствующие интернет-площадки брать согласие с автора на Обработку ИИС. Однако в таком случае все ранее загруженные в «Интернет» произведения окажутся без защиты, а обратной силы такое предписание иметь не должно, так как реализовать подобную меру на практике невозможно.

Внедрение правового режима свободного использования представляется практически реализуемым и отвечающим запросам широких общественных масс. Однако в таком случае сужаются пределы исключительного права на произведение и тем самым игнорируются имущественные интересы правообладателей в том смысле, что правообладатель никогда не сможет полностью контролировать дальнейшее использование результатов машинного обучения. К тому же, как было выявлено в настоящем исследовании, Обработка произведений ИИС — это процесс со сложной структурой; каждый этап может отвечать критериям для квалификации в качестве нарушения авторских прав. В силу высокой вероятности появления самостоятельного экономического значения использования произведения на каждом последующем этапе Обработки, норма, содержащая случаи свободного использования произведения, носила бы в подавляющей степени формальный характер и способствовала бы легализации свободного использования лишь в научно-исследовательских целях. На практике Обработка произведений ИИС зачастую не ограничивается исключительно научно-исследовательскими либо некоммерческими целями. Например, доступ к использованию современных популярных нейросетей нередко является возмездным⁷. Таким образом, данное решение можно использовать, но только в рамках комплексной реформы по легализации Обработки произведений ИИС, а не в качестве единичной изменения законодательства. Следовательно, требуется иной правовой режим Обработки произведений ИИС.

Принудительная лицензия на первый взгляд выглядит приемлемой мерой, учитывающей интересы правообладателя и широких общественных масс. Тем не менее, возникает аналогичная проблема, как с получением согласий на Обработку произведений ИИС в первом решении. Согласно положениям ст. 1239 ГК РФ в

⁷ Топ 100 платных нейросетей. Крупнейший каталог нейросетей с возможностью подбора и сравнения // AI Directory. URL: <https://aidir.ru/ai-price/platno/>. (дата обращения 30.09.2023).

случаях, предусмотренных ГК РФ, суд может по требованию заинтересованного лица принять решение о предоставлении этому лицу на указанных в решении суда условиях права использования результата интеллектуальной деятельности, исключительное право на который принадлежит другому лицу. Как следует из содержания данной статьи, для правомерной Обработки произведений ИИС потребуется соответствующее решение суда относительно каждого произведения. Конечно, ничего не мешает в рамках одного решения указать несколько объектов авторского права, однако в рамках Обработки произведений ИИС потребуется множество произведений, авторов или правообладателей которых далеко не всегда представляется возможным идентифицировать. Это обстоятельство также не позволяет определить лицо, которому подлежит выплата компенсации. Поэтому обращение в суд представляется лишней бюрократической процедурой, которая не способствует разрешению данной проблемы. Из этого можно сделать вывод, что применение принудительной лицензии не решает проблему легализации Обработки произведений ИИС.

На основании вышеизложенного представляется, что меры гражданско-правового характера для решения исследуемой проблемы неэффективны. Легализовать Обработку произведений ИИС следует в первую очередь на публично-правовом уровне. В силу общих технических методов обработки персональных данных и произведений (так как произведения принимают форму данных) предлагается использовать схожую логику при формулировании правовых норм для регулирования Обработки произведений ИИС.

Для начала следует сформулировать правовой статус оператора ИИС в контексте Обработки произведений ИИС. За основу можно взять положения Федерального закона от 27.07.2006 г. № 152-ФЗ «О персональных данных» (далее — ФЗ № 152) и положения GDPR (англ. «General Data Protection Regulation»; рус. Общий регламент по защите данных), так как предметом указанных актов является обработка данных, связанных с субъективными правами лиц.

Согласно дефиниции, указанной в ст. 3 Федерального закона «О персональных данных» (далее — ФЗ № 152), оператор — это государственный орган, муниципальный орган, юридическое или физическое лицо, самостоятельно или совместно с другими лицами организующие и (или) осуществляющие обработку персональных данных, а также определяющие цели обработки персональных данных, состав персональных данных, подлежащих обработке, действия (операции), совершаемые с персональными данными. Примечательно, что в GDPR роль оператора разделена на двух разных субъектов: «controller» (контроллер) и «processor» (обработчик). Согласно ст. 4 GDPR контроллер определяет цели и средства обработки, а обработчик, в свою очередь, по поручению контроллера занимается обработкой персональных данных. Обработчиком, в отличие от контроллера, может быть исключительно физическое или юридическое лицо. Таким образом, по ключевым вопросам статуса Оператора обработки персональных данных расхождений между GDPR и ФЗ № 152 не наблюдается, в российском законодательстве функции контроллера и обработчика включены в правовой статус Оператора.

Так как в рамках настоящего исследования делается акцент на российское законодательство, будет использован подход одного правового статуса — Оператора ИИС. Под Оператором ИИС предлагается понимать государственный орган, муниципальный орган, юридическое или физическое лицо, самостоятельно или совместно с другими лицами организующие и (или) осуществляющие создание ИИС (включая машинное обучение) и практическое применение по назначению. Таким образом, исходя из указанного определения можно выявить основные положения относительно статуса Оператора ИИС. Ключевым является презумпция ответствен-

ности Оператора ИИС за неправомерную (осуществленную с нарушением требований законодательства в области авторских прав) Обработку произведений ИИС.

Требуется закрепления и статус «Пользователя ИИС», так как с одной стороны данная категория лиц пользуется результатами Обработки произведений ИИС, а с другой стороны не имеет отношения к обучению модели ИИС и поэтому не может нести ответственность за использование полученных незаконным путем результатов, идентичных результатам творческого труда. Из вышеуказанного можно выявить ключевые признаки Пользователя: имеет законный доступ к ИИС; пользуется функциями ИИС по целевому назначению; не причастен к машинному обучению. Вопрос, может ли Пользователем выступать юридическое лицо в предмет настоящего исследования не входит, поэтому условно следует допустить такой вариант, однако при этом презюмировать, что правовая природа отношений между Оператором ИИС и Пользователем ИИС в сущности своей потребительская, если не доказано иное.

На основании изложенного в настоящем исследовании можно сформулировать нижеследующие принципы Обработки произведений ИИС, которые необходимо закрепить на законодательном уровне.

Во-первых, указанный ранее принцип ограниченного целевого назначения подходит для адаптации под Обработку произведения ИИС. Согласно данному принципу Обработка произведений ИИС должна осуществляться исключительно в целях, установленных Оператором ИИС. Каждое повторное использование тех же наборов данных должно проверяться на соответствие установленной цели. Этот принцип также подразумевает, что цели должны носить законный характер.

Во-вторых, вполне возможна адаптация принципа ограниченного хранения, заключающийся в том, что персональные данные должны храниться в доступной для идентификации субъектов этих данных форме, но не дольше, чем этого требует достижение поставленных целей обработки [10, с. 178–179]. Применительно к Обработке произведений ИИС данный принцип выражается в расширении правил воспроизведения произведений в памяти ИИС в процессе машинного обучения, в дополнение к положениям подп. 1 п. 2 ст. 1270 ГК РФ. Однако несмотря на видимое расширение пределов правомерного осуществления Обработки произведений ИИС, хранение обрабатываемых наборов данных с произведениями осуществляется исключительно в пределах конкретной цели — создании обученной модели машинного обучения, т.е. ИИС.

В-третьих, следует добавить принцип прозрачной отчетности, заключающийся в необходимости ведения какого-либо реестра или иной формы отчетности относительно количества и качества обрабатываемых наборов данных с произведениями, доступной для любого лица, вне зависимости от заинтересованности такого лица.

По результатам настоящего исследования можно сделать следующие выводы. Обработка произведений ИИС (в том числе в составе Больших данных), состоит из трех структурных элементов: TDM, машинное обучение (в узком смысле) и создание результатов, идентичных результатам творческого труда. Каждый из указанных процессов при определенных условиях, исследованных в настоящей статье, возможно квалифицировать как нарушение авторских прав. Ситуация усугубляется тем, что при Обработке Больших данных идентифицировать лиц, чьи права были нарушены, в крайней степени затруднительно даже с применением специальных технологий. Вычислительные мощности ИИС позволяют создавать результаты, идентичные результатам творческого труда, с чрезвычайно высокой скоростью, что способствует вытеснению авторов, не использующих ИИС, с рынков определенных произведений. При этом ИИС обучается на основе произведений вытесняемых авторов, что закрепляет факт нарушения имущественных прав.

Частно-правовые меры по решению проблем Обработки произведений ИИС представляются неэффективными, поэтому следует прибегнуть к публично-правовым механизмам. В качестве одной из необходимых мер предлагается внедрить в законодательство правовой статус Оператора ИИС и Пользователя ИИС, а также сформулировать принципы Обработки произведений ИИС. Указанными мерами не следует ограничиваться, на их основе представляется возможным сформулировать конкретные положения норм права, регулирующих правила Обработки произведений ИИС.

Список использованной литературы


1. Семин П.О. Правовые аспекты искусственного интеллекта и смежных технологий: права на контент, созданный с помощью машинного обучения / П.О. Семин. — EDN HPOJBV // Журнал Суда по интеллектуальным правам. — 2022. — № 2(36). — С. 21–32.
2. Карпычев В.Ю. Правовое регулирование больших данных: пусть будет / В.Ю. Карпычев. — DOI 10.18572/1812-3929-2022-4-68-73. — EDN KFXBKW // Юрист. — 2022. — № 4. — С. 68–73.
3. Дмитриев А.С. Big Data, 4v: Volume, Velocity, Variety, Value / А.С. Дмитриев // Мониторинг общественного мнения. — 2015. — № 2(126). — С. 156–159.
4. Афанасьева Е.Н. Введение в индустрию 4.0: основы формирования цифрового будущего : монография / Е.Н. Афанасьева. — Томск : ТУСУР, 2021. — 93 с.
5. Пилецкая А.В. Искусственный интеллект и большие данные / А.В. Пилецкая. — EDN PNNGHD // Молодой ученый. — 2019. — № 50 (288). — С. 20–22.
6. Christensen K. A European solution for Text and Data Mining in the development of creative Artificial Intelligence / K.A. Christensen // Publicerad i Stockholm IP Law Review. — 2021. — No. 2. — P. 18–33.
7. Кольздорф М.А. Свободное использование объектов авторских и смежных прав при обработке больших данных (Big Data) / М.А. Кольздорф. — EDN YOYBOG // Закон. — 2021. — № 5. — С. 142–164.
8. Свободное использование объектов авторских и смежных прав в условиях развития цифровых технологий на примере технологий TDM / А.И. Тиунова, Ю.К. Павлова-Семион, С.К. Романова [и др.]. — EDN GRFWQI // Журнал Суда по интеллектуальным правам. — 2022. — № 1(35). — С. 29–58.
9. Грачева Д.А. Свободное использование авторских и смежных прав в условиях развития цифровых технологий в Российской Федерации / Д.А. Грачева. — DOI 10.17323/tis.2023.17371. — EDN MPYQCS // Труды по интеллектуальной собственности. — 2023. — Т. 45, № 2. — С. 44–52.
10. Мотовилова Д.А. Ключевые положения Директивы (ЕС) № 2019/790 об авторском праве и смежных правах на едином цифровом рынке / Д.А. Мотовилова. — EDN DMXOEN // Журнал Суда по интеллектуальным правам. — 2019. — № 26. — С. 106–120.
11. Крылова М.С. Принципы обработки персональных данных в праве Европейского Союза / М.С. Крылова. — DOI 10.17803/1994-1471.2017.83.10.175-181. — EDN ZRIRQV // Актуальные проблемы российского права. — 2017. — № 10 (83). — С. 175–181.
12. Тюнин М.В. Свойство искусственного интеллекта вписываться в общую систему правового регулирования интеллектуальной собственности в ЕАЭС / М.В. Тюнин. — DOI 10.17323/tis.2021.13503. — EDN KIMIUB // Труды по интеллектуальной собственности. — 2021. — Т. 39, № 4. — С. 37–41.
13. Артений Л.С. Искусственный интеллект в авторском праве / Л.С. Артений. — EDN LBRCCT // Вестник науки и образования. — 2019. — № 7-1. — С. 76–81.


References

1. Syomin P.O. Legal Aspects of Artificial Intelligence and Related Technologies: Rights to Content Created With Machine Learning. *Zhurnal Suda po intellektual'nym pravam = Journal of the Intellectual Property Rights Court*, 2022, no. 2, pp. 21–32. (In Russian). EDN: HPOJBV.
2. Karpychev V. Yu. The Legal Regulation of Big Data: Just In Case. *Yurist = Jurist*, 2022, no. 4, pp. 68–73. (In Russian). EDN: KFXBKW. DOI: 10.18572/1812-3929-2022-4-68-73.


3. Dmitriev A.S. Big Data, 4v: Volume, Velocity, Variety, Value. *Monitoring obshchestvennogo mneniya = Monitoring of Public Opinion*, 2015, no. 2, pp. 156–159. (In Russian).
4. Afanas'eva E.N. *Introduction to Industry 4.0: The Basics of Shaping the Digital Future*. Tomsk, TUSUR Publ., 2021. 93 p.
5. Piletskaya A.V. Artificial Intelligence and Big Data. *Molodoi uchenyi = Young Scientist*, 2019, no. 50, pp. 20–22. (In Russian). EDN: PNNGHD.
6. Christensen K. A European solution for Text and Data Mining in the development of creative Artificial Intelligence. *Publicerad i Stockholm IP Law Review*, 2021, no. 2, pp. 18–33.
7. Kol'zendorf M.A. Free Use of Copyright and Related Rights In Big Data Processing. *Zakon = Law*, 2021, no. 5, pp. 142–164. (In Russian). EDN: YOYBOG.
8. Tiunova A.I., Pavlova-Semion Yu.K., Romanova S.K., Bryukhanov P.V., Obolonina Yu.O. Free Use Of Copyrighted and Related Rights Works in the Context of the Development of Digital Technologies on the Example of TDM Technologies. *Zhurnal Suda po intellektual'ny'm pravam = Journal of the Intellectual Property Rights Court*, 2022, no. 1, pp. 29–58. (In Russian). EDN: GRFWQI.
9. Gracheva D.A. Free Use of Copyright and Related Rights in the Context of the Development of Digital Technologies in the Russian Federation. *Trudy po intellektual'noi sobstvennosti = Works on Intellectual Property*, 2023, vol. 45, no. 2, pp. 44–52. (In Russian). EDN: MPYQCJ. DOI: 10.17323/tis.2023.17371.
10. Motovilova D.A. Key Provisions of Directive (Eu) 2019/790 on Copyright and Related Rights in the Digital Single Market. *Zhurnal Suda po intellektual'ny'm pravam = Journal of the Intellectual Property Rights Court*, 2019, no. 26, pp. 106–120. (In Russian). EDN: DMX-OEH.
11. Krylova M.S. Principles of Processing Personal Data in the European Union Law. *Aktual'nye problemy rossiiskogo prava = Topical Problems of Russian Law*, 2017, no. 10, pp. 175–181. (In Russian). EDN: ZRIRQV. DOI: 10.17803/1994-1471.2017.83.10.175-181.
12. Tyunin M.V. The Property of Artificial Intelligence to Fit Into the General System of Legal Regulation of Intellectual Property. *Trudy po intellektual'noi sobstvennosti = Works on Intellectual Property*, 2021, vol. 39, no. 4, pp. 37–41. (In Russian). EDN: KIMIUB. DOI: 10.17323/tis.2021.13503.
13. Artenii L.S. Artificial Intelligence in Copyright Law. *Vestnik nauki i obrazovaniya = Herald of Science and Education*, 2019, no. 7-1, pp. 76–81. (In Russian). EDN: LBRCCT.


Информация об авторах

Афанасьева Екатерина Нодариевна — кандидат юридических наук, PhD, доцент, кафедра информационного, гражданского права и правового обеспечения инновационной деятельности, старший научный сотрудник проектной научно-исследовательской лаборатории «Развитие средств и методов электронного государства», Томский государственный университет систем управления и радиоэлектроники, г. Томск, Российская Федерация, afeka@inbox.ru,  <https://orcid.org/0000-0002-3885-4357>, SPIN-код: 8587-9508, AuthorID РИНЦ: 590763, ResearcherID: O-8344-2014.

Фурман Дмитрий Дмитриевич — магистрант программы «Право интеллектуальной собственности и цифровые технологии», Новосибирский государственный университет, г. Новосибирск, Российская Федерация, dmitryfurman@mail.ru,  <https://orcid.org/0009-0005-4356-9816>.

Authors

Ekaterina N. Afanasyeva — PhD in Law, Associate Professor, Department of Information, Civil Law and Legal Support of Innovation; Senior Researcher, Project Research Laboratory “E-Government Tools and Methods”, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation, afeka@inbox.ru,  <https://orcid.org/0000-0002-3885-4357>, SPIN-Code: 8587-9508, AuthorID RSCI: 590763, ResearcherID: O-8344-2014.

Dmitry D. Furman — Master's Degree Student, Program “Intellectual Property Law and Digital Technologies”, Novosibirsk State University, Novosibirsk, Russian Federation, dmitryfurman@mail.ru,  <https://orcid.org/0009-0005-4356-9816>.

Для цитирования

Афанасьева Е.Н. Правовые аспекты обработки искусственными интеллектуальными системами объектов авторского права в процессе машинного обучения / Е.Н. Афанасьева, Д.Д. Фурман. — DOI 10.17150/2411-6262.2023.14(4).1643-1656. — EDN YOZDTD // *Baikal Research Journal*. — 2023. — Т. 14, № 4. — С. 1643–1656.

For Citation

Afanasyeva E.N., Furman D.D. Legal Issues of Copyright Objects Processing by AI Systems in the Process of Machine Learning. *Baikal Research Journal*, 2023, vol. 14, no. 4, pp. 1643–1656. (In Russian). EDN: YOZDTD. DOI: 10.17150/2411-6262.2023.14(4).1643-1656.