Толочко Ольга Николаевна

ПРАВОВЫЕ ПРОБЛЕМЫ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ

профессор кафедры государственного управления юридического факультета Белорусского государственного университета, доктор юридических наук, доцент, o.tolochko@mail.ru

Аннотация. Статья посвящена актуальным правовым проблемам использования технологий машинного обучения в юриспруденции. В настоящее время в мире разработаны и используются десятки таких технологий, однако многие аспекты их работы остаются не урегулированными. Автор обращает внимание на необходимость научно-правовых исследований в указанной области, в том числе контроля за качеством обучающих баз данных, присутствия человека на всех стадиях разработки алгоритмов, профилактики дискриминационных и иных нежелательных решений, принимаемых посредством систем искусственного интеллекта.

Ключевые слова: машинное обучение; право; предиктивный анализ; юриспруденция; искусственный интеллект.

Термин «машинное обучение» (machine learning) был введен в 1959 году американским исследователем искусственного интеллекта, сотрудником IBM и профессором Стэнфордского университета Артуром Сэмюэлом, который, помимо этого, разработал первую в мире самообучающуюся компьютерную программу игры в шашки.

Машинное обучение следует отличать от простого программирования. Машинное обучение представляет собой одну из ветвей искусственного интеллекта, базирующуюся на тренировке (обучении) алгоритмов на массивах определенных данных. Анализируя решения большого числа похожих задач, информационные системы начинают самостоятельно выявлять закономерности и предлагать собственные варианты их решений. Таким образом, машинное обучение построено на концепции, что компьютеры могут самообучаться и делать то, на что они не были запрограммированы изначально.

В настоящее время создано и применяется огромное количество продуктов, основанных на технологиях машинного обучения. Наиболее широко такие технологии используются в IT, торговле, транспорте, финансах, здравоохранении, образовании, сельском хозяйстве и промышленности. В последние годы стремительно развиваются также ML legal tech – технологии машинного обучения для юриспруденции.

Число таких технологий быстро растёт. Одним из первых считается американский проект eDiscovery, созданный для анализа и приоритезации юридических документов. Например, одна сторона судебного дела присылает другой стороне документы для ознакомления, – в американском судебном процессе таких документов может быть до миллиона. Машина анализирует их

и предлагает к внимательному прочтению лишь часть из них, которые представляются ей наиболее важными. На базе eDiscovery созданы и работают также технологии анализа и менеджмента контрактов, финансовых расследований и др. В нашей стране они не получили широкого распространения по ряду причин, одной из которых является отсутствие необходимости анализировать огромные массивы документов в судебном процессе – решения могут строиться на небольшом количестве доказательств, представленных истцом (в гражданском процессе) или обвинением (в уголовном).

В последние годы в США повсеместно применяются системы, оценивающие вероятность того, что обвиняемый или подсудимый вновь совершит правонарушение.

Разработанная IBM для нужд полиции ML-модель «Blue CRUSH» предсказывает места совершения будущих преступлений и показывает их на карте города. Эта технология, предназначенная для профилактики правонарушений, на основе статистики предоставляет полиции прогноз совершения потенциальных преступлений определенного вида в определенной местности и в определенный интервал времени. По имеющимся данным, применение её в Мемфисе (штат Теннеси) с 2006 по 2010 гг. позволило сократить количество серьезных преступлений более чем на 30% [1].

Несколько лет назад была разработана система Ross, обучающаяся на истории дел о банкротстве. Ross способна давать советы клиенту, обратившемуся за консультацией, причем диалог происходит в реальном времени, на естественном языке и без участия юриста. В основе системы лежат технологии IBM Watson. Похожего робота, консультирующего клиентов по закону о защите прав потребителей, недавно представила в России компания «Правовед».

Еще одним примером российской разработки на базе технологий IBM является проект Право.ру «Caselook», представляющий собой систему для поиска и анализа судебной практики с элементами машинного обучения. Эта система позволяет решать следующие задачи:

– разбивка споров на категории (аренда, заём, купля-продажа недвижимости и т.п.);

– поиск судебных споров по конкретному участнику;

– поиск в резолютивной части решения;

– поиск по результату рассмотрения (удовлетворено, частично удовлетворено, отказано в удовлетворении);

– определение вероятности исхода дела по конкретной категории споров в конкретном суде конкретным судьёй и др.

Наиболее далеко в применении ML legal tech продвинулся, по всей вероятности, Китай, где внедрена и используется разработанная Академией наук технология машинного обучения System of Systems (SoS), которая сканирует судебные дела, рекомендует судье законодательные акты, составляет процессуальные документы и даже исправляет то, что она считает ошибками в

приговоре. В соответствии с предписанием Верховного Народного Суда КНР каждый судья по каждому делу обязан консультироваться с системой. Если судья не согласен с рекомендациями, то он должен представить ей письменное объяснение для учёта и проверки. Система обучалась на 17 тысячах дел, рассмотренных в судах Шанхая с 2015 по 2020 годы. С 2022 г. она самостоятельно выносит обвинения, основываясь на тысяче характерных черт, которые сама выявляет в судебных документах по конкретным уголовным делам. По сообщениям из китайских источников, внедрение SoS в судебную систему позволило сократить среднюю нагрузку на судью более чем на 33% и сэкономить 1,7 млрд рабочих часов в течение 2 лет. Кроме того сообщается, что за 2019–2021 годы было сэкономлено более 45 млрд долларов, что эквивалентно половине всех гонораров адвокатов в Китае за 2021 год. Система может выдвигать обвинения по восьми наиболее распространённым в КНР преступлениям: хищение, мошенничество, мошенничество с кредитными картами, организация азартных игр, нарушение правил дорожного движения, нанесение телесных повреждений и создание препятствий представителям властей. Разработчики заявляют, что в скором времени искусственный интеллект сможет выносить обвинения и по более сложным с правовой точки зрения делам [2].

Вместе с тем, широкое использование ML legal tech сопряжено с рядом правовых и этических проблем. Разумеется, это большое подспорье в работе, экономия человеческих и материальных ресурсов, снижение коррупционных рисков и т.д. Вместе с тем, применение технологий машинного обучения в юриспруденции должно тщательнейшим образом регулироваться в силу целого ряда обстоятельств.

Прежде всего, важным требованием к качеству самой технологии является полнота, ценность и достоверность данных, на основе которых обучается система. Чем больше данных загружено в алгоритм, тем результативнее будет обучение. Однако многие данные, которые собираются различными органами и организациями, ограничены к доступу: персональные данные; медицинские данные; геолокация и информация, содержащая тайну связи; сведения, которые могут использоваться для идентификации субъектов и объектов, и др. В связи с этим закономерен вопрос: должен ли доступ к необходимым для обучения данным для решения задач в социально важных сферах, предоставляться на особых, упрощённых основаниях?

Проблема соблюдения баланса между требованиями о защите персональных данных и необходимостью их использования для обучения систем искусственного интеллекта является в настоящее время одной из основных концептуальных проблем мировой юридической науки. Решается она в разных странах по-разному. Например, в России принятая в 2019 г. Национальная стратегия развития искусственного интеллекта на период до 2030 года [3] предусматривает обеспечение благоприятных правовых условий для доступа к данным, преимущественно обезличенным, включая данные,

собираемые государственными органами и медицинскими организациями (п. 49).

Однако обезличивание данных не всегда гарантирует их защиту. Известно довольно большое число случаев, когда сопоставление обезличенных данных между собой позволяло раскрыть субъектов этих данных, в т.ч. медицинских [4]. Если использовать данные под условием согласия, то проблемой является то, что не всегда можно точно определить субъект, которому принадлежат права на данные, – лицо? торговое предприятие / клиника / юридическая фирма…? государственный орган? При этом крупные корпорации имеют гораздо больше возможностей по обработке данных. Они имеют доступ к большему объёму собираемых данных и поэтому находятся в более выигрышном положении, что может приводить к доминирующему положению небольшого количества крупных компаний, которые собирают информацию о своих пользователях и в своих действиях часто руководствуются только собственными интересами и внутренней политикой.

Второй ключевой проблемой является невозможность однозначно оценить, каким образом система пришла к тому или иному решению. Это связано с включением в процесс обучения новых данных и использованием машиной собственного опыта. Далеко не всегда лица, использующие технологию, понимают, каким образом она работает и на основании чего принимает то или иное решение. Иными словами, действия системы не являются достаточно прозрачными для человека (проблема «чёрного ящика») и, таким образом, теряется возможность четко определить причинно-следственную связь, важную для установления фактов и квалификации действий.

При этом, чем сложнее технология, тем меньше у пользователей понимания принципов и методов её работы. В качестве примера в литературе приводится случай, связанный с использованием алгоритма Northpointe, который предназначен для прогноза рецидивов преступлений при назначении наказания [5]. В своей работе Northpointe использовал различные факторы, в т.ч. возраст, пол, образ жизни семьи, употребление наркотиков, наличие арестов. При этом такой критерий как раса в качестве одного из факторов в алгоритм заложен не был. Тем не менее, когда выводы Northpointe были проанализированы журналистами, выяснилось, что в случаях, когда подсудимым являлся афроамериканец, алгоритм удваивал вероятность повторного совершения преступления, хотя такой критерий изначально не был заложен в его механизм.

Таким образом, искусственный интеллект, анализируя конкретные юридические дела, может устанавливать ложные корреляции и опираться на них при формулировании своих выводов. В качестве путей решения этой ситуации Европейская Комиссия в своём документе, посвящённом развитию технологий искусственного интеллекта («Белой книге»), предлагает сохранять информацию о базах данных, использованных для обучения и тестирования системы, а также фиксировать в соответствующей документации все процессы,

связанные с программированием, тестированием и обучением искусственного интеллекта. Для особых случаев предлагается, в том числе, установить обязанность сквозной архивации данных, на которых происходило обучение [6, с. 19–20]. При всех плюсах подобного подхода, тем не менее, это увеличивает ресурсоемкость и стоимость машинного обучения, что, в свою очередь, может снизить интерес бизнеса к развитию технологии.

Проблема ресурсоемкости прослеживается и в других сферах использования технологий машинного обучения. Например, в целях противодействия терроризму операторов связи и сервисов интернет-коммуникаций обязали хранить не только метаданные (сведения о коммуникациях – длительность, геолокация, инициатор, получатель и др.), но и содержание самих коммуникаций (текст, звук, фото, видео). Однако хранить все эти данные очень дорого. Не решен также вопрос о том, кто будет обрабатывать эти данные и принимать решения по ним – оператор связи? Орган расследования? Некий специально назначенный субъект? Остается и целый ряд других правовых и технических проблем: сжатие данных, анализ метаданных и др.

Третьей проблемой является феномен, обозначенный как «предвзятость искусственного интеллекта» (AI bias). Хрестоматийным примером считается обучение бота Microsoft на сообщениях пользователей Твиттера, в результате чего он начал генерировать антисемитские и сексистские сообщения. Иными словами, данные, предоставляемые алгоритмам, могут в силу сложившихся практик содержать элементы дискриминации, которые при этом будет крайне трудно обнаружить. Так, программа PredPol, разработанная для прогнозирования наиболее вероятных мест совершения преступлений полицией Окленда, была обучена на данных о том, в каких местах города полиция арестовывала нарушителей ранее. По результатам анализа вероятность нарушений в районах, где проживают афроамериканцы, оценивалась в 2 раза выше, чем в иных районах. При этом по данным национальной статистики места нарушений были гораздо равномернее распределены по городу. Т.е. сам факт, что данное подразделение полиции ранее уделяло большее внимание конкретному району, совершенно не означает потенциально большей криминогенности этого района.

Следовательно, для решения проблемы необходимо обеспечить предоставление для обучения максимально объективных данных. Машина не должна обучаться на любой доступной информации. Европейская Комиссия предлагает разработать некий стандарт для данных, на которых будут обучаться системы искусственного интеллекта. Кроме того, в процесс максимально должен быть вовлечен человек. Однако эти задачи легче ставить, нежели решать.

В целом следует с сожалением констатировать, что подходы к построению принципов машинного обучения, а также общая концепция использования технологий машинного обучения в юриспруденции пока что остается вне поля зрения отечественной правовой науки. Специальное правовое

регулирование использования технологий распознавания лиц, прогнозной аналитики и других технологий искусственного интеллекта также почти отсутствует, что, в свою очередь, тормозит развитие отечественных разработок. Такая ситуация может привести к использованию алгоритмов, созданных за рубежом, – например, китайских. Польза от этого неочевидна; хотелось бы всё-таки развития отечественного сектора ML legal tech с максимальным учётом потребностей страны и ее граждан.

Таким образом, технологии машинного обучения как достижение научно-технического прогресса могут не только приносить пользу, но и создавать угрозы нарушения ключевых прав человека: права на справедливое судебное разбирательство, свободу слова, доступность информации, тайну личной жизни, безопасность, равенство и недискриминацию. Риски могут быть снижены активным вовлечением в процесс их разработки и использования учёных-правоведов и, в конечном счете, качественным правовым регулированием.

Список использованных источников:

1. Журавлев, М. Право в эпоху развития цифровых технологий / М. Журавлев // Высшая школа экономики [Электронный ресурс]. – Режим доступа: https://www.hse.ru/mirror/pubs/share/216482270. – Дата доступа: 20.10.2022.

2. China’s AI-Enabled ‘Smart Courts’ To Recommend Laws & Draft Legal Docs; Judges To Take Consult AI Before Verdict // The Eurasian Times [Electronic resource]. – Mode of access: https://eurasiantimes.com/chinas-ai-enabled-smart-court-to-recommend-laws-judges/. – Date of access: 20.0.2022.

3. Национальная стратегия развития искусственного интеллекта на период до 2030 года: утв. указом Президента РФ от 10 октября 2019 года № 490 // Гарант [Электронный ресурс]. – Режим доступа: https://base.garant.ru/72838946/. – Дата доступа: 20.10.2022.

4. Наумов, В.Б. Правовые проблемы машинного обучения / В.Б. Наумов, Е.В. Тытюк // Образование и право. – 2020. – № 6. – С. 219–232.

5. A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear // The Washington Post [Electronic resource]. – Mode of access: https://washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas. – Date of access: 20.0.2022.

6. Белая книга «Об искусственном интеллекте - Европейский подход по совершенствованию и повышению доверия» (On Artificial Intelligence - A European approach to excellence and trust): документ Европейской Комиссии от 19 февраля 2020 г. // European Commission [Electronic resource]. – Mode of access: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. – Date of access: 20.0.2022.