

НЕЙРОСЕТИ: ПРИМЕНЕНИЕ, ВОПРОСЫ ЭТИКИ И ПРАВА

И. А. Филипова, irinafilipova@yandex.ru
Нижегородский государственный университет имени Н. И. Лобачевского,
г. Нижний Новгород, Россия

Аннотация. Использование искусственных нейронных сетей на практике в последние годы стало быстро расти, они все чаще применяются в бизнесе, в медицине и даже в государственном управлении. Такое распространение нейросетей как инструмента принятия решений меняет содержание ряда общественных отношений, тем самым ставя перед юристами новые задачи, касающиеся создания адекватного происходящим изменениям правового регулирования. На основе анализа теоретических положений и практических особенностей, относящихся к обучению искусственных нейронных сетей в целях дальнейшего их использования в качестве инструмента по поиску оптимальных решений, в статье сделаны выводы о ключевых вопросах этико-правового характера, без разрешения которых негативные эффекты от использования нейросетей проявятся значительно сильнее. Один из таких вопросов – необходимость «честности» обучения искусственной нейронной сети, чтобы она не воспроизводила общественные предрассудки и предпочтения разработчиков, дискриминируя отдельные социальные группы. Еще один вопрос касается прозрачности принятия решений нейросетью, отсутствие понимания логики в выборе решения не способствует установлению доверия со стороны людей. Не менее важен и вопрос о сохранении конфиденциальности данных: как защитить персональные данные, если они нужны нейросети для принятия решения? Результаты исследования позволяют четче сформулировать вопросы, требующие первоочередного внимания правоведов, чтобы вести поиск решений по ним юридическим сообществом. Ведь использование нейронных сетей практикующими юристами в своей профессиональной деятельности в рамках направления LegalTech параллельно с изучением теоретических моментов позволяет приблизиться к оптимальному решению поставленных вопросов.

Ключевые слова: машинное обучение, глубокое обучение, алгоритм, нейронная сеть, искусственный интеллект, генеративный искусственный интеллект, математическая модель, нейротехнологии, цифровые технологии, персональные данные, предиктивная аналитика, право, регулирование, транспарентность, этика, безопасность искусственного интеллекта.

Для цитирования: Филипова И. А. Нейросети: применение, вопросы этики и права // Вестник ЮУрГУ. Серия «Право». 2023. Т. 23. № 4. С. 76–81. DOI: 10.14529/law230411.

Original article
DOI: 10.14529/law230411

NEURAL NETWORKS: APPLICATION, ETHICS AND LAW ISSUES

I. A. Filipova, irinafilipova@yandex.ru
Nizhny Novgorod State University named after N. I. Lobachevsky,
Nizhny Novgorod, Russia

Abstract. The use of artificial neural networks in practice has grown rapidly in recent years; they are increasingly used in business, medicine, and even government administration. This proliferation of neural networks as a decision-making tool changes the content of a number of social relations, thereby posing new tasks for lawyers regarding the creation of legal regulation adequate to the changes taking place. Based on the analysis of theoretical provisions and practical features related to the training of artificial neural networks for the purpose of their further use as a tool for finding optimal solutions, the article draws conclusions about key issues of an ethical and legal nature, without the resolution of which the negative effects of the use of neural networks will manifest themselves significantly stronger.

One of these issues is the need for “fairness” in training an artificial neural network so that it does not reproduce social prejudices and preferences of developers, discriminating against certain social groups. Another question concerns the transparency of decision-making by a neural network; the lack of understanding of the logic in choosing a solution does not contribute to establishing trust on the part of people. No less important is the question of maintaining data confidentiality: how to protect personal data if the neural network needs it to make a decision? The results of the study allow us to more clearly formulate issues that require the priority attention of legal scholars in order to search for solutions to them by the legal community. After all, the use of neural networks by practicing lawyers in their professional activities within the LegalTech direction, in parallel with the study of theoretical aspects, allows them to get closer to the optimal solution to the issues raised.

Keywords: machine learning, deep learning, algorithm, neural network, artificial intelligence, mathematical model, neurotechnology, digital technologies, personal data, predictive analytics, law, regulation, transparency, ethics, artificial intelligence safety.

For citation: Filipova I. A. Neural networks: application, ethics and law issues. *Bulletin of the South Ural State University. Series “Law”*. 2023. vol. 23. no. 4, pp. 76–81. (in Russ.) DOI: 10.14529/law230411.

В последнее десятилетие нейросети получили широкое распространение на практике, все больше задач решается с их помощью. Одним из основных трендов 2023 года стал резкий всплеск внимания к генеративным нейросетевым моделям, таким как ChatGPT, Bard, DALL-E, Midjourney и т.д. Что такое нейросети, для чего они используются, каковы перспективы их применения на будущее и какие проблемы с точки зрения права это вызывает – вот те вопросы, которые будут рассмотрены в данной статье.

Искусственная нейронная сеть – это математическая модель нервной системы живого существа. Как нервная система человека – биологический прототип нейросети – реагирует на раздражители, так и нейросеть за счет входного слоя (сенсоров) получает сигналы из внешней среды, после чего в следующем слое (или слоях) анализирует их, а в выходном (реагирующем) слое выдает реакцию (решение). Внутренние скрытые слои нейросети состоят из ассоциативных нейронов, способных выявлять связи между поступившими через сенсорные нейроны данными.

«История нейросетей» началась задолго до текущего десятилетия. Впервые модель искусственного нейрона и нейронной сети была предложена в 1943 году нейрофизиологами У. Макаллоком и У. Питтсом [7, с. 115], ими было высказано предположение, что такая нейронная сеть способна к обучению и обобщению (может демонстрировать признаки интеллекта), именно поэтому данных исследователей называют одними из разработчиков первых систем искусственного интеллекта.

В 1948 году Н. Винер опубликовал свой труд «Кибернетика или управление и связь в животном и машине» [13], в котором представил сложные биологические процессы с помощью математических моделей, а уже в 1949 году Д. Хебб предложил первый алгоритм обучения нейросетей. В 1958–1960 гг. Ф. Розенблаттом была сконструирована нейросеть «Перцептрон» – кибернетическая модель мозга, реализованная в виде электронной машины (нейрокомпьютера), данная нейросеть имела один скрытый слой. «Перцептрон» мог решать задачи классификации и регрессии (предсказания), но в целом круг задач, которые он способен решить, был сильно ограниченным. Далее разработчики нейросетей стали их усложнять, в 70–80-е гг. XX века появились многослойные нейронные сети, способные находить решения для более трудных задач. Примером может служить задача кластеризации объектов, при которой, в отличие от задачи классификации, изначально неизвестно точное число групп, по которым должны быть распределены объекты.

К 80-м гг. XX века можно говорить о складывании машинного обучения как самостоятельного перспективного направления в создании слабого искусственного интеллекта. Машинное обучение – это процесс, в результате которого машины (компьютеры) способны демонстрировать поведение, которое не было в них явно запрограммировано ранее. Машинное обучение производится путем применения специальных алгоритмов самостоятельного поиска решений для разных задач (за счет выведения закономерностей на основе использования статистических дан-

ных). Соответственно нейросети уже к этому времени – хорошо известное специалистам направление, но возможности их обучения были скромными, так как отсутствовали достаточные вычислительные мощности (объем памяти, скорость обработки информации) и необходимые для эффективного обучения нейросетей большие датасеты, то есть наборы данных (изображений животных, транспорта, людей и т.д.).

Выход на новый уровень по вычислительным мощностям связан с распространением графических процессоров (видеокарт) к началу XXI века, крупные датасеты появились чуть позже, примерно с 2010 года. После этого словосочетание «глубокое обучение» (*Deep Learning*) как подвид машинного обучения (*Machine Learning*) становится известным за пределами сообществ программистов. Помимо увеличения вычислительных мощностей и появления крупных датасетов, развитию направления нейросетей способствовало и накопление опыта в построении различных разновидностей нейросетей или, как принято говорить, их архитектур. К настоящему времени основными типами нейросетей являются полносвязные, сверточные, рекуррентные сети и трансформеры, сочетающие элементы предыдущих трех типов.

Несмотря на то, что термин «нейронная сеть» может навести на мысль, что это направление относится к нейротехнологиям, данный вывод будет скорее ошибочным. Согласно отчету Международного биоэтического комитета ЮНЕСКО об этических вопросах нейротехнологий [9], опубликованному в декабре 2021 года, в группу нейротехнологий в широком смысле, действительно, можно включить любую технологию, которая оказывает фундаментальное влияние на то, как люди понимают мозг и мышление, тем не менее, в узком смысле нейротехнологии определяются как область устройств и процедур, позволяющих измерять и анализировать химические и электрические сигналы в нервной системе, взаимодействовать с нервной системой для изменения ее активности. Цель нейротехнологий – прямое влияние на мозг для исправления или улучшения его функций через воздействие на нервную систему, в то время как создание искусственных систем, способных выполнять максимально широкий круг задач на уровне, не уступающем человеческому, а в идеале значительно превосходя-

щем его – это цель другой группы технологий (искусственного интеллекта). С этой точки зрения к нейротехнологиям относятся лишь такие подгруппы технологий как нейропротезирование, нейроинтерфейсы, нейростимуляция и т.д. [2, с. 33].

Искусственная нейронная сеть – это математическая модель на основе нервной системы, способная выполнять задачи, в том числе творческого характера, что позволяет говорить об искусственном интеллекте (и о технологиях искусственного интеллекта, таких как компьютерное зрение или распознавание речи, что, собственно, и делают различные нейросети). Таким образом, направление нейросетей входит в группу технологий искусственного интеллекта. Впрочем употребление определения «сквозные» применительно к основным группам цифровых технологий позволяет говорить о наличии тесных взаимосвязей между этими группами. Кроме того, не стоит забывать, что разработкой первых искусственных нейронных сетей занимались нейрофизиологи. Кстати, развитие искусственного интеллекта с самого начала идет по двум основным направлениям – на основе двух подходов: восходящего (или биологического) и нисходящего (семиотического), основанного на формальной логике. Первый из подходов, что отражено и в названии, включает изучение нейрофизиологических и психологических механизмов мыслительной деятельности человека с целью воспроизвести эти механизмы техническими средствами так, чтобы «поведение» системы совпадало с поведением человека в данной ситуации.

Какова сегодня сфера применения искусственных нейронных сетей? Сверточные нейросети используются для распознавания фото-, видео-, иных изображений, они позволяют проводить идентификацию личности, создавать дипфейки и т.д. Рекуррентные сети используются для создания чат-ботов, генерации текстов и распознавания речи, например для перевода аудиофайлов в текстовые. Наиболее востребован и активно развивается такой тип нейросетей, как трансформеры, они могут выполнять наиболее широкий круг задач. В частности, популярное сейчас направление – развитие сверхглубоких мультимодальных нейросетевых моделей можно рассматривать как движение к сильному искусственному интеллекту, когда интеллектуальные способности системы повышаются за

счет наращивания числа параметров (превосходя их число в мозге человека), кроме того, система воспринимает модальности (зрение, распознавание речи, обоняние, движение и т.д.), включая «чувства», недоступные человеку, например, инфракрасное зрение, а используемые для обучения датасеты содержат немислимые для людей объемы данных. Такие генеративные модели некоторые исследователи уже предлагают признать «восходом» больших языковых моделей и зарождением общего (сильного) искусственного интеллекта [14, р. 8].

В условиях информационного общества, где экспоненциально возрастают объемы новых данных, причем, если раньше они создавались людьми, то чем дальше, тем больше генерируются различными техническими системами и устройствами, «работа» нейросетей становится необходимой для обработки и анализа данных, получаемых в процессе функционирования «умных» домов, «умных» машин, разнообразных «умных» гаджетов и целых «умных» заводов.

С учетом того, что для обучения искусственной нейронной сети критически важными являются вычислительные мощности и большие наборы данных (оба фактора наличествуют и демонстрируют многократный рост), обучение нейросетей будет совершенствоваться, а результаты обучения позволят использовать их и в тех областях, в которых сегодня это кажется малоприменимым. Если естественный интеллект (человек, не дополненный продуктами нейротехнологий) не в состоянии анализировать «сваливающиеся» на него гигантские и продолжающие увеличиваться потоки информации, то искусственный интеллект (в данном случае – нейросеть) делает это вместо человека.

Использование нейросетей на практике уже сегодня позволяет обозначить ряд вопросов, требующих осмысления с точки зрения этики и создания соответствующего правового регулирования во избежание негативных последствий растущего применения нейросетей для людей.

Первым из подобных вопросов может быть назван вопрос об этичности и «честности» обученных моделей нейросетей. Еще в 1967 году исследователь в области *Data Science* М. Конвей сформулировал «Закон Конвея», который гласит: когда организации проектируют системы, данные системы копи-

руют их структуру коммуникаций [4, с. 28]. Это означает, что создаваемые и обучаемые нейросети также отражают ценности людей, их разрабатывающих. Любые предпочтения и стереотипы, свойственные команде разработчиков, изначально закладываются в алгоритмы машинного обучения. Таким образом, возникает вопрос: как закон может обеспечить защиту от дискриминации, «заложенной» в нейросеть в процессе обучения? Можно привести пример, демонстрирующий проблему. В 2014 году компания *Amazon* заказала группе разработчиков создание программы по подбору сотрудников. Созданная программа стала отдавать приоритет мужчинам. Как позже выяснилось, обучение алгоритма основывалось на базе резюме всех сотрудников компании за предыдущее десятилетие, а доля мужчин в выборке оказалась гораздо больше. Этот случай не единичен, проблема предвзятости при выборе решения, свидетельствующая о дискриминации отдельных групп людей, уже проявилась на практике в результате использования нейросетей в разных областях – в трудоустройстве, кредитовании, страховании [10, с. 3].

Допустимо ли применять нейросети для выстраивания «социального рейтинга» (социальной оценки) граждан на основе их поведения? Технически нейросети могут отлично справиться с этой задачей. Судя по позиции европейского законодателя (*Artificial Intelligence Act*) [8], такое ранжирование должно быть в принципе запрещено, и наоборот, китайский законодатель, предусмотревший обязательный расчет «социального рейтинга», влияющего на доступ к общественным благам, считает иначе.

Следующий вопрос касается того, как соотносить требование о транспарентности (прозрачности) принятия решений искусственным интеллектом с тем, что нейросеть приходит к определенному решению, учитывая гораздо больше факторов, чем смог бы это сделать человек. С одной стороны, можно зафиксировать в законе требование: нейросеть может использоваться, только если логика ее решений объяснима. Так, Конституционный совет Франции в 2018 году постановил, что государственные органы вправе полагаться на алгоритмические процессы при принятии решений лишь в случае, когда они способны подробно и в понятной форме раскрыть способ, которым этот алгоритм пришел к такому ре-

шению (*Decision No. 2018-765 DC, June 12, 2018*) [5]. Представляется весьма затруднительным согласовать подобные требования с наличием скрытых слоев нейросети, выдающей решения по принципу «черного ящика». С другой стороны, если нейросеть способна дать верное решение по лечению смертельного заболевания, к примеру, и применение результатов этого решения на практике показывает ее «правоту», но неочевидно, почему сформулировано именно это решение, можно ли отказываться от ее использования на том основании, что не понятно, как нейросеть пришла к определенному выводу?

По мнению многих юристов, объяснимость как свойство искусственного интеллекта должна стать обязательным условием использования этих систем на практике, чтобы люди могли доверять нейросетям как инструменту принятия решений. В то же время методы глубокого обучения, используемые при обучении нейросетей, по сути, противоречат требованию алгоритмической прозрачности как запросу со стороны общества. Есть и другой аспект – отсутствие юридической прозрачности, то есть возможности доступа к коду программы, пока он защищен как интеллектуальная собственность или коммерческая тайна, в результате чего невозможен общественный контроль над работой алгоритма [1].

Еще один не менее важный вопрос связан с защитой персональных данных. При использовании нейросетей повышается угроза утраты конфиденциальности данных, ведь методы *Big Data* (накопление, управление и передача данных), *Data Mining* (структурирование информации и поиск скрытых связей) в совокупности с *Machine Learning* (аналитикой и прогнозированием), подвидом которого является глубокое обучение нейросетей, позволяют «вытащить» полезную информацию из данных гораздо в больших количествах, чем людям было доступно раньше. С учетом того, что чем больше данных, тем эффективнее будет производиться обучение нейросети, причем важно не только количество, но и качест-

во данных (то есть иногда необходимы именно персональные данные, например, при проведении исследований медицинского характера), к дискуссии о пределах защиты персональных данных будут возвращаться в ближайшие годы все чаще.

Помимо перечисленных выше вопросов, связанных с потенциальным регулированием использования нейросетей на практике (в каких сферах это допустимо, а в каких нет, при каких условиях и т.д.), есть вопросы и к применению нейросетей самим правом. Внимание юристов к использованию нейросетей и к проблемам, в связи с этим возникающим, стало проявляться задолго до широкого распространения нейросетевых моделей. Уже в 90-е гг. XX века зарубежные юристы задумались о применении нейросетей в юриспруденции [12, с. 129], в первую очередь, из-за возможностей предиктивной аналитики. Стоит отметить, что мнения разных правоведов о потенциале использования нейросетей правом нередко различались. Так, Д. Хантер из Университета Кембриджа, скорее, критически оценивал применение нейросетей [6, с. 130], в то время как М. Эйкенхед из Даремского университета (Великобритания) был более оптимистичен [3, с. 31]. Впрочем не стоит забывать, что «бум» развития нейросетей начался гораздо позднее, после 2010 года, когда обучение искусственных нейронных сетей вышло на новый уровень.

Современные правоведы отмечают резкое увеличение использования методов машинного обучения, в том числе *Deep Learning*, для решения юридических задач [11, с. 561], следовательно, интерес юристов к нейросетям в ближайшее время будет только расти, а это в свою очередь позволит более внимательно отнестись к изучению проблемных вопросов этического и правового характера, связанных с применением искусственных нейронных сетей. Данный вывод может быть подкреплен востребованностью нейросетевых моделей для осуществления предиктивной аналитики в быстро развивающейся сфере *LegalTech*.

Список источников

1. Дремлюга Р. И. Право и этика искусственного интеллекта: учебное пособие. Владивосток: Изд-во ДВФУ, 2021. 94 с.
2. Филипова И. А. Нейротехнологии в праве и правоприменении: прошлое, настоящее и будущее // Правоприменение. 2022. Т. 6. № 2. С. 32–49.

3. Aikenhead M. The Uses and Abuses of Neural Networks in Law // Santa Clara High Technology Law Journal, 1996. Vol. 12. Iss. 1. pp. 31–70.
4. Conway M.E. How do Committees Invent? // Datamation, 1968. Vol. 14. Iss. 5, pp. 28–31.
5. Decision No. 2018-765 DC of 12 June 2018. Law related to the protection of personal data. URL: <https://www.conseil-constitutionnel.fr/en/decision/2018/2018765DC.htm>.
6. Hunter D. Out of Their Minds: Legal Theory in Neural Networks // Artificial Intelligence and Law, 1999. Vol. 7, pp. 129–151.
7. McCulloch W.S., Pitts W. A Logical Calculus of the Ideas Immanent in Nervous Activity // Bulletin of Mathematical Biophysics, 1943. Vol. 5. pp. 115–133.
8. Proposal for a Regulation of The European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts COM/2021/206 final, 21.04.2021. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206>.
9. Report of the International Bioethics Committee of UNESCO (IBC) on the Ethical Issues of Neurotechnology, SHS/BIO/IBC-28/2021/3, December 15, 2021. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000378724> (дата обращения: 07.06.2023)
10. Rodrigues R. Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities // Journal of Responsible Technology, 2020. Vol. 4. Art. 100005.
11. Villata S., Araszkiwicz M., Ashley K. et al. Thirty Years of Artificial Intelligence and Law: The Third Decade // Artificial Intelligence Law, 2022. Vol. 30, pp. 561–591.
12. Warner D.R. The Role of Neural Networks in the Law Machine Development // Rutgers Computer & Technology Law Journal, 1990. Vol. 16, pp. 129–144.
13. Wiener N. Cybernetics: Or Control and Communication in the Animal and the Machine. New York: The Technology Press; Paris: Hermann et Cie, 1948. 194 p.
14. Yang Zh., Li L., Lin K., Wang J., Lin Ch.-Ch., Liu Z., Wang L. The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision) // Computer Science. Submitted on 29 Sep 2023. URL: <https://doi.org/10.48550/arXiv.2309.17421> (дата обращения: 02.10.2023).

References

1. Dremlyuga R. I. *Pravo i etika iskustvennogo intelekta* [Law and ethics of artificial intelligence]. Vladivostok, 2021, 94 p.
2. Filipova I. A. [Neurotechnologies in law and law enforcement: Past, present and future]. *Pravoprimerenie [Law Enforcement]*, 2022, Vol. 6, no. 2, pp. 32–49. (in Russ.)

Информация об авторе

Филипова Ирина Анатольевна, кандидат юридических наук, доцент кафедры трудового и экологического права юридического факультета, Нижегородский государственный университет имени Н. И. Лобачевского, г. Нижний Новгород, Россия.

Information about the author

Irina A. Filipova, Candidate of Sciences (Law), Associate Professor of the Department of Labor and Environmental Law, Faculty of Law, Nizhny Novgorod State University named after N. I. Lobachevsky, Nizhny Novgorod, Russia.

*Поступила в редакцию 25 августа 2023 г.
Received August 25, 2023.*