Искусственный интеллект

Проблемы правовой охраны персональных данных в процессе использования нейронных сетей

Аннотация. *Развитие высокотехнологичных отраслей экономики во многом обусловлено формированием системного правового регулирования складывающихся отношений в цифровой среде, в том числе отдельных сквозных технологий. Одними из важнейших таких технологий выступает искусственный интеллект и неразрывно связанные с ним нейронные сети. Данная статья посвящена анализу одной из ключевых проблем в сфере использования нейронных сетей — защите персональных данных. Авторы приходят к выводу, что в законодательстве о персональных данных должно быть закреплено, что обработка персональных данных с использованием нейронных сетей должна производиться только при условии письменного согласия субъекта персональных данных. Кроме того, любые изменения персональных данных, произведенные в результате процессов машинного обучения, должны регулярно согласовываться с субъектом персональных данных. Ему должна быть обеспечена возможность ознакомления с обновленными персональными данными.*

Ключевые слова: *нейронные сети, персональные данные, правовое регулирование, охрана.*

DOI: 10.17803/2311-5998.2023.102.2.071-079

Алексей Владимирович МИНБАЛЕЕВ,
заведующий кафедрой информационного права и цифровых технологий Университета имени О.Е. Кутафина (МГЮА), доктор юридических наук, профессор
alexmin@bk.ru
125993, Россия, г. Москва, ул. Садовая-Кудринская, д. 9

Екатерина Эдуардовна СТОРОЖАКОВА,
директор бюро переводов Translate2u
mail@2u.com.ru
125993, Россия, г. Москва, ул. Садовая-Кудринская, д. 9

ALEKSEY V. MINBALEEV,
Head of the Department of information law and digital technologies
of the Kutafin Moscow State Law University (MSAL),
Dr. Sci. (Law), Professor
alexmin@bk.ru
9, ul. Sadovaya-Kudrinskaya, Moscow, Russia, 125993

EKATERINA E. STOROZHAKOVA,
Director of the translation agency Translate2u
mail@2u.com.ru
9, ul. Sadovaya-Kudrinskaya, Moscow, Russia, 125993

Problems of legal protection of personal data in the process of using neural networks

Abstract. *The development of high-tech sectors of the economy is largely due to the formation of a systematic legal regulation of the emerging relations in the digital environment, including the regulation of certain end-to-end technologies. One of the most important such technologies is artificial intelligence and inextricably linked neural networks. This article is devoted to the analysis of one of the key problems in the use of neural networks — the protection of personal data. The authors conclude that the legislation on personal data should stipulate that the processing of personal data using neural networks should be carried out only with the written consent of the subject of personal data. In addition, any changes to personal data made as a result of machine learning processes must be regularly coordinated with the subject of personal data. He should be provided with the opportunity to get acquainted with the updated personal data.*
Keywords: *neural networks, personal data, legal regulation, protection.*

Нейронные сети являются сегодня ключевым инструментом и технологией, направленной на обеспечение использования искусственного интеллекта и машинного обучения. В современном гражданском обороте все чаще возникают вопросы договорного сопровождения использования тех или иных нейронных сетей в процессе функционирования систем искусственного интеллекта. Происходит это чаще всего через различные создаваемые приложения, устанавливаемые клиентами в компьютерных устройствах.

Активно используются и смарт-контракты, с помощью которых автоматически происходит обработка той или иной информации нейронными сетями, например автоматические актуализации по тому или иному запросу на предмет поиска определенного варианта действий на основе произошедших изменений и дополнений в онтологии за счет как поступившей информации, так и произведенного машинного самообучения системой. Обеспечение договорного процесса в сфере услуг, связанных с искусственным интеллектом, облачной платформой и

нейронными сетями, сегодня является весьма актуальным и вызывает на практике много вопросов и проблем[1].

Сегодня постепенно приходит осознание того, что сами нейронные сети не могут регулироваться в отрыве от технологий искусственного интеллекта. В то же время отдельные аспекты правовой природы нейронных сетей, как и всех сквозных технологий, необходимо учитывать в процессе правотворческой и правоприменительной деятельности[2]. К сожалению, сегодня приходится констатировать, что исследований[3], посвященных нейронным сетям, ничтожно мало, в то время как тема является очень важной с позиции юридического анализа.

С точки зрения защиты персональных данных при использовании нейронных сетей возникает несколько проблем.

1. Существуют проблемы законности обработки тех или иных персональных данных, которые используются в онтологии или при разработке алгоритма, а также законности использования тех или иных персональных данных при формировании запроса к нейронной сети; обеспечения и защиты прав граждан на персональные данные и их использование. В данном случае в основном используются классические механизмы защиты и, если отсутствует согласие на обработку персональных данных и она осуществляется вне обязательных требований, установленных федеральным законом, то имеется факт нарушения.

Персональные данные могут содержаться как в онтологии системы (по сути, в данном случае мы можем говорить об аналогии с охраной персональных данных в определенной базе данных или информационной системе и прибегать к соответствующим инструментам охраны).

Персональные данные могут содержаться и в самом алгоритме, и эта ситуация гораздо сложнее, поскольку чаще всего в алгоритме заложены косвенные персональные данные. Термин «алгоритм» часто используется для обозначения приложений искусственного интеллекта, например, через такие термины,

---

[1] Drew Stevens (2018) Neural Networks and Advanced AI Contract Issues // URL: https://drewstevenslaw.com/neural-networks-and-advanced-ai-contract-issues/ (дата обращения: 10.12.2022).

[2] См.: Право и иные регуляторы в развитии цифровых технологий / А. В. Минбалеев, А. В. Мартынов, Г. Г. Камалова [и др.]. Саратов : Амирит, 2022 ; Проблемы и вызовы цифрового общества: тенденции развития правового регулирования цифровых трансформаций : колл. монография : в 2 т. / Е. Н. Абанина, О. Ю. Авдевнина, Р. В. Амелин [и др.] Саратов : Саратовская государственная юридическая академия, 2020 ; Информационно-технологическое обеспечение юридической деятельности (LegalTech) : учебник / А. В. Минбалеев, Т. А. Полякова, М. Б. Добробаба [и др.] М. : Проспект, 2023.

[3] См.: *Бойченко И. С.* Модели правового регулирования нейросетей // Образование и право. 2019. № 1. С. 235—237 ; *Он же.* Формирование правового регулирования нейронных сетей // Формирование системы правового регулирования обеспечения информационной безопасности в условиях больших вызовов в глобальном информационном обществе. Вторые Бачиловские чтения : сборник научных трудов, Москва, 8 февраля 2019 г. М. ; Саратов : Амирит, 2019. С. 218—223 ; *Полякова Т. А.*, *Химченко А. И.* Правовые проблемы обеспечения информационной безопасности при использовании облачных технологий // Правовая информатика. 2013. № 2. С. 12—16.

как «алгоритмическое принятие решений». Однако понятие алгоритма является более общим, чем понятие искусственного интеллекта, поскольку оно включает в себя любую последовательность однозначно определенных инструкций для выполнения задачи, в частности, но не исключительно, с помощью математических вычислений.

Чтобы быть выполняемыми компьютерной системой, алгоритмы должны быть выражены с помощью языков программирования и стать, таким образом, машинно-исполняемыми программами. Алгоритмы могут быть очень простыми, указывая, например, как упорядочивать списки слов в алфавитном порядке или как найти наибольший общий делитель между двумя числами (например, так называемый алгоритм Евклида). Они также могут быть очень сложными, такими как алгоритмы шифрования файлов, сжатия цифровых файлов, распознавания речи или финансового прогнозирования. В данном случае как раз проблема и состоит в том, что персональные данные будут находиться в зашифрованном виде и не доступны простому субъекту персональных данных, который не в состоянии их выявить и проверить.

Очевидно, что не все алгоритмы включают искусственный интеллект, но каждая система искусственного интеллекта, как и любая компьютерная система, включает в себя алгоритмы, часть которых имеет дело с задачами, непосредственно касающимися функций искусственного интеллекта. Алгоритмы искусственного интеллекта могут включать различные виды эпистемологических или практических рассуждений (обнаружение паттернов и формы, применение правил, составление прогнозов или планов), а также различные способы обучения.

Все это может осуществляться с использованием отдельных персональных данных, например, при формировании алгоритмов, связанных с таргетинговой рекламой, основанной на вкусах и предпочтениях потребителей. Причем в большинстве случаев, конечно, персональные данные будут обезличены, но в ряде случаев — нет. Логично, что и нейронная сеть, функционирующая на базе таких алгоритмов, может использовать персональные данные, в том числе нарушая права тех или иных субъектов.

Нейронная сеть, используя данные пользователей, может улучшить себя, разработав новые эвристики (предварительные стратегии решения проблем), модифицируя свои внутренние данные или даже генерируя новые алгоритмы. Например, система искусственного интеллекта для электронной коммерции может предоставлять скидки потребителям, отвечающим определенным условиям (применять правила), предоставлять рекомендации (например, изучать и использовать корреляции между особенностями пользователей и их покупательскими привычками), оптимизировать управление запасами (например, разрабатывать и внедрять лучшие торговые стратегии). При этом, естественно, закладывается необходимость обработки персональных данных потребителей[4].

[4] Подробнее см.: *Минбалеев А. В.* Проблемы социальной эффективности и защиты прав человека при использовании искусственного интеллекта в рамках социального скоринга // Вестник Южно-Уральского государственного университета. Серия : Право. 2020. Т. 20. № 2. С. 96—101.

Хотя система искусственного интеллекта включает в себя множество алгоритмов, ее также можно рассматривать как единый сложный алгоритм, объединяющий алгоритмы, выполняющие ее различные функции, а также верхние алгоритмы, которые управляют функциями системы путем активации соответствующих алгоритмов более низкого уровня. Например, бот, который отвечает на запросы на естественном языке, будет включать в себя организованную комбинацию алгоритмов для обнаружения звуков, захвата синтаксических структур, извлечения соответствующих знаний, построения выводов, генерации ответов и т.д.

2. Наиболее проблемным является вопрос о создании новых данных, полученных в результате самообучения нейронной сети. В системе, способной к обучению, особо важным компонентом будет не изученная алгоритмическая модель, т.е. алгоритмы, которые непосредственно выполняют задачи, возложенные на систему (например, составление классификаций, прогнозов или решений), а, скорее, алгоритмы обучения, которые модифицируют алгоритмическую модель, чтобы она лучше выполняла его функцию. Например, в системе классификатора, которая распознает изображения с помощью нейронной сети, решающим элементом является алгоритм обучения (тренажер), который изменяет внутреннюю структуру алгоритмической модели (обученная нейронная сеть). изменяя его (изменяя его внутренние связи и веса) таким образом, чтобы он правильно классифицировал объекты в своей области (например, животных, звуки, лица, позы и т.д.).

Алгоритм обучения системы (ее тренажер) использует обучающий набор для построения алгоритмической модели: нейронную сеть, дерево решений, набор правил и т.д. Алгоритмическая модель предназначена для сбора соответствующих знаний, изначально заложенных в обучающий набор, а именно корреляций между случаями и ответами. Затем эта модель используется алгоритмом прогнозирования для предоставления, как всегда надеются разработчики, правильных ответов на новые случаи, имитируя корреляции в обучающем наборе.

Если примеры в обучающий набор, наиболее близкий к новому случаю (с точки зрения соответствующих функций), связан с определенным ответом, тот же ответ будет предложен для нового случая. Например, если картинки, наиболее похожие на новый ввод, были помечены как кошки, то и новый ввод будет помечен таким же образом; если прошлые кандидаты, характеристики которых наилучшим образом соответствуют характеристикам нового кандидата, были связаны с отклонением, система предложит отклонить также нового кандидата; если прошлые работники, которые ближе всего подходят к новому кандидату, показали хорошие (или плохие) результаты, системы предсказывают, что и заявитель будет действовать аналогичным образом[5]. Все это требует особого подхода с позиции выработки новых персональных данных. Необходимо учитывать согласие и иные законные основания для подобной обработки персональных данных, а также возможность их использования в измененной в ходе машинного обучения форме.

[5] The impact of the General Data Protection Regulation (GDPR) on artificial intelligence / European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 641.530 June 2020. P. 4, 11.

В последнее время глубокое обучение, основанное на многоуровневых нейронных сетях, было очень успешно внедрено — особенно, но не исключительно — там, где необходимо распознавать паттерны и связывать их с классификациями и решениями (например, при обнаружении объектов на изображениях, распознавании звуков и их источников, постановке медицинского диагноза, переводе текстов, выборе стратегий в играх и т.д.). Нейронные сети состоят из набора узлов, называемых нейронами, расположенных в несколько слоев и соединенных ссылками. Они так называются, поскольку воспроизводят некоторые аспекты нервной системы человека, действительно состоящую из взаимосвязанных специализированных клеток, биологических нейронов, которые получают и передают информацию.

Нейронные сети и были разработаны в предположении, что искусственный интеллект может быть достигнут путем воспроизведения человеческого мозга, а не путем моделирования человеческого мышления, т.е. что искусственное мышление естественным образом возникнет из искусственного мозга. Хотя может встать вопрос: в какой степени искусственные нейронные сети и человеческий мозг действительно имеют сходные структуры и процессы. Каждый нейрон получает сигналы (числа) от подключенных нейронов или извне, и эти сигналы усиливаются или уменьшаются по мере пересечения входящих связей, в соответствии с весами последних.

Нейрон применяет некоторые вычисления к входным данным, которые он получает, и, если результат достигает порогового значения нейрона, нейрон становится активным, посылая сигналы подключенным нейронам или снаружи из сети. Активация начинается с узлов, получающих внешние входные данные, и распространяется по сети. Обучение сети происходит путем информирования сети о том, являются ли ее ответы (ее выходные данные) правильными или неправильными. Если ответ сети неверен, алгоритм обучения обновляет сеть, т.е. корректирует веса соединений, так что в следующий раз, когда сети будут предоставлены эти входные данные, она выдаст правильный ответ.

В случае нейронной сети алгоритм обучения модифицирует сеть до тех пор, пока она не достигнет желаемого уровня производительности, в то время как результатом обучения — алгоритмической моделью — является сеть в ее окончательной конфигурации.

Как отмечалось ранее, алгоритм обучения способен модифицировать нейронную сеть таким образом, чтобы сеть была способна предоставлять наиболее подходящие ответы. При контролируемом подходе к обучению обученная сеть будет воспроизводить поведение в обучающий набор; в соответствии с подходом к обучению с подкреплением сеть будет придерживаться поведения, которое максимизирует ее эффективность (пример — бонусные баллы, связанные с прибылью от инвестиций или с победами в играх).

Важная проблема состоит еще и в том, что «нейронная сеть не дает объяснений своим результатам. Можно определить, как данный результат был получен в результате активации сети и как эта активация в ответ на заданный ввод была определена соединениями между нейронами (и весами, присвоенными таким соединениям в результате обучения сети) и математическими функциями, управляющими каждым нейроном. Однако эта информация не показывает логического

обоснования, значимого для людей: она не говорит нам, почему был дан определенный ответ. В связи с этим важно понимать, что субъекту персональных данных оператор не может указать, почему получена именно такая информация.

Существует множество подходов к объяснению поведения нейронных сетей и других непрозрачных систем (также называемых «черными ящиками»). Некоторые из этих подходов рассматривают систему, подлежащую объяснению, и строят объяснения соответствующим образом. Другие подходы строят объяснения на основе внешнего поведения сети: они рассматривают только связь между входными данными, предоставляемыми сетью, и результатами, которые она обеспечивает, и, соответственно, строят аргументы или другие объяснения.

Однако достижения в объяснении нейронных сетей, понятных человеку, до сих пор были весьма ограниченными. К сожалению, во многих областях системы, функционирование которых в меньшей степени является системой машинного обучения, наилучший баланс также зависит от области, в которой используется система, и от важности затрагиваемых интересов. Когда речь идет о публичных действиях и на карту поставлены ключевые интересы человека (например, как в судебных решениях), объяснение имеет первостепенное значение.

Однако даже тогда, когда систему можно рассматривать только как «черный ящик», некоторый критический анализ ее поведения все еще возможен. С помощью анализа чувствительности, т.е. систематической проверки того, изменяется ли результат при изменении значения определенных входных характеристик, оставляя все остальные характеристики неизменными, мы можем понять, какие функции определяют выходные данные системы. Например, проверяя, изменяется ли прогноз системы, предназначенной для оценки кредитоспособности, если мы изменяем место рождения или жительства заявителя, мы можем определить, имеет ли эта входная функция отношение к выходным данным системы. Следовательно, мы можем задаться вопросом, не является ли система чрезмерно дискриминирующей людей в зависимости от их этнической принадлежности или социального статуса, который может быть связан с местом рождения или проживания[6].

Особую обеспокоенность вызывает вопрос о защите информации, сохраняемой нейронными сетями, в том числе персональных данных. Нейронные сети могут сохранять части набора данных, используемые для обучения и генерации текста. Вредоносные объекты могут проникнуть в сеть для получения конфиденциальных данных, таких как номера кредитных карт и номера социального страхования. «В идеале, даже если бы обучающие данные содержали редкую, но конфиденциальную информацию о некоторых отдельных пользователях, нейронная сеть не запоминала бы эту информацию и никогда не выдавала бы ее как завершение предложения»[7].

Нужно также учитывать, что это может быть преднамеренным и непреднамеренным запоминанием. Когда речь идет о непреднамеренном запоминании,

[6] The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. P. 14—15.

[7] *Nicholas Carlini, Chang Liu, Úlfar Erlingsson, Jernej Kos, Dawn Song*. (2019) The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Network // URL: https://doi.org/10.48550/arXiv.1802.08232.

модели глубокого обучения (в частности, генеративные модели), по-видимому, часто запоминают редкие детали обучающих данных, которые совершенно не связаны с предполагаемой задачей, в то время как модель все еще изучает базовое поведение (т.е. в то время как потеря теста все еще уменьшается). Такое непреднамеренное запоминание редких деталей обучения может вызвать серьезные проблемы с конфиденциальностью при использовании конфиденциальных данных для обучения моделям глубокого обучения.

Самое тревожное, что такое запоминание может произойти даже для примеров, которые присутствуют в обучающих данных всего несколько раз, особенно когда эти примеры являются выбросами в распределении данных; это верно даже для языковых моделей, которые используют современные методы регуляризации для предотвращения традиционных форм переобучения и перетренированности. На сегодняшний день не существует хорошего метода, помогающего практикующим специалистам измерить степень запоминания моделью аспектов обучающих данных[8]. В случае преднамеренного запоминания речь должна идти об административной или уголовной ответственности.

Сложности, связанные с выявлением фактов использования персональных данных в нейронных сетях и их модификаций, обусловливают ограничения в возможности защиты прав на персональные данные, в том числе их гражданско-правовой защиты и возможного возмещения причиненного ущерба и компенсации морального вреда. В связи с этим важно обеспечить законодательные требования к информированию субъектов персональных данных.

В связи с этим нам представляется, что в законодательстве о персональных данных должно быть закреплено, что обработка персональных данных с использованием нейронных сетей должна производиться только при условии письменного согласия субъекта персональных данных. Кроме того, любые изменения персональных данных, произведенные в результате процессов машинного обучения, должны регулярно согласовываться с субъектом персональных данных. Ему должна быть обеспечена возможность ознакомления с обновленными персональными данными.

БИБЛИОГРАФИЯ

1. *Бойченко И. С.* Модели правового регулирования нейросетей // Образование и право. — 2019. — № 1. — С. 235—237.
2. *Бойченко И. С.* Формирование правового регулирования нейронных сетей // Формирование системы правового регулирования обеспечения информационной безопасности в условиях больших вызовов в глобальном информационном обществе. Вторые Бачиловские чтения : сборник научных трудов. Москва, 8 февраля 2019 г. — М. ; Саратов : Амирит, 2019. — С. 218—223.
3. Информационно-технологическое обеспечение юридической деятельности (LegalTech) : учебник / А. В. Минбалеев, Т. А. Полякова, М. Б. Добробаба [и др.]. — М. : Проспект, 2023. — 368 с.

[8] *Carlini Nicholas, Chang Liu, Úlfar Erlingsson, Jernej Kos, Dawn Song*. Op. cit.

4. *Минбалеев А. В.* Проблемы социальной эффективности и защиты прав человека при использовании искусственного интеллекта в рамках социального скоринга // Вестник Южно-Уральского государственного университета. — Серия : Право. — 2020. — Т. 20. — № 2. — С. 96—101.
5. *Полякова Т. А., Химченко А. И.* Правовые проблемы обеспечения информационной безопасности при использовании облачных технологий // Правовая информатика. — 2013. — № 2. — С. 12—16.
6. Право и иные регуляторы в развитии цифровых технологий / А. В. Минбалеев, А. В. Мартынов, Г. Г. Камалова [и др.]. — Саратов : Амирит, 2022. — 338 с.
7. Проблемы и вызовы цифрового общества: тенденции развития правового регулирования цифровых трансформаций : колл. монография : в 2 т. / Е. Н. Абанина, О. Ю. Авдевнина, Р. В. Амелин [и др.]. — Саратов : Саратовская государственная юридическая академия, 2020. — 204 с.
8. *Carlini Nicholas, Chang Liu, Úlfar Erlingsson, Jernej Kos, Dawn Song.* (2019) The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks // URL: https://doi.org/10.48550/arXiv.1802.08232.
9. *Drew Stevens.* (2018) Neural Networks And Advanced AI Contract Issues // URL: https://drewstevenslaw.com/neural-networks-and-advanced-ai-contract-issues/ (дата обращения: 10.12.2022).
10. The Impact of the General Data Protection Regulation (GDPR) on artificial intelligence / European Parliamentary Research Service Scientific Foresight Unit (STOA) PE 641.530 — June 2020.