

**ЕЛЕНА ЛАРИНА
ВЛАДИМИР ОВЧИНСКИЙ**

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ



ЭТИКА И ПРАВО



Елена Ларина, Владимир Обнинский
Искусственный интеллект. Этика и
право

Введение

Реализация **Стратегии развития технологий искусственного интеллекта в Российской Федерации до 2030 года** предполагает интеграцию всего мирового опыта на данном направлении.

В январе 2019 года на сайте ООН был представлен *первый Доклад Всемирной организации интеллектуальной собственности (ВОИС), посвященный развитию искусственного интеллекта (ИИ)*.

С середины 50-х годов прошлого века изобретатели и исследователи подали заявки на почти 340 тысяч изобретений на основе ИИ. Более половины всех изобретений, связанных с ИИ, были запатентованы после 2013 года.

Эксперты отмечают, что в последнее время увеличилось число патентных заявок на использование ИИ в робототехнике. В 2013 году их было 622, а в 2016-м – уже 2272. Ведущее место занимает транспортный сектор: в 2016 году на его долю приходилось 8764 патента. 6684 поданных заявки были связаны со сферой коммуникаций. В медицине, где ИИ активно применяется, например, в робото-хирургии, число заявок в 2016 году достигло 4112, что на 40 % больше, чем в 2013 году. В секторе персональных устройств, вычислительной техники и взаимодействия человека с компьютером число заявок в 2016 году составило 3977, что на 36 % больше, чем в 2013 году.

В числе лидирующих направлений использования ИИ – *сфера безопасности*. Применение ИИ *в юриспруденции* находится в самой нижней строке рейтинга. Тем не менее, этому направлению в США, странах Евросоюза, Японии, Южной Кореи и Китае уделяется не меньшее значение, чем применению ИИ в сферах бизнеса, промышленности, транспорте, медицине и IT.

Российские и зарубежные исследователи проблем использования ИИ в правосудии и правоохранительной деятельности нередко задаются вопросами об *этичности и правомерности* этого процесса. Высказываются предположения о том, что использование в правосудии ИИ таит опасность сделать человека, его права и свободы уязвимыми, а само правосудие бесчеловечным и формальным. Как отмечал на

одной на одном из научных обсуждений в 2018 году *начальник Управления конституционных основ уголовной юстиции Конституционного Суда России Олег Вагин*, «страшно представить, что решения в отношении человека, общества или человечества будут приниматься исходя из заданности информационных технологий, следуя совету «умного» искусственного интеллекта, полагаясь на который человек сам перестает думать, анализировать, прогнозировать свое будущее, а потому деградирует в качестве разумного существа, станет беспомощным и уязвимым, зависимым и от «машины», и от того лица, которое ею управляет, ее обслуживает, модернизирует, корректирует программное обеспечение, а в конечном счете, возможно и манипулирует поведением и решениями отдельных лиц, а через них интересами многих, тем самым превращая людей в марионеток, заложников ИТ технологий и средств их реализации».

Для того чтобы развеять эти сомнения в мировом сообществе делаются первые шаги. Так, в декабре 2018 года *Европейской Комиссией* одобрена **Европейская Этическая Хартия использования искусственного интеллекта в судебной и правоохранительной системах.**

Кроме того, в феврале 2019 года *Центр Европейских Политических исследований при Евросоюзе* опубликовала доклад **об этических, правовых и политических принципах регулирования развития и применения искусственного интеллекта относительно любых направлений деятельности.**

В США *Партнерство по искусственному интеллекту (PAI)*, в которое входит более 80 корпоративных разработчиков и пользователей ИИ, в начале 2019 года опубликовало **отчет об алгоритмических инструментах оценки рисков в системе уголовного правосудия США.**

25 мая 2019 года *Организация экономического развития и сотрудничества (ОЭСР)* поддержала **Принципы ответственного управления надежным ИИ.**

В России противники применения ИИ в судебной системе часто апеллируют к тому, что у нас не прецедентное право, поэтому нельзя основываться в судебных решениях на аналоги уголовных дел. При этом закрываются глаза на то, что судьи всегда основываются на своем

прошлом опыте по делам об аналогичных составах преступления, опыте региональных судей, решениях Конституционного Суда РФ и постановлениях Президиума Верховного Суда РФ, которые по сути своей – разновидность прецедентного права.

Но технологический «ветер перемен» коснулся и российских судов. Выступая 12 февраля 2018 года на совещании – семинаре судей судов общей юрисдикции и арбитражных судов России *председатель Правительства РФ Дмитрий Медведев* подчеркнул, что в стране продолжается работа по **цифровизации судебной системы**. В банке автоматизированной системы ГАС «Правосудие» уже более 80 млн. дел. В 2018 году суды общей юрисдикции приняли более 500 тыс. заявлений в электронном виде, и более 250 тыс. заявлений поступило в арбитражные суды. Активно ведется работа по размещению в интернете решений судов. В базе данных Верховного суда уже более 600 тыс. дел.

Это только первые шаги, которые должны вести к использованию российскими судами ИИ. Многие задачи, которые выполняют юристы и адвокаты, очень трудоемкие, утомительные и дорогостоящие. При этом существует риск ошибок и неточностей в работе специалистов.

Вопрос о делегировании таких задач ИИ очень актуален. Сервисы по автоматизации поиска юридической информации и анализа документов могут сэкономить до 80 % времени юристов и до 90 % стоимости их консультаций.

Еще один драйвер внедрения ИИ в юридическую сферу – большие данные (bigdata). Это огромный массив накопленной человечеством информации: судебные решения, иски, договоры, консультации. Благодаря росту мощностей современных компьютеров, совершенствованию алгоритмов и математических моделей, появляется возможность на базе этих больших данных создавать принципиально новые продукты.

Создание и использование ИИ и БД не является попыткой заменить человека (судью, прокурора, следователя, полицейского и адвоката) в процессе принятия решений по всем направлениям судопроизводства.

Главное цель это максимальная объективность на основании БД и сокращение сроков на принятие решений или получения достоверной информации с помощью ИИ. На уже упомянутом совещании –

семинаре судей Дмитрий Медведев особый акцент сделал на нагрузке судей. Если в 2017 году 33 тыс. судей страны рассмотрели 28 млн. дел, то в 2018 году уже более 31 млн. дел. Причем прогнозируется постоянный рост числа дел для рассмотрения.

Внедрение ИИ и БД – это уже вынужденная мера для обеспечения качественной и эффективной судебной защиты прав наших граждан, обеспечения доступа граждан к правосудию и их безопасности.

Глава I
Использование ИИ в судебных и
правоохранительных системах

§ 1. Общий обзор

Судебная и правоохранительная сферы относятся к числу областей с наиболее продвинутым использованием искусственного интеллекта (ИИ). В 2018 г. в *Соединенных Штатах* «роботы-адвокаты» уже оказали услуги более чем на 1 млрд, долларов. Они помогают клиентам *готовить первичную документацию и различного рода юридические документы* для представления в государственные органы.

Юридические стартапы, специализирующиеся на разработке новых юридических услуг, продали юристам, юридическим службам и страховщикам услуг на 2,3 млрд, долларов. Появились также частные компании, оказывающие юридическим и физическим лицам экзотическую услугу, а именно, *прогнозирование решений судов присяжных по различным делам*. Данная услуга при всей своей спорности лежит в рамках законодательства штатов и поэтому получает все большее развитие.

В ряде государств, прежде всего, в Нидерландах, Германии, Франции, Австрии, Польше уже разработан чрезвычайно продвинутый подход к использованию практических приложений ИИ, прежде всего, *в системе электронного документооборота*, в судах и правоохранительных органах. В некоторых государствах, а именно в *Нидерландах, Франции, Австрии* системы ИИ, использующие глубокие нейронные сети на основе машинного обучения, взяты на вооружение правоохранительными органами крупных городов и используются для *предиктивной аналитики*, а также цифровых расследований.

В Великобритании главное внимание уделено использованию ИИ *для поиска информации в юридических архивах и хранилищах*, а также повышения эффективности электронного документооборота судебной и правоохранительной сфер.

С 2019 года китайские суды уже используют в судебных процессах и исполнении судебных решений технологии искусственного интеллекта для улучшения услуг и развития интеллектуальных судов. Об этом проинформировали в Верховном народном суде КНР в марте 2019 года.

По словам директора информационного центра Верховного народного суда Сюй Цзяньфэна, народные суды в стране будут использовать приложения для электронной подачи заявлений, продвигать мобильные электронные судебные процессы, а также создавать мобильные микросуды, позволяющие интеллектуальным приложениям охватывать все аспекты судебных процессов.

Что касается создания «умных» судов, народные суды будут сосредоточены на точном исполнении судебных решений, создании голограмм людей, находящихся под следствием, а также формировании связи с системой социального кредитования.

Одновременно с помощью технологии распознавания лиц народные суды Китая будут совершенствовать комплексные услуги и создавать базу данных для предоставления услуг по проверке личности и совместному использованию информации о гражданстве, адвокатах, предприятиях и судьях.

Китайская онлайн-судебная система, одна из публичных судебных платформ страны, обнародовала более 60 миллионов судебных документов. На платформе зарегистрировано к началу 2019 года более 20 миллиардов посещений.

В Аргентине прокуратура округа Буэнос-Айрес в 2018 году подвела промежуточные итоги эксперимента по использованию ИИ по ряду категорий гражданских и административных дел. Оказалось, что местными судьями были утверждены 100 % решений, принятых за 2018 год с помощью ИИ.

Приложение *Prometea*, как сообщало издательство *Bloomberg*, было создано 29-летним программистом Игнасио Раффой в партнерстве с офисом окружного прокурора аргентинской столицы. Раффа обучил двуязычную программу (она распознает английский и испанский языки), используя цифровую библиотеку документов: *Prometea* проанализировала порядка 300 тыс. отсканированных судебных решений с 2016-го по 2017 год, в том числе 2000 постановлений (в Аргентине окружные прокуроры составляют решения, а председательствующие судьи по делам либо отклоняют их и пишут свои собственные, либо просто одобряют их).

Теперь, как только новое дело попадает в прокурорскую систему, *Prometea* сопоставляет фактуру с наиболее релевантными решениями в своей базе данных – и это позволяет программе примерно за 10 секунд

угадать, как суд отреагирует на ситуацию. Речь пока идет об относительно простых случаях, например, спорах о выдаче лицензии на перевозки такси, отказах в выплате компенсации учителям, купившим школьные принадлежности за свой счет, и т. п. В результате применения приложения прокурорские работники оказались освобождены от больших массивов рутинной деятельности. По словам главы офиса, *15 его юристов сейчас за шесть недель справляются с таким объемом работы, на который прежде им требовалось около шести месяцев*, ошибки при просмотре сгенерированных компьютером файлов встречаются весьма редко. Окружной прокурор, в частности, поведал, что его сотрудники, однажды подававшие рядовую жалобу на вождение в нетрезвом виде, должны были внести в нее «39 деталей 111 раз». Теперь, формируя подобный документ в новой программе, они вносят все данные только один раз.

Высоко оценили приложение и судьи, которыми к настоящему времени одобрено 33 из 33 предложенных им решений (сейчас оно используется еще в 84 других рассматриваемых делах). По словам проводившего презентацию Prometea профессора Оксфордского университета *Эсекьяля Гонсалеса*, программу нельзя считать полноценной заменой интеллекту судьи-человека, однако она очень помогает служителям Фемиды, которые буквально «похоронены» в бумажной работе.

По данным Европейской Комиссии на декабрь 2018 года, *судьи в государствах-членах ЕС широко и повсеместно применяют традиционный софт и сервисы, связанные с документооборотом, и вообще не используют прогнозное программное обеспечение, базирующееся на ИИ*, в своей профессиональной деятельности.

По оценкам Европола, правоохранительные органы Европы, возможно, за исключением *Нидерландов*, отстают как минимум на 10–12 лет от Соединенных Штатов, Великобритании и Китая в использовании и прогнозной или предиктивной аналитики в практической правоохранительной деятельности, включая полицейских непосредственно на земле. В целом, использование ИИ в судебной и правоохранительной системах государств-членов ЕС крайне незначительно и относится скорее к перспективным, чем к текущим задачам.

Согласно материалам государств-членов ЕС, на начало 2019 года на государственном уровне *ни в одной из стран не реализуется государственная программа цифровизации судебной системы и национальных правоохранительных органов.* Все успехи европейского ИИ, за исключением *Франции* и выходящей из Европейского Союза Великобритании, связаны исключительно с *разработками частного сектора, нацеленными на европейские адвокатские и в целом юридические фирмы.*

По мнению экспертов ЕС, отставание Европы от лидеров в области ИИ имеет не только очевидные минусы, но и некоторые плюсы. В частности, при разработке и внедрении ИИ в судебную сферу можно использовать зарубежный опыт феномена «предсказательной юстиции».

В настоящее время, особенно в *Соединенных Штатах, Великобритании, отчасти в Китае, Южной Корее и Японии,* все шире используется *система предсказательного судопроизводства* и предсказательных полицейских действий. Использование предсказательной аналитики в судебной и правоохранительной сферах сталкивается в этих государствах со значительной критикой. Соответственно производители систем ИИ публикуют достаточно большой объем данных, программ, статистики, характеризующих работу их систем. В рамках ЕС обсуждается вопрос создания межстрановой комиссии, которая могла бы на междисциплинарной основе обобщить результаты дискуссии и подготовить базис для создания европейских систем ИИ.

Столь быстрое развитие предиктивной аналитики связано не только с практическими нуждами, но и объясняется соревнованием США и Китая в области ИИ, используемого для поведенческого управления и контроля. Прогноз – важнейшая фаза поведенческого контроля, а соответственно в значительной степени предопределяет его эффективность.

Анализ американской и азиатской практики применения предиктивной судебной аналитики позволяет увидеть еще одну опасность. Не секрет, что все большее распространение в мире, в том числе в Европе, приобретает практика создания частных инвестиционных пулов, которые финансируют адвокатов в судебных процессах, где потерпевший выступает против корпораций. Особенно

это связано с табачной, алкогольной, медицинскими отраслями, производителями товаров для населения. Если ранее подобного рода пулы формировались исключительно на основе экспертного анализа, то теперь они покупают услуги предиктивной судебной аналитики, рассчитывающей шансы на то или иное завершение судебного процесса.

Эксперты ЕС полагают, что в перспективе это может привести к усилению зависимости судей, присяжных и других участников судебного процесса от заранее сделанных прогнозов относительно итогов процесса, которые базируются не на каких-то предпочтениях и личных оценках, а исключительно на объективных статистических расчетах.

В этой связи в ЕС поставлена задача проанализировать, насколько целесообразно с точки зрения соблюдения прав человека разрешить или нет осуществлять подобную деятельность на территории ЕС.

Все более жестко разворачивающаяся гонка между США и Китаем в области ИИ и больших данных, которые, по сути, на наших глазах из различных областей сливаются в одну, подкрепленную интернетом вещь, позволяет разглядеть еще одну опасность. Уже сегодня очевидно, что тот, кто обладает **АЮТ (ИИ + большие данные + интернет вещей)** на порядок сильнее и могущественнее того, у кого этого нет. Использование полицией и судебной властью, включая прокуратуру, АЮТ делает одну из сторон судебного разбирательства на порядок более мощной и вооруженной, чем другая. Де факто, по мнению экспертов ЕС, это с большой вероятностью подрывает право граждан на независимый и беспристрастный суд, а также равенство сторон в судебном разбирательстве.

Эксперты ЕС полагают, что *бесконтрольное использование ИИ и АЮТ в судебной и правоохранительной областях без преувеличения в короткие сроки может уничтожить важнейший европейский принцип соблюдения прав человека и состязательный характер судебного процесса.* Использование АЮТ одной стороной процесса, по их мнению, сродни превращению состязания в преодоление местности между людьми в соперничестве человека и автомобиля.

Проще всего было бы запретить использование АЮТ в судебной сфере, тем более что пока в государствах ЕС ИИ применяется только в сфере повышения эффективности судебного документооборота.

Однако практика использования новых технологий в мире показывает, что *подобного рода запреты не работают*, и любая технология получает соответствующее распространение. К тому же *решения ИИ уже сегодня используются криминалом и организованной преступностью*^[1]. Соответственно, если правоохранительные органы и судебные системы на основе законодательных решений будут лишены АЮТ, тем самым они будут поставлены в заведомо проигрышную позицию по отношению к криминальным структурам.

Вопрос применения ИИ в судебной и правоохранительных сферах – это не вопрос технологий и информационной политики, а политическая и отчасти социально-психологическая проблема. ИИ – это вычислительная система, делающая выбор на основе статистической информации. Как правило, главная проблема, которая решается в подобных системах, это – обеспечение отражения в статистической информации фактического положения дел. Наиболее эффективный ИИ будет принимать или рекомендовать решения на основе реальных связей, зависимостей и отношений, которые выявлены на основе анализа реального положения дел.

Однако, хорошо известно, что политика базируется не на реальном положении дел, а на компромиссе между интересами, возможностями и силой, которой обладают различные группы, слои, страты общества и паттерны давления. Политическая система компромиссов имеет свое отражение и в судебной системе. Любое судебное решение – это ситуативное решение, когда закон применяется к определенной конкретной ситуации. Особенно это очевидно в британском прецедентном праве. Однако в более неявном виде ситуационный принцип реализуется и в так называемом континентальном праве.

Главная проблема использования ИИ в судебной практике это то, что ИИ базируется не на ситуационной логике, а на вычислительных процедурах. ИИ не умеет работать с контекстами. Соответственно, *в каждом государстве политические силы должны прийти к компромиссу и установить, в каких сферах и в каких аспектах судебной системы и правоохранительной деятельности можно использовать ИИ, а в каких использование этого инструмента опасно для политического консенсуса.*

Вопрос об использовании ИИ в судебных системах был впервые рассмотрен на уровне Европейской Комиссии в апреле 2018 г.

Результаты опроса свидетельствуют о том, что министерства юстиции государств-членов ЕС сообщили, что широко пользуются как на федеральном, так и на низовом уровне инструментами ИИ. Однако при обработке результатов опроса удалось установить, что практически во всех случаях под инструментами ИИ понимались либо корпоративные информационно-аналитические системы, т. е. фактически хранилища документации, оснащенные визуализаторами и поисковиками, либо стандартные статистические пакеты, обрабатывающие стандартные цифровые данные. Ни первые, ни вторые программные комплексы не являются ИИ, а относятся к предыдущей стадии интеллектуального софта – datamining.

Приведенные ниже сведения основаны не на результатах опроса, а на исследовании в частном секторе, проведенном экспертами ЕС. Прежде всего, следует сразу уточнить, что выбраны приложения, где уже сегодня участвует хоть в какой-то степени ИИ. Под ИИ в данном случае понимаются платформы, обеспечивающие анализ и прогнозирование в судебной и правоохранительной сферах на основе обработки больших данных с использованием машинного обучения, нейронных сетей и методов распознавания образов с выводом результатов в визуальном или любом ином понятном для конечных пользователей виде. Выделяется несколько направлений практического использования ИИ в судебной и правоохранительной сферах. Конкретно речь идет о:

- продвинутых семантических корпоративных поисковых системах;
- системах поиска и анализа ситуаций, содержащихся в юридических документах;
- системах поддержки разработки юридических документов на основе автоматического генерирования шаблонов;
- системах предиктивной судебной аналитики, ориентированных на адвокатские компании;
- юридических чатботах, осуществляющих информирование сторон в ходе первоначальных контактов по тем или иным делам на естественных языках.

Первым государством ЕС, которое заявило, что приступило к изучению возможностей машинного обучения для повышения качества судебных решений, стала *Латвия*. Основная цель на

сегодняшний день – это запуск не позднее конца 2020 г. охватывающей все судебные решения страны системы прогнозной аналитики для составления предварительной оценки людских и материальных ресурсов, которые могут быть потрачены на ведение того или иного процесса, начиная от принятия решения о начале судопроизводства и вплоть до вынесения окончательного вердикта. Фактически в Латвии собираются запустить *систему предиктивной судебной аналитики*, рассчитанную не на частные компании, выступающие на стороне, как правило, защиты, а в интересах государства – в основном стороны обвинения.

Наиболее широко система ИИ на сегодняшний день в государствах ЕС применяются в двух странах – *Нидерландах и Франции*. В остальных государствах ЕС отсутствуют национальные решения в области использования ИИ в судебной и правоохранительных сферах. В Нидерландах на сегодняшний день ИИ используется для предиктивной судебной и правоохранительной аналитики, для *углубленного datamining текстовых документов* и для обработки документов, включая генерирование шаблонов, в рамках электронного документооборота.

Французы создали *внутренний государственный семантический поисковик*, используемый судебными работниками и правоохранителями для работы с государственными базами данных и архивами судебных решений. Также во Франции юридические компании широко используют *системы превентивной судебной аналитики* и чатботы, поддерживающие наряду с французским, английский и арабский языки. В остальных странах национальных систем ИИ для судопроизводства выявить не удалось.

§ 2. Европейская политика в области больших данных, относящихся к судебной и правовой системам

Доступность данных является важнейшим условием развития ИИ, позволяющим ему выполнять определенные задачи, ранее выполняемые людьми. Чем больше доступных данных, тем больше ИИ может совершенствовать свои модели, улучшать их способность к распознаванию тех или иных ситуаций и прогнозированию. Поэтому подход раскрытия больших данных, относящихся к судебным решениям, является обязательным условием использования ИИ в судебной системе.

По мнению экспертов ЕС, растущее отставание большинства государств ЕС от мировых лидеров в области ИИ связано не со слабостью европейских компаний, специализирующихся в области интеллектуального софта и не с отсталостью европейской математики, а с гораздо более жесткими и запретительными по сравнению с Соединенными Штатами, Китаем и рядом других стран законами ЕС и отдельных стран в части использования, хранения и извлечения любых видов данных, имеющих отношение к гражданам Европейского Союза.

Компьютерные данные являются «нефтью» XXI века. Их использование производит новое богатство. Глобальные успехи в цифровой индустрии за последние десятилетия подтвердили огромный потенциал этой сферы деятельности, как локомотива четвертой производственной революции. Эксперты ЕС признают, что европейские страны в большинстве своем проигрывают цифровую гонку, уступая не только Соединенным Штатам и Китаю, но и Японии, Южной Корее, вышедшей из состава ЕС Великобритании и т. п.

В большинстве государств ЕС традиционно значительная часть правоохранительных данных и некоторые судебные данные закрыты и требуют для доступа определенного статуса. Совершенно другая ситуация в Соединенных Штатах и странах Британского Содружества наций, где *практически все судебные и большинство правоохранительных данных стали открытыми данными, а соответственно оказались доступными для обработки ИИ.*

Европа фактически проспала революцию, связанную с переводом больших данных в формат общедоступных машиночитаемых данных. Соответствующий стандарт был разработан неправительственной организацией *«Партнерство для открытого правительства»*. В него входят цифровые гиганты и более 70 государств-членов, включая США, Великобританию, Россию, Японию, многие государства-члены ЕС.

Целью этой организации является повышение прозрачности общественной деятельности, последовательный перевод закрытых государственных данных в открытые, доступные аналитике и прогностике не только со стороны государства, но и общества. Следует констатировать, что после мощного старта в начале десятых годов, во второй половине текущего десятилетия подавляющая часть государств ЕС, прекратила активное участие в этой организации и приостановила процесс перевода больших данных в открытые машиночитаемые данные.

В этой связи в мае 2018 г. Европейская Комиссия приняла решение о развертывании *европейской инициативы «За открытые машиночитаемые данные»*. В рамках ЕС ответственность за общеевропейскую программу взяла на себя Франция. Именно эта страна является на сегодняшний день одним из мировых лидеров в области перевода больших данных в открытые, машиночитаемые данные. Во Франции существует и эффективно действует специальный портал, на котором размещается вся сертифицированная публичная информация, содержащая не только законодательные и нормативные тексты, но также прецедентное право, информацию о назначениях на публичные должности и огромные массивы статистической информации, которые за государственный счет переводятся в формат машиночитаемых данных.

К настоящему времени в значительной степени благодаря усилиям Франции, в государствах-членах ЕС проходит унификация требований к данным. Открытые машиночитаемые данные включают только те данные, которые могут быть автоматическим образом извлечены из портала в интернете и без потерь и искажений храниться в структурированных компьютерных базах данных.

Открытые данные, объединенные с другими структурированными и неструктурированными данными, составляют то, что называется

большими данными. Консультативный Комитет Конвенции 108 Совета Европы определяет большие данные как «растущую технологическую способность собирать, обрабатывать и извлекать новые, в том числе прогнозные, знания из большого объема быстроизменяющихся и разнообразных данных». Как показывает определение открытые и большие данные не следует путать со средствами их обработки. ИИ – это ни в коем случае не сами по себе большие данные, а комплексы или платформы их обработки на основе использования сложных статистических алгоритмов, машинного обучения и нейронных сетей. По сути, в большинстве случаев *ИИ сегодня определяется не через функцию, либо особое содержание, а как своего рода зонтичное определение совокупности конкретных IT технологий, типа нейронных сетей, распознавания образов, факторного анализа, корреляционно-регрессионного метода и т. п.*

В настоящее время европейскими лидерами по открытым данным в судебной сфере являются Франция и Нидерланды. Они еще в 2016 г. приняли законодательство, делающее обязательным публикацию открытых данных не только о решениях судов различной юрисдикции, но и судебных дел по закончившимся процессам. В настоящее время во Франции и Нидерландах в обязательном порядке, за исключением особенных случаев, определенных законом, публикуются не только судебные решения, но и сопутствующие документы. Это позволяет развивать судебный ИИ на основе анализа открытых данных.

Однако в 2017–2018 гг. во Франции произошел ряд инцидентов, которые поставили под сомнение принцип открытой публикации материалов судебных дел. В настоящее время этот вопрос рассматривается Европейским судом по правам человека. Лишь после вынесения его решения будет окончательно ясно, есть ли будущее у публикации всех материалов судебных дел, как открытых данных, или в перспективе публиковаться будут лишь судебные решения.

Эксперты ЕС делают вывод, что, несмотря на некоторые многообещающие разработки, *создание полностью эффективного автоматического механизма распознавания ситуации на основе анализа судебных решений и дел несет большие риски для соблюдения конфиденциальности и личных свобод граждан.*

В 2017–2018 гг. в соответствии с решениями Совета Европы был проведен масштабный опрос среди сотрудников судов; какое

направление они считают наиболее многообещающим с точки зрения использования возможностей ИИ для анализа машиночитаемых данных. На выбор предлагалось четыре варианта:

- использование ИИ для извлечения знаний из хранящихся в электронном виде решений судов и судебной документации;
- использование ИИ для прогнозирования линии защиты в ходе судебных процессов;
- использование ИИ для поиска прецедентных решений судов или близких к рассматриваемым в том или ином процессе случаев в прошлом. Фактически речь идет о распознавании ситуации;
- использование ИИ как служебного инструмента для организации первичного общения, автоматизации и генерации шаблонов юридических документов, улучшения качества переводов с редких иностранных языков в тех случаях, когда в процессах участвуют их носители.

Результаты обследования более чем 1700 судебных работников из 12 государств ЕС следующие: более 50 % в качестве первоочередной, наиболее интересующей функции выделили относительно простые служебные функции ИИ, связанные с чат-ботами, визуализацией, переводами и т. п. Более 20 % выразили заинтересованность в наличии инструмента, позволяющего распознавать аналоги ситуаций, рассматриваемых в том или ином судебном процессе в прошлом, и созданию подборок документов по данным ситуациям. Еще 20 % опрошенных сказали, что им ИИ не интересен и не нужен. Оставшиеся 10 % примерно пополам разделились между поисковой и прогностической функциями ИИ, которые, например, в Соединенных Штатах являются главными функциями ИИ в судебной и правоохранительной сферах.

Учитывая, что служебные функции ИИ, по сути, выполняются традиционным интеллектуальным софтом, можно утверждать, что более 80 % судебных работников не видят необходимости в использовании ИИ в своей работе, а соответственно не видят необходимости в повышении уровня своих знаний в области информационно-коммуникационных технологий за пределами умения работать с поисковиками и традиционным аналитическим софтом.

Для достижения справедливого баланса между эффективностью правоохранительных и судебных систем – с одной стороны, и приватности – с другой, необходимо провести тщательнейший анализ целесообразности публикации в рамках судебных решений имен и фамилий свидетелей, их адресов, показаний и т. п. Ключевым вопросом, который предстоит разрешить системам ИИ для судебной отрасли, который неизбежно появится, является вопрос персональных данных.

К персональным данным в рамках европейского законодательства относятся привязанные к имени, фамилии номера телефонов, адреса места жительства и электронной почты, дата рождения, имена детей, имена и фамилии родственников, места работы, проведения отдыха и т. п. В большинстве стран запрещено не только бесконтрольное использование персональных данных, но и хранение личных данных на частных серверах, в базах данных без письменного согласия человека. Проблема заключается в том, что в решениях судов, а тем более в судебной документации все эти данные обязательно содержатся. Соответственно, используя самые простые методы анализа, из судебной документации могут быть извлечены огромные массивы персональных данных. Кроме того, некоторые персоналии могут быть идентифицированы по косвенным данным и анализу перекрестных ссылок.

Наряду с персональными данными, существуют особо чувствительные данные, которые в соответствии с европейскими конвенциональными требованиями не могут быть раскрыты ни перед кем, включая государственные органы, без специального решения суда. Речь идет *об этническом или расовом происхождении, политических взглядах, членстве в профсоюзах, религиозных и иных убеждениях, любой другой информации, касающейся физического, психического здоровья, половой жизни и личных увлечений, которые по законодательству считаются нераскрываемыми интимными подробностями.*

Проблема заключается в том, что эти нераскрываемые данные подчас содержатся в материалах судебных заседаний, составляют их важную часть и не защищены в текстах специальным образом. В соответствии с этим подавляющее большинство судей, прокуроров в государствах-членах ЕС выступают за законодательный запрет на

уровне Европейской Комиссии, переводить судебную документацию, за исключением общедоступных судебных решений, в разряд открытых данных, и более того, требуют запретить пользоваться ими всем, за исключением специально поименованных категорий участников состязательных судебных процессов. Выбирая между ИИ и личной приватностью, по мнению судебных работников, необходимо выбрать приватность.

При этом более 80 % опрошенных судебных работников ЕС полагают необходимым законодательно ограничить или строго кодифицировать условия применения ИИ в правоохранительной работе с тем, чтобы избежать вторжения полиции в частную жизнь законопослушных граждан.

§ 3. Профессиональные судебные проблемы и ИИ

Ни для кого не секрет, что значительную роль в вынесении того или иного судебного решения даже в рамках англосаксонского права играет *личность судьи*. Адвокатские фирмы издавна занимались *прогнозированием исхода дела в зависимости от состава суда*. Эта деятельность была важнейшим элементом адвокатской работы задолго до появления ИИ. До появления ИИ адвокатские фирмы в основном анализировали прошлые дела того или иного судьи, и на основании этой информации пытались делать прогноз относительно решения текущего процесса. Также адвокатские фирмы в последние 10–15 лет разработали сложнейшие алгоритмы, сочетающие использование ИИ и человеческого экспертного анализа, для подбора присяжных в Суды присяжных.

Вполне очевидно, что *появление ИИ позволяет резко повысить уровень аналитики и прогностики для адвокатов*. В этом случае в оборот будут вовлекаться не только прежние решения судьи, но и исследоваться их корреляции со всеми мыслимыми и немыслимыми обстоятельствами, имеющими отношения к судьям, о которых имеется информация. Например, в конце 2018 г. в Соединенных Штатах Лаборатория ИИ Северо-восточного университета опубликовала результаты анализа ИИ факторов, влияющих на решения судей об условно-досрочном освобождении преступников, получивших относительно недолгие – до шести лет – сроки. К удивлению судебных работников, правоохранителей и самих исследователей, самым сильным фактором, обуславливающим отклонение в положительную сторону решений об условно-досрочном освобождении, является *фактор точного времени рассмотрения дела*. Выяснилось, что в послеобеденные полтора часа число положительных судебных решений оказалось на 19 % больше, чем в среднем по дню. Ни один другой фактор не влиял на американских судей, как положительные эмоции от хорошего обеда. Вторым по значимости фактором является время и день недели. После обеда и до вечера в пятницу выносилось оправдательных приговоров на 12 % больше, чем в среднем.

Первоначальная гипотеза исследователей состояла в том, что таковыми факторами станут молодой возраст подсудимых или наличие у женщин-подсудимых детей. Однако сколько-нибудь заметного влияния на решения эти факторы не оказали. *Хороший обед и близость уикенда перевешивали у американских судей все остальные факторы.*

Понятно, что чистая прогностика никого не интересует и главный смысл состоит в том, чтобы найти уязвимости в характере, профессиональных навыках или стиле ведения процесса у судей, и использовать их в интересах подзащитных. Представляется, что распространение в Европе подобных систем предиктивной судебной аналитики усилит искажение равенства и состязательности в судебном процессе, даст преимущество защите перед обвинением, богатым юридическим фирмам перед независимой адвокатурой, состоятельным обвиняемым перед малоимущими.

Существует реальная опасность того, что будет подорван краеугольный принцип юстиции – принцип справедливого судебного разбирательства. Анализ использования дорогостоящих предиктивных судебных программ ведущими юридическими фирмами в Соединенных Штатах, Великобритании и Франции показывает, что адвокаты, располагающие такими системами, в среднем на 17–22 % чаще выигрывают дела, чем не располагающие. Повышение шансов на 1/5 буквально за два-три года свидетельствует о появлении конкурентного преимущества у адвокатов, имеющих системы предиктивной судебной аналитики.

Соответственно заслуживает внимания вопрос о запрете использования такого рода систем, по крайней мере, до тех пор, пока государственная сторона судебного процесса, в том числе прокуратура, не будут располагать аналогичными по эффективности системами. В противном случае это будет напоминать дуэль, когда один вооружен автоматом, а другой – старым мушкетом.

Кроме того, этот вопрос заставляет по-новому подойти к определению того объема данных по судебным процессам, которые не просто становятся доступными публике, но и хранятся. Представляется, что необходимо разделить два вопроса: *публикацию судебных решений и предоставление компьютеризированных данных о судебной практике.*

Целью открытых данных является обеспечение автоматической обработки прецедентного права в рамках недорогой операционной лицензии, приобретение которой доступно каждому. Соответственно, ЕС занимает крайне осторожную позицию относительно рекомендаций по расширению объема публикуемых данных судебной практики. Более того, эксперты ЕС ставят вопрос о необходимости еще раз рассмотреть вопрос *о целесообразности указания фамилии и имени судей, принявших то или иное решение*. С одной стороны, к такому порядку, действующему уже почти столетие, европейцы привыкли. С другой стороны надо ограничить возможность предиктивного анализа деятельности судей, поскольку, как показывает практика, в конечном счете, предиктивный анализ является способом управления поведением судей. Что же до управления поведением судей, то это – уголовное преступление во всех государствах ЕС.

Щекотливость ситуации состоит в том, что в настоящее время обязательность упоминаний имен и фамилий судей в судебных решениях является обязанностью государств-членов ЕС в соответствии с принципами публичного разбирательства, изложенными в статье 6 Европейской Конвенции о правах человека в части порядка обеспечения объективной беспристрастности судей.

Какие практические меры могут быть осуществлены для защиты судей в мире ИИ? Ответ на этот вопрос зависит от специфических особенностей судебной системы каждого конкретного государства. Например, в *Швейцарии*, где все судьи избираются непосредственно населением, публикация судебных решений является обязательным условием их ответственности судей перед гражданами-избирателями. Эта информация в обязательном порядке доступна в Швейцарии в кантональных и общешвейцарской онлайн судебной базе данных.

В то же время существуют и ограничительные правила. Например, *Европейский суд по правам человека на своем портале позволяет осуществлять поиск решений по именам и фамилиям судей, однако не допускает подсчет статистики, относящейся к конкретному судье*. Можно узнать разовые решения, но нет возможности получить по каждому судье выборку всех решений, которые он вынес. С другой стороны в странах, где судьи не выбираются гражданами, а назначаются государством, бывают случаи политического давления на судей и вынесение несправедливых решений. Соответственно, в таких

случаях ограничения информационного доступа общественности к судебной информации еще более усилят авторитарные течения и ослабят независимость судов.

Следует констатировать, что на сегодняшний день не только отсутствуют проверенные практикой рекомендации по решению данного вопроса, но и не проводилось, по крайней мере, в Европе, сколько-нибудь обширного сравнительного анализа способов защиты независимости судей в мире ИИ и больших данных.

§ 4. Характеристики ИИ применительно к судебным решениям

Обработка естественного языка и машинное обучение – две базовых технологии, лежащих в основе использования ИИ в судопроизводстве. В большинстве случаев системы ИИ в юриспруденции нацелены на то, чтобы выявить корреляции между различными параметрами, содержащимися во вводных и в итоговом решении. В качестве вводных параметров используются, например, содержащиеся в заявлениях о разводе сведения о доходах супругов, существовании супружеской измены, детях и т. п. Использование таких моделей позволяет развернуть прогнозную аналитику для предвидения будущего судебного решения – типа программ предиктивной судебной аналитики, наиболее востребованных юридическими фирмами.

В качестве введения следует кратко рассмотреть *функции, которые выполняют прогностическое программное обеспечение*. Это обеспечение исчисляет вероятности успеха или, напротив, неудачи дела в суде. Эти вероятности устанавливаются посредством статистического моделирования предыдущих решений с использованием двух широких областей информатики: обработки естественного языка и машинного обучения. В большинстве случаев использование именно этих двух областей в настоящее время и называется ИИ. Хотя по факту, это – не вполне ИИ, а продвинутый интеллектуальный софт, его, тем не менее, можно назвать «слабым» ИИ.

Поскольку термин «искусственный интеллект» сегодня используется не только в IT отрасли, то это породило множество двусмысленностей. Термин «искусственный интеллект» вошел в наш повседневный язык для описания программных комплексов, позволяющих компьютерам побеждать чемпионов-людей в игре Го, шахматы, водить машины, вести с людьми диалоги, в которых собеседнику трудно распознать, кто перед ним: человек или машина.

В самой IT отрасли предпочитают определять ИИ через используемые технологии. В этом понимании *слабый ИИ – это программные комплексы, использующие машинное обучение,*

нейронные сети, многомерный статистический анализ для распознавания образов, программы, обеспечивающие общение с компьютером на естественном языке и софт, позволяющий интегрировать в единую систему обработки информации потоки числовых, текстовых, аудио, видео и сигнальных данных. Сегодня ИИ – это не научное определение и даже не определение какой-то функции, а скорее *зонтичный термин для совокупности практически применяемых технологий.*

Сами специалисты различают слабый и сильный ИИ. Слабый ИИ – это некий набор сложных компьютерных программ. Что же касается сильного ИИ, который пока не создан, это – самообучающиеся машины, способные автоматически постигать мир в целом, и самостоятельно не только развивать софт, но и при необходимости достраивать свой хард.

Относительно ИИ в сфере правосудия. Системы юридической предиктивной аналитики направлены на определение всеми участниками судебного процесса наиболее вероятного исхода рассмотрения дела при имеющихся вводных. Теоретически такие системы не только облегчают жизнь участникам судебного процесса, но и могут помочь судьям и/или присяжным в принятии решений.

Предиктивный юридический софт на основе первичных данных, используя специфические критерии, определяет вероятность успеха (неуспеха) спора и возможные его имущественные итоги.

С самого начала необходимо разделить предсказания и прогнозы. Предсказания – это суждения о будущем событии на основании любых факторов – от интуиции до предчувствия. Что же касается прогнозирования, то это всегда – результат обработки исходных данных с использованием модели, разработанной на основе прошлых. Прогноз – это всегда счет. Это – область математики и статистики. Успешность прогнозирования де факто зависит от трех обстоятельств:

- *во-первых*, от качества исходных данных по судебным процессам. Чем данных больше и чем они точнее и полнее, тем больше шансов получить значимые результаты;
- *во-вторых*, от используемой прогностической модели. В подавляющем большинстве случаев для построения прогностической модели используются нейронные сети, либо многомерные статистические программы. На основе анализа прошлых данных эти

программы выясняют наиболее важные параметры, определяющие судебные решения, или так называемые веса критериев. С помощью этих весов нейронные сети делают прогноз;

- *в-третьих*, успех прогноза зависит от типичности судебного решения. Прогноз – это не предвидение. Он всегда базируется на прошлом, на данных об уже состоявшихся судебных процессах. Чем более настоящее похоже на прошлое, чем текущий процесс больше напоминает уже закончившиеся, тем больше шансов на качественный прогноз. Грубо говоря, в устойчивом мире прогнозирование более эффективно, чем в турбулентном, неустойчивом.

Надо понимать, что, несмотря на использование термина «интеллект» вычислительные машины немногим отличаются от арифмометра. Компьютер всегда считает и никогда не понимает. Соответственно, в основу предиктивной аналитики заложена гипотеза о возможности перевести все факторы и обстоятельства в числовой вид. Важно иметь в виду, что это – иллюзия. Обсчитывается только часть факторов. Самые важные факторы – эмоции, мотивы и побуждения – обсчитываются крайне плохо, либо вообще не обсчитываются. *Люди, зачастую, принимают решения, которые сами в последующем искренне объяснить не могут.* Тем более их невозможно рассчитать.

Согласно беседам, проведенным экспертами ЕС с производителями предиктивных судебных систем, выяснилось любопытное обстоятельство. Самым неотработанным блоком этой системы являются предиктивные возможности относительно решений судов присяжных за пределами учета тривиальных факторов пола, расы, имущественного дохода и т. п. Чем больше судебный процесс регулируется законом и чем меньше он носит прецедентный характер, тем больше шансов на успех предиктивного юридического софта. В этом смысле иронией выглядит факт, что примерно 85 % предиктивного юридического софта производят американские и британские фирмы, т. е. фирмы стран, где используется прецедентное законодательство.

Никакого сильного ИИ не только в настоящее время, но и в обозримом будущем не появится. Эволюция ИИ в течение последних примерно 10 лет идет не в сторону интеграции различных систем ИИ,

а, напротив, в направлении создания все более специализированных слабых ИИ, позволяющих использовать конкретные программные решения для строго определенных классов задач.

Существует *три основных тенденции развития ИИ в области судопроизводства и правоохранительной сферы:*

1. Чем дальше, тем больше ИИ будут функционально специализироваться и, по сути, смыкаться с традиционным интеллектуальным софтом, представляя следующую ступень его развития;

2. Главным направлением развития специализированного ИИ будет усиление возможностей комбинаторной аналитики;

3. Усилия разработчиков будут направлены на то, чтобы машины за максимально короткий период времени могли вычленивать из первичных больших данных максимальный набор переменных и установить между ними всеми по принципу «каждый с каждым» корреляционные связи. Собственно, главная задача специализированного ИИ и состоит в обнаружении нетривиальных связей, которые, тем не менее, влияют на те или иные процессы.

В настоящее время выделяется *три основных направления развития специализированного ИИ для судопроизводства:*

во-первых, это – создание систем распознавания прецедентов. Каждый судебный процесс может быть описан через ограниченное число параметров. В зависимости от законодательства той или иной страны при совпадении определенного числа параметров можно говорить о наличии прецедента и полагать, что решение по новому судебному делу будет достаточно близко к решению по прецеденту. В настоящее время, например, в Соединенных Штатах в гражданских процессах до 80 % оплаты юридической фирме клиентами идет на покрытие затрат на поиск этих самых прецедентов. Для европейских стран этот показатель ниже, но, например, во Франции он составляет около 50 % от гонорара юридических фирм.

Специалисты в области больших данных и работы с текстовыми массивами информации быстро обнаружили, что именно для этих юридических вопросов существует большое число уже имеющихся решений, связанных с ИИ. Фактически речь идет о том, чтобы имеющиеся мощные системы анализа неструктурированной текстовой

информации соединить с программами распознавания образов и получить прецедентный софт. На сегодняшний день именно эта сфера является самым большим рынком для использования ИИ в судопроизводстве в Соединенных Штатах, Великобритании и странах Британского Содружества наций. В Европе этот рынок только формируется;

- *во-вторых*, это – создание предиктивных юридических систем. О преимуществах, возможностях и ограничениях этих систем говорилось выше. Здесь стоит отметить лишь то, что, как это ни парадоксально, в странах, базирующихся на континентальном законодательстве, точность прогнозов выше, чем в странах прецедентного права. Слабым местом ИИ является прогнозирование именно одиночного, а не массового человеческого поведения. Соответственно на сегодняшний день системы ИИ не способны в удовлетворительной степени прогнозировать решения судов присяжных;

- *в-третьих*, специализированный ИИ активно развивается по направлению семантического поиска. Хорошо известно, что мало какая сфера человеческой деятельности связана с анализом огромных массивов документации, написанной на естественном или близком к нему юридическом языке, чем судопроизводство. Сегодня до 70 % затрат времени у наиболее высокооплачиваемых и квалифицированных юристов уходит на поиск тех или иных документов, примеров, а также создания на основе этих документов аналитических справок по персонам и т. п. С этой задачей как нельзя лучше справляется интеллектуальный софт, а именно корпоративные модификации семантических поисковых систем с блоками визуализации и генерации отчетов. По сути, такой софт был создан в Соединенных Штатах, Великобритании, во Франции еще в нулевые годы и продавался под грифом *datamining* и *textmining*. Сегодня этот софт успешно продается как частным юридическим фирмам, так и государственным судебным органам, как ИИ.

В целом, по состоянию на конец 2018 года *ИИ, как интегральное решение, является мифом. Практически повсеместно ИИ – это зонтичное наименование класса программно-аппаратных решений, которые существуют уже долгое время и успешно решают задачи определенного класса. Что действительно новое, так это тот факт, что в*

текущем десятилетии благодаря достижениям в производстве исходных материалов резко *повысилась скорость процессоров, а также появились мощные графические процессоры*. Благодаря этому оказалось возможным полностью проявить в практических делах потенциал математических методов, разработанных несколько десятилетий назад. По сути, *сегодняшний ИИ – это соединение графических процессоров, новых типов чипов с отработанными и прошедшими проверку временем программными математическими методами*. Также как в других отраслях, ИИ в значительной степени – это маркетинговый прием, такой же, как несколько лет назад был *datamining*, а до этого текстовый поиск неструктурированных данных. Что касается ИИ, о котором писали еще в прошлом веке, то это – сильный ИИ. Сильный ИИ на сегодняшний день является столь же недостижимой мечтой, что и 25 лет назад.

Только на первый взгляд может показаться, что приведенные выше соображения – это абстрактное обсуждение, далеко отстоящее от нужд судебной системы. Это не так. *Одной из главных причин слабого внедрения интеллектуального софта в судебной системе европейских стран является боязнь судебных работников нового, их опасения, что они окажутся просто неспособными освоить новые программно-аппаратные комплексы*.

Судебные работники, подобные другим неспециалистам в IT сфере, оказались «замороженными» компьютерным маркетингом и *всерьез решили, что появился ИИ, который их не сегодня-завтра заменит*. Необходимо не просто четко и ясно констатировать, но и донести до судебных работников простую мысль: *никакого сильного работающего ИИ нет и в обозримом будущем не будет*. Что касается новых решений, то они являются очередным этапом усовершенствования того софта, который судебные работники используют в своей профессиональной деятельности и частной жизни уже давно. В этом случае удастся сломать реально имеющиеся психологические барьеры по широкому использованию эффективных интеллектуальных программ и систем в судопроизводстве.

§ 5. Может ли ИИ моделировать судебные решения

ИИ базируется на том, что машинное обучение улучшает классификации, которые первоначально были созданы разработчиками, а также дополняет эти классификации новыми параметрами, выявленными машиной. Континентальные правовые системы далеки от рациональности, которая воплощается в гражданском кодексе Наполеона, принятом в 1804 г. Единый некогда кодекс в рамках Европы претерпел значительные изменения. В результате, сложились национальные правовые системы, которые хоть и имеют общее правовое ядро, тем не менее, значительно различаются и более того, противоречат друг другу. В настоящее время эти различия стремительно увеличивается. Право, ранее существовавшее десятилетия, теперь должно поспевать за переменами. Поскольку динамика технологического развития в европейских странах различна, различны и темпы изменения правовых систем.

Главным вопросом, с которым столкнулись IT эксперты в области юриспруденции, стало разложение закона на однозначно трактуемые логические правила. Старые интеллектуальные системы могли успешно работать с 200–300 логическими правилами. В настоящее время выяснилось, что *для того, чтобы описать закон с точностью, необходимой для практического использования решений на основе компьютерной логики, нужно оперировать с 1400–1800 правилами.*

Современные программно-аппаратные комплексы позволяют работать с системами, описываемыми до 10 000 правил. Именно в этом, а не в надуманном мышлении роботов, и состоит главный, принципиальный прогресс, достигнутый специализированными ИИ по сравнению с традиционным интеллектуальным софтом. Современный ИИ позволяет гораздо более точно распознавать ситуации, и что возможно еще более важно, без потерь описать законы на логическом, а соответственно и программном, языках.

До последнего времени из-за неточного перевода закона на алгоритмический язык невозможно было создать систему моделирования судебных рассуждений или размышлений. В

ближайшем будущем это станет возможным. Согласно французским и нидерландским исследователям – лидерам в этой области в Европе – в последние два-три года выявились из раза в раз повторяющиеся дефекты. Оказалось, что общеевропейское и страновые законодательства внутренне противоречивы. Если раньше правовой язык не представлялось возможным перевести на алгоритмический из-за неспособности свести право к логическим правилам, то в настоящее время эта проблема, как ни удивительно, осталась, но изменилась причина. Раньше у компьютера не хватало мощности, чтобы оперировать достаточным количеством правил для описания закона. Теперь же установлено, что закон не переводится в логические правила из-за внутренних противоречий, существующих в общеевропейском и национальных законах.

Эти противоречия не видны даже опытным правоведам. Они проявляются при рассмотрении того или иного вопроса в контексте нескольких законодательных актов. Отсюда возникает отдельная огромная задача. Вполне вероятно, что в ближайшее время юристы, если они преодолеют предубеждения по отношению к ИИ, и государственные деятели поставят перед сферой ИТ задачу осуществить сквозную проверку европейского и национального законодательства на противоречивость, наличие неточностей и лакунов, не позволяющих перевести закон в логические правила и создать внутренне непротиворечивую систему национального и общеевропейского законодательства.

Представляется, что если бы такая задача была поставлена, а тем более решена, это стало бы беспрецедентным вкладом ИТ технологий в юриспруденцию и значительно превзошло бы по своим последствиям любое другое использование ИИ в судебной сфере. Такая работа могла бы быть разбита на несколько этапов.

Во-первых, используя уже имеющиеся мощные языки логического программирования, типа Prolog, было бы необходимо написать максимально возможное количество логических правил для описания законодательства на машиночитаемом языке.

Во-вторых, перевести законы, записанные на естественном языке в законы, записанные на логическом машиночитаемом языке.

В-третьих, выделить в соответствии с принципами языка Prolog логические правила, как элементарные смысловые единицы

законодательства. Записать на языке логических правил законы.

В-четвертых, выявить методами стандартного логического моделирования противоречия между логическими правилами и интерпретировать их как противоречия внутри отдельных законодательных актов или между законодательными актами.

В-пятых, осуществить обратную интерпретацию выявленных противоречий, переведя их с языка логических правил и программной логики на естественный язык.

На основе этой работы устранить противоречия и лакуны как внутри законодательных актов, так и между актами, создав национальное и общеевропейское логически внутренне непротиворечивые законодательства^[2].

Помимо распознавания ошибок и противоречий в законодательстве работа с логическими моделями позволяет перевести на новый уровень предиктивную судебную аналитику. Об этом свидетельствует работа головной организации Великобритании по разработке логических юридических моделей и их использования для анализа права – Университетского Колледжа в Лондоне. Помимо своей основной работы Колледж создает системы предиктивной судебной аналитики. Программа 2015 г., созданная без учета логических моделей, давала прогнозы решений Европейского суда по правам человека с точностью до 57 %. С начала 2018 г. опробовалась новая программа, базирующаяся на логических правилах, как основе законов. Сегодня эта модель предсказывает исход конкретного дела в Европейском суде с точностью до 82 %.

Как представляется, участники судебного процесса, сталкиваясь с противоречиями между законами или в законе, чаще принимают оправдательные решения, нежели обвинительные. Соответственно этот факт был заложен в модель, и она стала давать гораздо более точные, чем раньше, прогнозы.

§ 6. Может ли ИИ объяснить поведение судей

Объяснение поведения судей, в частности, *своевременное выявление предвзятости*, является чрезвычайно важной и одновременно сложной задачей. Кроме того, один и тот же факт может интерпретироваться как предвзятость и как юридически оправданное мнение судьи. Едва ли не наиболее ярким примером такого рода амбивалентных ситуаций являются решения судов относительно заботы о детях. Гораздо чаще она поручается матерям, а не отцам. При этом, общей позицией европейского правосудия является то, что подобное решение – это не проявление предвзятости, а выбор судьи в пользу лучшего варианта, исходя из совокупности реальных обстоятельств.

С научной точки зрения объяснение человеческого поведения сводится к определению причинных механизмов, которые приводят к определенному поведению в конкретном контексте. Лионский университет в 2017 г. провел исследование относительно результатов передачи заботы о детях одному из родителей. Обследование велось на материале города Лион и нескольких прилегающих местностей. Выяснилась достаточно парадоксальная картина. Суды в 87 % случаев вынесли решение в пользу матери, и лишь 13 % – в пользу отца. Анализировались решения, вынесенные в 2016 г. Затем исследователи проверили всех детей, переданных соответственно матерям или отцам, на факты приводы в полицию, исключения из школы, других вопиющих нарушений, которые брались из официальной статистики. Было установлено, что 14 % из переданных матерям детей совершили за два года те или иные проступки, зафиксированные официально. Для детей, переданных отцам, этот показатель составил 3 %.

На обсуждении, которое исследователи из Лионского университета провели с судьями, последние категорически отвергали свою предвзятость. В качестве сильнейшего аргумента в свою пользу они приводили тот, что если долю детей-нарушителей из материнских и отцовских семей привести к общему знаменателю, то окажется, что в отцовских семьях нарушения все равно были чаще, а поэтому решения в пользу матерей судьи выносили верно.

Однако специалисты ИТ обратили внимание судей на весьма шаткий характер их аргументации. Они мешали факты и гипотезы. Факт

состоял в том, что гораздо больше нарушений в материнских семьях, а гипотеза – в том, что если бы в материнские и отцовские семьи передавали бы детей поровну, а нарушали они порядок так, как было обнаружено по факту, тогда бы отцы оказались гораздо более худшими воспитателями. С точки зрения научной статистической теории экспериментов однозначно правы были IT специалисты, а не судьи.

Этот простой эксперимент показал три очень интересных обстоятельства.

Во-первых, сами по себе фактические данные никак не влияют ни на решения судей, ни на оценку решений общественным мнением.

Во-вторых, понятие предвзятости – это нечеткое, не установленное законодательно понятие. Это – оценочное суждение, выносимое общественным мнением. Поскольку общественное мнение само крайне манипулируемо, то ссылки на предвзятость или непредвзятость судей в строгом статистическом или научном плане не имеют никакого отношения к реальной ситуации.

В-третьих, если последствия одного судебного решения предсказать на основе статистических данных не представляется возможным, то для совокупности статистика дает эффективный инструмент прогноза. Последнее очень важно. Оно демонстрирует силу и слабость статистики. Статистика – мощный аналитический и прогнозный инструмент. Но на основании статистических данных нельзя принимать решения в каждом конкретном случае хотя бы потому, что статистика – это наука, а соответственно инструментарий для выявления тенденций массовых явлений.

Поскольку подавляющее большинство, если не сказать все, методы ИИ базируются на статистике, необходимо четко и однозначно сказать обществу, что судья-ИИ невозможен. Соответственно направление моделирования индивидуального поведения судьи методами ИИ – пустая трата средств. Если судья находится под подозрением, то задачу подтверждения или опровержения этих подозрений должен решать не ИИ, а конкретная следственная работа.

§ 7. Как ИИ должен применяться в гражданском, коммерческом и административном правосудии

Уже сегодня, а тем более завтра, специализированный ИИ обеспечит достаточно высокий уровень прогнозирования судебных решений. С другой стороны, использование ИИ за пределами прогнозирования, textmining и распознавания прецедентов вряд ли может дать что-то полезное судопроизводству в ближайшее время.

Посмотрим на эту тему с позиций специфики разных направлений судопроизводства. Как мы видим, главный вопрос, возникающий при использовании ИИ в юриспруденции, заключается не в абстрактных рассуждениях, полезен он или вреден, а в конкретном анализе, будет ли использование ИИ способствовать лучшему правосудию или нет.

При всех несовершенствах и недостатках ИИ, представляется важным его внедрение в судопроизводство на стороне судей и обвинения, хотя бы для того, чтобы сохранить, а в отдельных случаях вернуть, равенство и состязательность судебного процесса. Например, ведущие американские юридические фирмы в 85 % коммерческих и гражданских процессов, которые они ведут не только в Америке, но и в Европе и в целом по миру, уже сегодня используют ИИ, прежде всего, для прогнозирования, облегчения работы с документами и распознавания прецедентов. Ситуация, когда одна сторона вооружена ИИ, а другая дискутирует, плох он или хорош, негативно сказывается на правосудии. Надо признать, что у европейских судов нет альтернативы использованию ИИ. Это верно для любой страны, вне зависимости от того, применяется ли в ней континентальное или прецедентное право, как организованы судебные процессы и т. п.

Сегодня юридические фирмы широко используют ИИ, особенно в предиктивной его составляющей, не только для того, чтобы решить, браться или не браться за тот или иной процесс, как строить работу, но и для поощрения заключения сделок еще на досудебной стадии, чтобы избежать долгого и неопределенного судебного испытания. Некоторые страховые компании во Франции, Нидерландах, Германии уже используют прогнозные юридические системы во взаимоотношениях с

клиентами для того, чтобы определить размеры страховых платежей и риски, связанные с взаимоотношениями с клиентами.

§ 8. Новые компьютерные инструменты для расчета издержек судебных процессов

В условиях бюджетных дефицитов, характерных для большинства европейских стран, и слабого экономического роста особую остроту приобретают вопросы, связанные с компенсациями и издержками ведения судебных процессов. В настоящее время лишь в одном государстве ЕС ИИ используется для оптимизации судебных расходов и расчетов компенсаций по результатам судебных процессов. Этой страной является Франция. Больше нигде в континентальной Европе вычислительная техника в этих целях не применяется.

С участием исследовательских учреждений, по заказу Европейской Комиссии был проведен анализ причин, по которым ИИ не используется для, казалось бы, наиболее подходящих для него задач, а именно определения степени обоснованности расходов на ведение судебных дел и прогноза компенсаций. Выяснилась достаточно поразительная картина. Главной причиной является тот факт, что, несмотря на многочисленные отчеты и высказывания руководителей судебных органов, практически нигде, кроме Франции и Нидерландов, не имеется выполненных в одних и тех же масштабах архивов судебной документации в электронном виде. По-прежнему хранение происходит частично на бумажных, частично на электронных носителях и в разных форматах. Кроме того, отсутствует четкая увязка судебной и бухгалтерской, финансовой документации по судебным процессам.

Все это свидетельствует о том, что для большинства европейских стран в ближайшее время использование ИИ в судопроизводстве – это неподъемная и неактуальная задача. Им необходимо пройти свой путь по внедрению стандартов электронного документооборота в судебном деле, создания хранилищ структурированной и неструктурированной судебной информации, согласования судебной и бухгалтерской отчетности.

Как показывает опыт, такая работа может потребовать как минимум пяти-семи лет. До этого использование ИТ технологий в судебном деле

будет ограничено обычным офисным софтом и корпоративными поисковиками, работающими с архивами.

§ 9. Проблема высоких издержек и ИИ

В настоящее время две страны ЕС и еще одно государство за пределами ЕС имеют государственные программы повышения информационной вооруженности судопроизводства и внедрения в эту сферу интеллектуального софта. Этими государствами являются Великобритания, Нидерланды и Латвия. Кроме того, до 2021 г. на полностью электронный документооборот в государственной сфере и сфере любых юридических взаимоотношений перейдет Эстония. Это позволит стране стать первой в Европе, а возможно и в мире страной, где все правовые акты, все договоры, все отношения будут иметь электронный характер и соответственно без труда как большие данные смогут быть обработаны ИИ.

§ 10. Интернет-услуги и сфера судопроизводства

В настоящее время сложилась парадоксальная ситуация. В целом, по государствам ЕС более 17 % товарооборота приходится на интернет-сервисы, рассчитанные как на население, так и на бизнес. При этом, судебные процессы по спорам в интернете продолжают вестись так же, как судебные процессы в реальной жизни. Доходит до смешного. Интернет-компании специально содержат значительную численность персонала в юридических отделах, чтобы переводить электронные документы в бумажные для представления в суды.

Только пять европейских стран приняли и осуществляют программу создания системы электронных судебных услуг, включающих, в том числе, проведение распределенных и удаленных судебных процессов в онлайн с участием современных коммуникаторов. Такие программы реализуются в рамках общенациональной программы электронного государства в Эстонии, плана автоматизации сферы судопроизводства и правоохранительной деятельности в Латвии, программы цифровизации правоохранительной, судебной и регулирующей сфер в Нидерландах и как отдельная программа, реализуемая на уровне муниципалитетов и территорий в Великобритании. В программно-аппаратном компоненте задел для перехода к электронному судопроизводству, включая автоматизированный расчет издержек на проведение судебных процессов и компенсаций, как основы досудебного урегулирования, имеется во Франции. Однако, из-за политических разногласий между правящими силами и оппозицией этой страны, подкрепить имеющийся задел принятой на государственном уровне программы использования ИИ в судебной и правоохранительной сфере, пока не удалось. Поэтому французские компании более активно работают на судебном рынке Канады и отчасти Великобритании, чем на родной территории.

§ 11. Основные гарантии, которые должны быть подтверждены в гражданском, коммерческом и административном судопроизводствах

Предоставление инструментов ИИ для разрешения споров в режиме онлайн не должно влиять на право беспрепятственного доступа к суду и равной состязательности в судебных процессах. Например, в гражданских делах каждый истец имеет право передавать в суд любой спор, касающийся его гражданских прав и обязанностей.

В 2015 г. парламентская Ассамблея Совета Европы приняла резолюцию «Доступ к правосудию через интернет: потенциал и проблемы». Согласно этой резолюции любой участник судебного процесса в том случае, если используются элементы ИИ, должен иметь полный доступ к большим данным, которые обрабатываются ИИ, а также алгоритмам, на основании которых принимается решение. При этом создатели судебного ИИ обязаны разъяснить суть алгоритма любому участнику судебного процесса.

По факту это означает, что значительная часть наиболее эффективных систем ИИ, занимающаяся прогнозированием или распознаванием сложных сущностей, не может быть применена в судебной сфере европейских стран. Проблема состоит в том, что эффективность ИИ напрямую связана со сложностью нейронных сетей, которые, как правило, занимаются распознаванием. Поскольку нейронные сети постоянно проходят машинное обучение, то зачастую разработчики алгоритма просто не знают промежуточные алгоритмы, влияющие на конечный вывод. Эти алгоритмы формируются в процессе обучения и постоянно меняются.

Сегодня в целом ряде видов деятельности складывается парадоксальная ситуация, когда наиболее эффективные программы с точки зрения распознавания и принятия решений делают это недоступным для понимания людьми без глубокого математического образования образом. Поскольку объяснить алгоритмы без использования сложных разделом математики невозможно, то в

соответствии с резолюцией Европейской Комиссии, любой участник процесса может заблокировать применение наиболее эффективных методов ИИ в этом процессе из-за их необъяснимости.

Поскольку, как показывает пример Соединенных Штатов, и отчасти Великобритании, использование ИИ в судопроизводстве, в конечном счете, неотвратимо, поскольку любая имеющаяся технология рано или поздно используется, необходимо законодательно решить главную проблему.

Главная проблема использования ИИ в судопроизводстве состоит в подрыве равенства сторон в пользу государственных учреждений, компаний со значительными средствами, людей, имеющих образовательные и профессиональные навыки в сфере высоких технологий. При этом использование ИИ делает судебную систему еще более непонятной, а потому враждебной для значительных групп населения, которые не обладают необходимыми знаниями или средствами для найма консультантов. Поскольку для большинства европейских стран реальный горизонт внедрения ИИ в судебную и правоохранительную сферы составляет от пяти до десяти лет, важно использовать эти годы для создания равных возможностей использования ИИ в судопроизводстве для всех участников судебного процесса. Как это сделать сегодня не вполне ясно. Однако, как показывает практика, сама постановка проблемы и привлечение к ней внимания повышает, в конечном счете, шансы на ее решение.

Использование прогнозной аналитики порождает и еще одну проблему. Об этом свидетельствует опыт Соединенных Штатов. В Америке часты случаи, когда юридические фирмы, осуществив прогноз по тому или иному делу, отказываются иметь дела с истцами или ответчиками, у которых малы шансы на успех. Тем самым люди, нуждающиеся в юридической консультации, в конечном счете, лишаются ее. Это подрывает основы судопроизводства.

Вероятно, внедрение ИИ потребует не только изменения норм, регулирующих судебные дела, но и принятия нового кодекса профессиональной этики для адвокатов.

§ 12. Вопросы, характерные для уголовного правосудия: предупреждение правонарушений, риск рецидива и оценка уровня опасности

Использование ИИ в уголовном судопроизводстве *создает риск возрождения детерминированных доктрин в ущерб доктринам индивидуализации санкций*, которые получили широкое распространение в большинстве европейских стран, начиная с 1945 г. Использование ИИ в уголовных делах ставит особо острые проблемы. Они связаны с возможностью предсказывать правонарушения. В течение 30 лет во многих европейских странах шли дискуссии на этот счет. Итогом дискуссий стали конкретные законодательные нормы.

Например, в Италии п.2 ст. 220 Уголовно-процессуального кодекса прямо запрещает использование привычных или профессиональных признаков, а также анализ психологических черт, как меток предрасположенности человека к преступлению. Во Франции доктрина «Новая социальная защита» предусматривает введение системы социальной реабилитации, предотвращающей преступления за счет ликвидации условий для его совершения, связанных прежде всего с неблагоприятными условиями жизни, отсутствием доступа к образованию, плохих условий занятости и т. п.

Этот подход заложен во французском, бельгийском и испанском законодательствах.

Иными словами, официальная позиция большинства судебных органов европейских стран, нашедшая отражение, в том числе в формах, регулирующих судопроизводство, состоит в отказе от концепции предрасположенности к совершению преступления. Таким образом, возникает чрезвычайно серьезное противоречие между сложившейся в европейском судопроизводстве традицией отказа от учета предрасположенности в совершении преступления на основе анализа психологических, социальных, медицинских, профессиональных и иных факторов, и базовой технологии ИИ – нейросетями и машинным обучением.

Суть этой технологии как раз состоит в том, чтобы первоначально сгруппировать все данные по максимальному числу факторов,

влияющих на ту или иную функцию, например, совершения преступления, а затем рассчитать уровень влияния каждого фактора и их комбинации на конечную функцию, т. е. вероятность совершения преступления.

По своей глубинной сути, нейронные сети – это программы максимально глубокого комбинаторного анализа на взаимосвязь различных факторов и осуществления прогноза на этой основе. Программа сама по себе находится по другую сторону добра и зла, этичного или неэтичного, корректного или некорректного. Не существует этичных или неэтичных программ. Есть программы, позволяющие реализовать адекватные реальности модели, а есть – неадекватные, не соответствующие фактическому положению дел.

В настоящее время в странах ЕС с большой долей национальных меньшинств, особенно Франции, Бельгии, Германии, Дании и Швеции активистские и общественные организации пытаются заблокировать использование предиктивных моделей в полицейских расследованиях и судебной аналитике. В ходе дискуссий в интернете и социальных медиа активисты высказываются в том плане, что нейронные сети неизбежно покажут большую вероятность совершения преступления у бедных, молодых, без достаточного образования представителей меньшинств. Они не оспаривают, что ИИ точно посчитает факторы. Их аргумент состоит в том, что коренные жители Европы виновны перед меньшинствами за колониальный период и поэтому полицейские и судьи не могут подходить с одной меркой к коренному населению и недавним мигрантам. Мерки должны быть разными. Более мягкие – для мигрантов, перед которыми якобы виноваты коренные европейцы. Именно в этом состоит сегодня главный предмет дискуссии относительно использования ИИ, прежде всего, в полиции, а также в судебной системе.

Обработка больших данных по преступности методами ИИ и, прежде всего, глубоких нейронных сетей с машинным обучением предоставляет обществу реальную картину факторов, влияющих на совершение конкретных преступлений, причем в каждом конкретном месте. Обвинять программы в неполиткорректности – глупость. Более того, глупость – не учитывать программные результаты, а тем более прогнозы в правоохранительной деятельности и судебной сфере.

При этом эксперты ЕС полагают, что *нормативный запрет на анализ причин преступности, реализованный в настоящее время в европейской судебной системе, не сокращает, а увеличивает уровень преступности*. В большинстве европейских стран существуют многоступенчатые запреты на предиктивную и профилактическую деятельность полиции. Эти запреты подталкивают полицию не предупреждать, а быстро раскрывать уже совершенные преступления. Между тем, в условиях широкого использования преступниками вредоносного софта, генных инструментов преступности и т. п. такой подход оказывается слишком опасным для общества. Единственно возможным становится превентивный подход, базирующийся не на раскрытии преступления, а на его предупреждении.

Сегодня законодательства подавляющего большинства европейских стран ориентировано на запаздывающее, а не на опережающее реагирование на преступление. Вполне вероятно, что именно тема превентивной борьбы с преступностью станет в ближайшие годы основной темой общественных дискуссий в сфере законодательства. *Без перехода от парадигмы наказания преступников к парадигме предупреждения преступлений предиктивный полицейский софт применить невозможно*. А если ограничить полицию в использовании предиктивного софта, то она неизбежно будет проигрывать преступникам, которые такой софт уже используют в поисках наиболее прибыльных и наименее рискованных направлений, сфер и мест деятельности.

§ 13. Инструменты, используемые следственными органами на стадии расследования

Инструменты, получившие название «предиктивного полицейского софта» уже широко используются в Соединенных Штатах, Канаде, Китае, Великобритании, Японии, Нидерландах и отчасти Франции. Эти системы базируются на аккумулировании данных из всех источников, всех форматов и позволяют выявлять неочевидные связи и прогнозировать действия. Наиболее известный программный комплекс такого типа, это – *Palantir* в Соединенных Штатах.

В целом, в европейской полиции большая часть компьютерных инструментов используется не для предотвращения преступных действий, а для эффективного судебного преследования преступников. Предиктивная аналитика в настоящее время находится на вооружении в ФБР, в полицейских управлениях большинства городов США, Великобритании, в Гонконге, Сингапуре и Китае. Как показывает практика, наиболее успешные полицейские прогнозы системы дают в отношении краж со взломом, уличного насилия, угона транспортных средства, уличной торговли наркотиками – по направлению низкоуровневой преступности, а также финансовых преступлений и мошенничеств, криминального оборота наркотиков и торговли животными и птицами – в высокоуровневой преступности.

Предиктивные системы позволяют устанавливать горячие локации и места наиболее возможного совершения преступлений, а также давать полицейским на земле наводки на конкретных лиц, вовлеченных в организованную преступность, которые осуществляют действия, интерпретируемые как подготовка к преступлению. Такие системы, как правило, используют в качестве исходного сырья больших данных разноформатные данные – тексты, видео, аудио, данные по транзакциям и т. п., -которые обрабатываются в единой интегрированной среде.

В то же время этот софт не свободен от недостатков. Главными из них является эффект «порочных кругов или самосбывающихся пророчеств». Программы ориентируют полицию на районы,

подвергающиеся максимальному риску. Полиция усиливает патрулирование этих районов, выявляет больше преступлений и соответственно система на следующем уровне еще более определенно показывает на те же самые районы. Таким образом, одним районам уделяется практически все внимание, а в других, лишенных полицейского присмотра, создаются условия для правонарушений.

Кроме того, в Соединенных Штатах меньшинства пытаются затормозить использование предиктивного полицейского софта по причине тирании алгоритма, который опираясь на данные о прошлом, дает прогнозы о будущем. В этой связи ранее уже указывалось, что при отсутствии предиктивной аналитики любые попытки полиции обвинить в возможном преступлении представителя меньшинств, сразу получают отпор как проявление фашизма, расизма, предвзятости и т. п. Полицейские, не желая втягиваться в длительные процессы по превышению полномочий или возможной предвзятости, все чаще игнорируют мелкие преступления и оставляют преступников безнаказанными. Машину обвинить в предвзятости невозможно. Соответственно, внедрение предиктивного софта резко повышает не только качество полицейской работы, но и независимость полицейских от, как правило, неадекватного общественного мнения. По сути, это – не общественное мнение, а специфический вид лоббизма.

По мнению экспертов ЕС, поскольку в Европе проблема меньшинств стоит, возможно, даже острее, чем в Соединенных Штатах, надо быть заранее готовым к конфликту с гражданскими активистами на этой почве. *Желательно закрепить нормы использования предиктивного полицейского софта непосредственно в законодательстве.* В противном случае у европейской полиции никогда не будет этого мощнейшего инструмента борьбы с организованной преступностью.

В отличие от судебных работников, полицейские вместе с сотрудниками разведывательных служб являются энтузиастами и заинтересованными эксплуатантами больших данных и ИИ. Поскольку компании-производители не только видят большой интерес полиции к системе ИИ, но и активно привлекаются ей для решения конкретных задач, именно в сфере полиции действуют сегодня наиболее эффективные платформы ИИ.

Помимо уже упомянутого Palantir, необходимо отметить используемую полицией Великобритании платформу ИИ *Connect*. Эта платформа в режиме реального времени аккумулирует миллиарды транзакций, осуществляемых в банках, находящихся под юрисдикцией Британской короны, и на основе нейронных сетей распознает среди них транзакции и цепочки, которые могут указывать на подготовку или уже совершившиеся преступления. Не только британские полицейские, но и полицейские других стран, получающие время от времени доступ к системе называют ее лучшим инструментом для поддержки финансовых расследований в мире. Эта система реализует важнейший принцип деятельности любого опытного полицейского – хочешь найти преступника, иди по следам его денег.

Также высокую оценку со стороны полицейских получила *международная база данных по сексуальной эксплуатации детей, управляемая Интерполом*. Она помогает идентифицировать жертв и преступников, в том числе путем распознавания предметов, выявленных на фотографиях и видеоизображениях, и даже использовать для поиска преступников фоновые шумы в видеороликах.

Эти две системы показывают, что при наличии заинтересованности уже достигнутый уровень развития ИИ позволяет создавать софт, оказывающий значительную помощь полицейской работе в режиме реального времени. Причем, этот софт, имеющий мощный графический интерфейс, понятен и дружелюбен для пользователей на земле – патрульных машинах и т. п.

Представляется, что еще одним фактором, обеспечивающим триумф предсказательного софта в полиции, стало следующее обстоятельство. В подавляющем большинстве стран уже более 10 лет происходит сокращение бюджетов правоохранительных органов. Перед полицией, особенно на низовом уровне, во весь рост встала проблема сохранения высокой эффективности работы при снижении уровня ассигнований. Хорошо известно, что профилактика и предотвращение преступлений в подавляющем большинстве случаев эффективнее и дешевле, чем их расследование. Соответственно полиция с большей охотой занимается пресечением деятельности преступников на стадии подготовки правонарушения. При этом судебная система охотнее имеет дело с судебными спорами относительно уже совершенных

преступлений. В подавляющем большинстве стран отсутствуют правовые нормы, связанные с серьезным наказанием за подготовку преступления. Это порождает в последние годы коллизии, когда во все возрастающем числе случаев люди, относительно которых полиция собрала неопровержимые доказательства по подготовке преступления, в конечном счете, избегают судебного наказания, поскольку судебные нормы ориентированы на уже совершенные, а не готовящиеся преступления.

По мнению экспертов ЕС, необходимо на уровне Европейской Комиссии начать дискуссию о *целесообразности совершенствования европейского законодательства с упором на предотвращение, а не наказание преступлений*. Если подобного рода дискуссию не провести, и не осуществить соответствующие меры, то зазор между деятельностью полиции и работой правоохранителей будет из года в год увеличиваться.

На сегодняшний день в Европе предиктивная аналитика в уголовных процессах, по сути, игнорируется судьями. В 2012 г. Кембриджский университет по собственной инициативе разработал и предоставил в распоряжение судебных работников и полиции инструменты оценки риска повторных преступлений, так называемый *HART*. Исследования, проведенные в 2013–2017 гг., показали, что для низкоуровневой уличной преступности HART дает правильные прогнозы в 98 % случаев, а применительно к компьютерной и другой высокоуровневой преступности во впечатляющих 88 % случаев. Система в течение трех месяцев после освобождения из мест заключения дает прогноз рецидива на ближайшие два года. Достигнутые результаты поражают, поскольку не имеют аналогов в мире. Поэтому британские и голландские правоохранительные органы всю используют HART и надеются, что в ближайшем будущем повышенное внимание к лицам высокой группы риска удержит последних от новых преступлений. При этом судьи, в том числе и британские, отказались использовать HART как консультанта при вынесении приговоров. Общая позиция судей состоит в том, что правосудие слишком серьезный предмет, чтобы допускать в эту сферу ИИ.

HART – это единственный прогностический инструмент, используемый в настоящее время в Европе. Анализ показал, что ни в одной стране, кроме Франции в настоящий момент нет планов

разработать мощный предиктивный инструмент, ориентированный на уголовные расследования. Есть основания полагать, что государства-члены ЕС уже в ближайшие годы будут все более и более отставать от Соединенных Штатов и Китая. В самой же Европе лишь три страны, ориентированные в основном на Соединенные Штаты, будут стараться реализовать прогностический уголовный софт. Это – Голландия, Латвия и Эстония, которые в основном будут использовать зарубежные разработки.

В 2018 г. в Соединенных Штатах имел место скандал, когда неправительственная организация Propublica раскрыла дискриминационный эффект алгоритма, используемого в программном обеспечении *Compas*. Целью софта является оценка риска рецидива на стадии, когда судья выносит приговор для отдельного лица. Аргумент противников системы состоял в том, что в алгоритм заложена гораздо большая вероятность рецидива для афроамериканцев и латинос, в том числе недавно получивших гражданство. В отличие от Европы, компания-разработчик софта совместно с Министерством юстиции США и правоохранительными органами заняли жесткую и неуступчивую позицию в дискуссии с активистами. Они сделали доступными для общественности все шаги алгоритма, вычисляющего коэффициент, и показали, что конечные значения риска зависят не от какой-то идеологической или иной дискриминации, а базируются исключительно на фактическом положении дел. Активистов же, превратившихся в лоббистов, общественное мнение, благодаря прессе, обвинило в попытках манипуляции американскими ценностями для того, чтобы сделать фактически неравным правосудие, создав преимущества для афроамериканцев и латинос в ущерб белым американцам.

§ 14. Проблемы использования прогнозов в уголовных делах

Как уже отмечалось, использование прогнозного инструментария в Европе заметно и все более отстает от ситуации в Северной Америке и странах Британского Содружества наций. Также установлено, что как сама полиция, так и общество гораздо активнее выступают и используют предиктивный полицейский софт в сравнении с судьями, в том числе с судьями в уголовных процессах.

Общеввропейской позицией является сегодня констатация, что ИИ дает огромные возможности полиции, но содержит и реальные риски, если ИИ используется без необходимых мер предосторожности и ограничений. Согласно результатам опроса парламентариев в девяти странах Европейского Союза, более 78 % из них активно ратуют за широкое использование предиктивного полицейского софта, особенно нацеленного на предупреждение преступлений. В то же время лишь 15 % опрошенных поддержали использование ИИ в судебной сфере, за исключением нормализации электронного документооборота, за который выступают все.

В комментариях высказывалась точка зрения, что в полиции использование ИИ нет альтернативы хотя бы потому, что это уже активно делает криминал. Что же касается судебной сферы, то здесь главное – обеспечить равенство сторон. На вопрос о том, не даст ли ИИ преимущества крупным юридическим фирмам, ответы в 85 % случаев были положительными. Самый неожиданный эффект – в том, что более половины парламентариев высказалось за мораторий использовании ИИ вообще в судебной сфере в ближайшие 15 лет.

Категорическими противниками активного использования ИИ в уголовных процессах являются сами судьи. Во Франции, которая является одной из самых продвинутых стран в использовании интеллектуального софта в судопроизводстве, лишь немногим более 15 % судей полагают, что ИИ им в ближайшие 10 лет необходим. Главным аргументом в пользу негативного отношения к ИИ является тот факт, что *значительная часть специфических человеческих черт, таких как настроение, контекст, ценности, эмоции не могут быть*

оцифрованы, и соответственно не будут приниматься во внимание машиной. А именно эти факторы являются, по мнению судей, чрезвычайно важными при вынесении вердикта, особенно в уголовных процессах.

Существенно различаются позиции экспертов, специалистов и общественности относительно применения ИИ в правоохранительной деятельности и в уголовном судопроизводстве. Подавляющее большинство не только профессионалов, но и гражданских активистов, хотя и с оговорками, но поддерживают широкое использование ИИ в полиции. Это одобрение базируется на мнении, согласно которому *инструменты ИИ являются не более чем вспомогательными компонентами работы следователя.* Более того, полиция с самых первых годов своей институционализации старалась максимально использовать бумажные аналоги нынешнего ИИ, а именно бумажные архивы, досье на преступников и преступные группировки, фотографии, постоянно собирала информацию как опосредовано, так и путем внедрения осведомителей.

В современных западных обществах сложился консенсус относительно того, что работа полиции всегда по существу носила и носит информационный характер, включая поиск, хранение, обработку и анализ информации относительно готовящихся или совершенных преступлений. Будучи де факто с самых своих истоков информационным видом деятельности, полиция как нельзя лучше подходит для использования IT инструментария. Поэтому во всех странах, где для этого есть технологические и финансовые возможности, полиция вместе с разведкой и банковской сферой являются основными секторами внедрения передовых информационных технологий.

Что же до судопроизводства, то, по крайней мере, в государствах ЕС не только подавляющая часть активистов, но и самих работников судопроизводства рассматривают ИИ исключительно как помощника в работе с документами, включая их архивирование, поиск и т. п. Этому способствуют три обстоятельства.

Первое. По странным причинам в экспертном сообществе и общественном мнении сложилась фантастическая картина, что в будущем судопроизводством будет заниматься именно ИИ, в конечном счете, будет выносить приговоры и определять судьбы.

Второе принципиальное возражение против использования ИИ заключается, как описано выше, в том, что использование продвинутых информационных технологий, в том числе ИИ, подрывает равенство перед законом и состязательность судебных процессов.

Наконец, *третья группа* возражений против широкого использования ИИ в уголовном судопроизводстве имеет, если можно так выразиться, статистический характер. Профессионалы понимают, что, по сути, специализированный ИИ – это мощнейший вычислитель, обрабатывающий статистические данные. Соответственно использование предиктивной, особенно судебной аналитики, может привести к тому, что прошлое поведение определенной группы людей с конкретными параметрами будет, в конечном счете, решить судьбу отдельного человека. Каждый подсудимый, хотя и имеет определенные, общие с другим людьми, характеристики, но оказывается на судебной скамье в результате комбинации уникальных жизненных обстоятельств. Единичная судьба никак не связана и не коррелируется с групповой статистикой. Пожалуй, в западноевропейских государствах ЕС именно возможность статистических данных влиять на судьбы отдельных людей и является главным аргументом общественности против использования ИИ в судебной сфере.

§ 15. Потенциал и ограничения предиктивной судебной аналитики

По мнению экспертов ЭС, *термин, пришедший из-за океана, «предиктивная судебная аналитика» необходимо отбросить*, поскольку он неоднозначен и вводит в заблуждение. Гораздо лучше использовать *термин «интеллектуальный аналитический судебный софт»*. Данный термин не вызывает нареканий у судей и общественности и принят в ИТ отрасли Европы.

Консенсусная точка зрения среди специалистов по софту для правоохранительной и судебной сфер состоит в следующем. Нейтральность алгоритмов – это миф. Машины еще не научились самостоятельно писать алгоритмы. Даже нейронные сети – это не самосовершенствующиеся программы, как многие считают, а программы, позволяющие менять оценку тех или иных факторов, используемых для распознавания чего-то, в зависимости от вновь поступающих данных. *ИИ – это произведенное человеком устройство, выполняющее команды, которые записаны на созданных людьми искусственных языках для выполнения тех или иных расчетных алгоритмов, порожденных не машиной, а конкретным человеческим разумом*. Т. е. все три главных компонента ИИ – это машины, языки и алгоритмы – созданы людьми.

Поскольку любые люди, и математики не исключение, имеют свои картины мира, взгляды, ценности, то так или иначе они отражают их в алгоритмах. Любой алгоритм, будучи порождением определенных математиков и разработчиков, несет в себе их внутренний мир. Когда люди выступают против ИИ, они в иррациональной форме демонстрируют свою неготовность вручить свою жизнь математике. Фактически замена судей алгоритмами не проходит в Европе, достаточно далекой от маркетингового бизнеса, потому, что ни государства, ни общество не готовы поменять профессионалов – судей, на математиков, которые ничего не понимают в законах. Все остальные соображения о боязни ИИ – это просто американские феномены, которые сами компьютерные фирмы используют для продвижения продукции.

10 лет назад никто не предлагал отказаться от компьютерных программ, например, для написания судебных документов или расчетов тех или иных данных. Ни одному здравомыслящему человеку не могла прийти в голову мысль наделять текстовой процессор Word или таблицы Excel свойствами искусственного разума. В условиях экономической стагнации IT компании вложили гигантские, исчисляемые миллиардами долларов, средства в маркетинг, рекламу и заморочили голову потребителю появлением чего-то похожего на человеческий разум у машин. При этом за эти годы алгоритмика не изменилась, выросла лишь производительность компьютеров и появились впервые в истории человечества большие данные.

К сожалению, историческая форма использования интеллектуального софта в судебной сфере завела общество и судебную систему в дебри бессмысленных дискуссий об ИИ. Эксперты ЕС считают, что с этим положением необходимо покончить. Прежде всего, они предлагают на общеевропейском уровне провести серию конференций и круглых столов, собрав их вместе с разработчиками инструментов, юристами, работниками государственных органов, средств массовой информации, для того, чтобы выработать четкую, ясную, соответствующую реалиям и помогающую делу терминологию. Судебные советы, профессиональные ассоциации и коллегии адвокатов также должны принять активное участие в этих круглых столах. Необходимо покончить с ситуацией, когда в юридической сфере мелькают случайные или рекламные IT термины, затрудняющие использование передовых технологий.

Необходимо создать *группы технического аудита* и предоставить им право исследовать предлагаемые частным сектором как ИИ прогнозноаналитические и иные программы. Главная же задача состоит в определении на основе широких дискуссий конкретных участков работы и сфер, где в ближайшие годы наиболее эффективно использование интеллектуального прогнозно-аналитического софта в судопроизводстве.

Глава II

Европейская Этическая Хартия использования искусственного интеллекта в судебной и правоохранительной системах

Признавая возрастающую важность ИИ в современном обществе, включая ожидаемые выгоды, Европейская комиссия на 31 пленарном заседании 3–4 декабря 2018 года в Страсбурге приняла *пять фундаментальных принципов, содержащихся в Европейской Этической Хартии, использования искусственного интеллекта в судебной и правоохранительной системах.*

Хартия предназначена для руководства в работе государственных и частных заинтересованных сторон, ответственных за создание и внедрение ИИ и его компонентов в судебной и правоохранительной сферах (методы машинного интеллекта включают машинное обучение или любые другие способы обработки больших данных). Хартия также касается лиц, принимающих решения в сфере законодательной или нормативной базы, а также оказывающих связанные с этим услуги.

Хартия поддерживает использование методов и компонентов ИИ для повышения эффективности и качества правосудия и правоохранительной работы. Однако она обращает особое внимание на необходимость максимально ответственного подхода к этой деятельности, обеспечивающего полный учет прав личности, изложенных в Европейской Конвенции о правах человека и Конвенции о защите данных, а также других нормативных материалах, связанных с гражданскими правами личности.

Использование ИИ для обработки судебных решений и правоохранительной деятельности поможет повысить уровень раскрываемости преступлений, поднять уровень доказательности направляемых в судебные инстанции обвинительных заключений и обеспечить новый уровень обоснованности судебных решений за счет снижения количества ошибок при их вынесении.

Применение методов ИИ в уголовных делах должно сопровождаться особыми оговорками с тем, чтобы предотвратить дискриминацию на основе использования конфиденциальных данных и обеспечить гарантии справедливого судебного разбирательства для любого человека, вне зависимости от пола, возраста, гражданства и других характеристик.

Для систем ИИ, используемых в судебных и правоохранительных органах, особое значение имеет проведение их независимой внешней экспертизы на предмет исключения влияния на алгоритмы случайно или специально подобранных данных, которые могут привести к дискриминационным решениям^[3].

Принципы Хартии должны регулярно применяться, контролироваться и оцениваться государственными и частными субъектами в целях постоянного совершенствования судебной и правоохранительной практики.

В этом отношении Европейская Комиссия рекомендовала, чтобы соответствующие субъекты регулярно проводили мониторинг хода практической реализации принципов Хартии. Независимые органы, упомянутые в Хартии, должны располагать результатами мониторинга и своевременно, под контролем соответствующих органов, вносить при необходимости изменения в программы ИИ, и установленные нормы применения ИИ в законодательной и правоохранительной деятельности.

Пять принципов Этической Хартии использования искусственного интеллекта в судебной и правоохранительной системах:

принцип первый – уважения фундаментальных прав. Необходимо убедиться, что разработка и использование инструментов и услуг ИИ совместимы с основными правами личности;

принцип второй – недискриминации. Уже на стадии разработки алгоритмов ИИ и больших данных, используемых для машинного обучения, необходимо блокировать возможность любой дискриминации между людьми или группами людей, которая может появиться в результате необдуманного использования статистических методов анализа необработанных больших данных;

принцип третий – качества и безопасности. В отношении разработки алгоритмов и использования данных, анализируйте источники больших данных, их структуру и содержание, а также

необходимо использовать математические модели, разработанные на междисциплинарной основе, учитывающей не только прямые статистические корреляции, но и социальные, культурные, экономические и другие факторы, вызывающие эти корреляции;

принцип четвертый – прозрачности. В судебной системе и правоохранительных органах ИИ может использоваться только в том случае, если его реализация обеспечивает прозрачность, включая доступность и понятность исходных больших данных и применение лишь тех типов нейронных сетей, в отношении которых внешний аудит может выявить критерии распознавания тех или иных совокупностей^[4];

принцип пятый – пользовательского контроля. При применении ИИ в судебной и правоохранительной сферах исключить нормативно предписывающий подход и обеспечить, чтобы все пользователи системы были информированы участниками и эксплуатировали ИИ только в том случае, если могут разобраться в алгоритмах и способах применения ИИ. В судебной системе также как в военной сфере не могут использоваться ИИ-черные ящики. Судебные работники всегда должны понимать и знать, на чем основываются выводы ИИ, которые предлагаются им как автоматизированное экспертное мнение.

§ 1. Принцип первый

Обработка судебных решений и данных должна служить четким целям и полностью соответствовать Европейской Конвенции о правах человека и Конвенции о защите персональных данных. (Конвенция о защите физических лиц в отношении автоматической обработки персональных данных ETSN108 с изменениями и дополнениями, внесенными в протокол CETS223).

Когда инструменты ИИ используются для решения споров или в качестве инструмента помощи в принятии судебных решений, а также в правоохранительной деятельности, важно обеспечить, чтобы ИИ не подрывал гарантии права на доступ к судье, а также право на равенство сторон и уважения к состязательному судопроизводству. При использовании ИИ правоохранительными органами, при предоставлении материалов в суд у стороны защиты в рамках проведения справедливого судебного разбирательства должно быть право на получение обоснования, принятого с использованием ИИ решения вплоть до раскрытия используемых больших данных и применяемых алгоритмах их разработки.

Также необходимо обеспечить безусловное сохранение принципа верховенства закона и независимости судьи и процессе принятия решений в обстановке, когда ИИ становится инструментом, вовлеченным в судебный процесс на стадии сбора доказательств, либо в фазе самого судебного процесса. Во всех случаях на этапах разработки алгоритмов ИИ необходимо соблюдать этические принципы, гарантирующие равенство граждан перед законом. В случае несоблюдения этических принципов необходимо отказываться от пусть эффективных и совершенных, но не соответствующих этим принципам алгоритмов машинного обучения, нейронных сетей и других математических методов ИИ.

§ 2. Принцип второй

Учитывая внутренне присущую методам машинного обучения и нейронных сетей способность закреплять классификации или распознавание новых данных на основе ранее выявленных закономерностей, касающихся отдельных лиц или групп лиц, необходимо, чтобы государственные и частные заинтересованные стороны уже на стадии разработки программных средств ИИ для судебных и правоохранительных сфер заблокировали возможность дискриминации определенных групп, слоев и отдельных граждан в результате применения алгоритмов.

Особое внимание следует уделить как на этапе разработки, так и на этапе развертывания, проверке применяемых алгоритмов на избыточный учет при принятии решений таких факторов, как раса, этническое происхождение, социально-экономическое состояние, политические взгляды, социальные связи, членство в профсоюзах и т. п. Следует прямо запретить в алгоритмах ИИ, используемых в судебной и правоохранительной системах, использование генетических данных, а также биометрических данных, связанных со здоровьем, сексуальной жизнью и сексуальной ориентацией. В случае если такая дискриминация выявляется при проверке применения алгоритмов, необходимо вплоть до усовершенствования алгоритмов отказаться от использования ИИ^[5].

Следует поощрять использование машинного обучения и междисциплинарного научного анализа для выявления фактического положения дел. Если фактическое положение дел может быть охарактеризовано как дискриминационная ситуация, то ИИ целесообразно максимально использовать как аналитический инструмент, а не инструмент поддержки принятия решений.

§ 3. Принцип третий

Разработчики модели машинного обучения должны иметь возможность широко использовать опыт работников судебной и правоохранительной систем, включая судей, прокуроров, юристов, следователей, оперативных работников и т. п., а также специалистов в области социальных наук, включая экономистов, социологов, психологов и т. п.

Для разработки ИИ, используемых в судебной и правоохранительной сферах, необходимо формировать смешанные проектные команды, реализующие междисциплинарный подход. При распределении подрядов на производство ИИ для судебной и правоохранительной сфер обязательным требованием должна быть сертификация разработчиков на прохождение курсов по этике ИИ.

Эффективность ИИ в значительной степени предопределяется качеством данных, используемых при машинном обучении. ИИ, применяемый в судебной и правоохранительной системах, должен обучаться на данных исключительно из сертифицированных источников.

§ 4. Принцип четвертый

Необходимо соблюдать баланс между интеллектуальной собственностью разработчиков на методы машинного обучения, используемые нейронные сети и т. п. и необходимостью обеспечения прозрачности, беспристрастности и отсутствия дискриминации в программах и алгоритмах ИИ.

Необходимо создать систему внешнего аудита программных средств ИИ. Целесообразно установить порядок, согласно которому обязательным условием использования ИИ в судебных и правоохранительных сферах является прозрачность их алгоритмической базы и объяснимость принятия решений на всех уровнях с целью недопущения использования дискриминационных факторов при распознавании объектов.

Наиболее предпочтительным является вариант полной технической прозрачности, когда в ИИ для судебной и правоохранительной сфер используют открытый исходный код, доступную документацию и человечески интерпретируемые алгоритмы нейронных сетей. Система также должны быть объяснена простым и понятным языком, прежде всего, для описания того, как получены результаты, чтобы обвиняемые, адвокаты и другие участники состязательного судебного процесса могли понять, и в случае необходимости оспорить, принятые ИИ рекомендации для решения дела.

Целесообразно ввести государственную сертификацию производителей ИИ, используемых в судебной и правоохранительной сферах.

§ 5. Принцип пятый

Пользовательский контроль должен быть увеличен, а не ограничен, на основе инструментов и сервисов ИИ. Профессионалы систем правосудия и правоохранительной деятельности должны в любой момент иметь возможность пересматривать судебные решения, принятые с применением ИИ, а также контролировать данные, используемые ИИ на всех этапах расследования и судебного процесса. Пользователь должен быть проинформирован ясным и понятным языком о том, каким образом принято решения ИИ, насколько они обязательны и имеют ли они право оспорить эти решения.

Каждый участник судебного процесса должен иметь полное право на получение в случае необходимости любых данных, используемых ИИ, для вынесения того или иного решения. Соответственно, участник не только должен иметь право на получение данных, но и на использование услуги обработки этих данных другим сертифицированным ИИ, чтобы выяснить, насколько обоснованными являются используемые в судебном процессе решения рекомендации со стороны ИИ, обслуживающего, например, сторону обвинения.

Глава III
Алгоритмические инструменты оценки
рисков в системе уголовного
правосудия США [\[6\]](#)

§ 1. Общие положения

В начале 2019 года был опубликован **Отчет об алгоритмических инструментах оценки рисков в системе уголовного правосудия США**. Он написан сотрудниками *Партнерства по ИИ (PAI)*. Работа по этой теме была первоначально продиктована законопроектом 10 Сената штата Калифорния (SB 10), который предписывал приобретение и *использование статистических инструментов и инструментов оценки риска машинного обучения для решений по предварительному заключению*. Впоследствии эта работа расширилась для оценки использования такого программного обеспечения по всей территории Соединенных Штатов.

Партнерство по искусственному интеллекту (PAI) – это некоммерческая организация, созданная для изучения и формулирования передового опыта в области технологий ИИ, для улучшения понимания общественностью проблем ИИ и в качестве открытой платформы для обсуждения и взаимодействия в отношении ИИ и его влияния на людей и общество.

Деятельность PAI определяется в сотрудничестве с его коалицией, насчитывающей более 80 членов. В том числе групп гражданского общества, корпоративных разработчиков и пользователей ИИ, а также многочисленных академических исследовательских лабораторий ИИ. PAI стремится создать пространство для открытого разговора, развития лучших практик и координации технических исследований, чтобы гарантировать использование ИИ на благо человечества и общества. Важно отметить, что PAI является независимой организацией.

В указанном отчете задокументированы *серьезные недостатки инструментов оценки риска в системе уголовного правосудия США, особенно в контексте досудебных задержаний*, хотя многие из наших наблюдений также применимы к их использованию в других целях, таких как условное осуждение и вынесение приговора. Несколько юрисдикций уже приняли законодательство, обязывающее использовать эти инструменты, несмотря на многочисленные глубокие проблемы и ограничения. Собирая взгляды сообщества исследователей в области ИИ и машинного обучения, PAI наметил десять в основном

невыполненных требований, которые юрисдикциям следует взвесить и рассмотреть перед дальнейшим использованием инструментов оценки рисков в системе уголовного правосудия.

Использование инструментов оценки риска для принятия справедливых решений о свободе человека потребует решения глубоких этических, технических и статистических задач, включая обеспечение того, чтобы инструменты были спроектированы и созданы для смягчения предвзятости как на уровне модели, так и на уровне данных, и чтобы были установлены надлежащие протоколы, содействовать прозрачности и подотчетности. Инструменты, доступные в настоящее время и находящиеся на рассмотрении для широкого использования, страдают от нескольких из этих сбоев, как описано в данном документе.

Эти недостатки выявлены в ходе консультаций с нашими экспертами, а также с помощью обзора литературы по инструментам оценки рисков и общедоступным ресурсам, касающимся используемых в настоящее время инструментов. Исследование было ограничено в некоторых случаях тем, что большинство инструментов не предоставляют достаточно подробную информацию об их текущем использовании, чтобы оценить их по всем требованиям в этом отчете. Юрисдикции и компании, разрабатывающие эти инструменты, должны реализовать большую прозрачность в отношении используемых данных и алгоритмов для решения этой проблемы в будущих исследовательских проектах. Тем не менее, многие из проблем, изложенных в этом отчете, относятся к любой попытке использовать существующие данные уголовного правосудия для обучения статистических моделей или для создания эвристики для принятия решений о свободе людей.

Проблемы использования этих инструментов в целом делятся на три категории:

- опасения по поводу достоверности, точности и предвзятости в самих инструментах;
- проблемы с интерфейсом между инструментами и людьми, которые взаимодействуют с ними;
- вопросы управления, прозрачности и подотчетности.

Хотя использование этих инструментов частично мотивировано желанием смягчить существующую человеческую погрешность в системе уголовного правосудия, это серьезное недопонимание, чтобы рассматривать инструменты как объективные или нейтральные просто потому, что они основаны на данных. Хотя формулы и статистические модели обеспечивают некоторую степень согласованности и воспроизводимости, они все же разделяют или усиливают многие слабые стороны процесса принятия решений человеком. Решения относительно того, какие данные использовать, как обрабатывать недостающие данные, какие цели оптимизировать и какие пороговые значения устанавливать, имеют существенное влияние на точность, обоснованность и предвзятость этих инструментов и, в конечном счете, на жизнь и свободу людей, которых они пытаются оценить.

Помимо технических проблем, при внедрении таких инструментов необходимо учитывать и *вопросы взаимодействия человека с компьютером*. Интерфейс «человек-компьютер» в данном случае относится к тому, как люди собирают и подают информацию в инструменты и как люди интерпретируют и оценивают информацию, которую генерируют инструменты. Эти инструменты должны соответствовать высоким стандартам интерпретируемости и объяснимости, чтобы пользователи (включая судей, адвокатов и служащих, в частности) могли понимать, как выполняются прогнозы инструментов, и принимать обоснованные решения на основе этих прогнозов. Чтобы улучшить интерпретируемость, такие предсказания должны явно включать информацию, такую как полосы ошибок, чтобы выразить неопределенность, лежащую в основе их предсказаний. Кроме того, пользователи должны посещать тренинги, которые учат, как и когда использовать эти инструменты надлежащим образом.

Более того, в той степени, в которой такие системы используются для принятия решений, которые меняют жизнь, инструменты и те, кто ими управляет, должны соответствовать высоким стандартам прозрачности и подотчетности. Данные, используемые для обучения инструментов и самих инструментов, должны подвергаться независимой проверке сторонними исследователями, адвокатами и другими соответствующими заинтересованными сторонами. Инструменты также должны получать постоянную оценку, мониторинг

и аудиты, чтобы гарантировать, что они работают должным образом и соответствуют обоснованным целям уголовной политики.

В свете этих вопросов, как общий принцип, эти инструменты не должны использоваться в одиночку для принятия решений о задержании или продолжении содержания под стражей. Учитывая насущную проблему массового лишения свободы, возможно, было бы разумно использовать эти инструменты для облегчения автоматического предварительного заключения большего числа лиц, но их не следует использовать для автоматического задержания лиц без дополнительных (и своевременных) индивидуальных слушаний. Более того, любое использование этих инструментов должно учитывать проблемы предвзятости, взаимодействия человека с компьютером, прозрачности и подотчетности, изложенные в этом отчете.

Инструменты оценки риска – это статистические модели, используемые для прогнозирования вероятности конкретного будущего результата. Такие прогнозы осуществляются путем измерения взаимосвязи между характеристиками индивидуума (например, его демографической информацией, криминальной историей или ответами на психометрическую анкету) и объединения числовых представлений этих характеристик в оценку риска.

Системы скоринга обычно создаются с использованием статистических методов и эвристики, применяемой к данным, чтобы рассмотреть, как каждая особенность способствует прогнозированию конкретного результата (например, неявка в суде).

Эти оценки часто используются для присвоения лицам разных категорий риска.

Хотя они, как правило, намного проще, чем глубокие нейронные сети, используемые во многих современных системах ИИ, инструменты оценки риска в уголовном правосудии являются основными формами ИИ. Некоторые используют эвристические структуры для получения своих оценок, хотя большинство используют простые методы машинного обучения для обучения прогностических моделей из входных наборов данных. Как таковые, они представляют парадигматический пример потенциальных социальных и этических последствий автоматического принятия решений ИИ.

Использование инструментов оценки рисков в процессах уголовного правосудия быстро расширяется, и лица, принимающие решения, как на федеральном уровне, так и на уровне штатов, приняли законодательство, предписывающее их использование. Это в основном произошло в рамках реформы, которая связана с чрезвычайно высокими показателями тюремного заключения в Соединенных Штатах, которые несоразмерны уровню преступности и международным и историческим исходным показателям. Сторонники этих инструментов выступают за их потенциал для рационализации неэффективности, снижения затрат и обеспечения строгости и воспроизводимости для жизненно важных решений. Некоторые адвокаты надеются, что эти изменения будут означать сокращение ненужного содержания под стражей и предоставят более справедливые и менее карательные решения, чем система освобождения под залог или системы, где лица, принимающие решения, такие как судьи, имеют полное усмотрение.

Это критически важные цели государственной политики, но есть основания полагать, что эти взгляды могут быть слишком оптимистичными. Остаются серьезные и нерешенные проблемы с точностью, достоверностью и смещением, как в наборах данных, так и в статистических моделях, которые управляют этими инструментами. Более того, эти инструменты также часто создаются для того, чтобы отвечать на неправильные вопросы, используются в плохо продуманных условиях или не подлежат достаточной проверке, аудиту и проверке. Эти проблемы почти универсальны в исследовательском сообществе ИИ и во всем PAI, хотя мнения расходятся в том, могут ли они быть реально решены путем улучшения инструментов.

Подавляющее большинство экспертов PAI, с которыми проводились консультации, согласились с тем, что *нынешние инструменты оценки рисков не готовы к использованию при принятии решений о задержании или продолжении задержания обвиняемых по уголовным делам без использования индивидуального слушания*. Одна из целей исследования состоит в том, чтобы сформулировать причины такого почти единодушного взгляда на вкладчиков и помочь наладить диалог с политиками, рассматривающими вопрос об использовании этих инструментов. Члены PAI и более широкое сообщество ИИ, однако, не имеют единого мнения о том, можно ли когда-либо совершенствовать

инструменты статистической оценки риска, чтобы справедливо задерживать или продолжать задерживать кого-либо на основе их оценки риска без индивидуального слушания. Для некоторых из наших членов проблемы остаются структурными и процедурными, а также техническими. Независимо от расхождений во взглядах на эти конкретные вопросы, в этом отчете кратко изложены технические проблемы, проблемы взаимодействия человека с компьютером и проблемы управления, которые коллективно определены сообществом.

Некоторые противоречия по поводу инструментов оценки риска проистекают из различных базовых показателей, по которым оцениваются инструменты оценки риска. У политиков есть много возможных исходных условий, которые они могут использовать при принятии решения о приобретении и использовании этих инструментов, включая следующие вопросы:

А. Достигают ли инструменты оценки риска абсолютной справедливости? Это вряд ли будет достигнуто какой-либо системой или учреждением из-за серьезных ограничений в данных, а также нерешенных философских вопросов о справедливости;

В. Являются ли инструменты оценки риска настолько справедливыми, насколько они могут быть основаны на доступных наборах данных? Это может быть достижимо, но только в контексте (а) выбора конкретной меры справедливости и (б) использования наилучших доступных методов для смягчения социальных и статистических ошибок в данных. Однако на практике, учитывая ограничения в доступных данных, это часто приводит к игнорированию систематических ошибок в данных, которые трудно устранить;

С. Являются ли инструменты оценки риска улучшением по сравнению с текущими процессами и лицами, принимающими решения? Инструменты оценки риска могут быть сопоставлены с эффективностью процессов, учреждений и практик принятия решений людьми, существовавшими до их внедрения, или аналогичными системами в других юрисдикциях без инструментов оценки риска. Такие оценки могут быть основаны на измеримых целях (таких как лучшее прогнозирование явки в суде или рецидив) или на отсутствии предрасположенности к человеческим предубеждениям. В этом

смысле инструменты оценки риска могут не достигать определенного понятия справедливости, а быть сравнительно лучше, чем статус-кво;

Д. Являются ли инструменты оценки риска улучшением по сравнению с другими возможными реформами системы уголовного правосудия? Другие реформы могут быть направлены на достижение тех же целей (например, повышение общественной безопасности, снижение вреда содержания под стражей и снижение издержек и бремени судебного процесса) при более низких затратах, большей простоте реализации или без учета интересов гражданских прав.

Базовые показатели А и В полезны для фундаментальных исследований алгоритмической справедливости и для эмпирического анализа производительности существующих систем, но они обязательно дают неоднозначные результаты из-за существования оправданных, но несовместимых определений справедливости. Тем не менее, они могут обеспечить полезную основу для понимания философских, правовых и технических вопросов с помощью предлагаемых инструментов.

Базовая линия С является одной из широко распространенных точек зрения экспертов, работающих в космосе. Это может быть уместно для политиков и юрисдикций, покупающих инструменты в соответствии с законодательными полномочиями вне их контроля, или в ситуациях, когда политические ограничения означают, что базовый уровень D неприменим. Тем не менее, мы должны подчеркнуть, что во всех обсуждениях, проведенных РАІ, базовый уровень D широко рассматривался как более фундаментально правильный и подходящий как цель выработки политики и стандарт оценки для инструментов оценки риска. Поэтому законодательные органы и судебные органы должны применять базовый уровень D всякий раз, когда это возможно для них.

§ 2. Минимальные требования к ответственному использованию инструментов оценки риска в уголовном правосудии

Что такое точность?

Точность представляет производительность модели по сравнению с принятой базовой линией или предварительно определенным правильным ответом на основе доступного набора данных. Чаще всего некоторые данные, используемые для создания модели, будут зарезервированы для тестирования и настройки модели. Эти зарезервированные данные обеспечивают новые оценки, которые помогают производителям инструментов избежать «переоснащения» в процессе экспериментов.

Точность измерения включает оценку того, выполнила ли модель наилучшую возможную работу по прогнозированию на тестовых данных. Сказать, что модель предсказывает неточно, значит сказать, что она дает неправильный ответ в соответствии с данными, либо в конкретном случае, либо во многих случаях.

Поскольку точность сфокусирована на том, как инструмент работает с данными, зарезервированными из исходного набора данных, он не решает проблемы, которые могут подорвать разумность самого набора данных (обсуждается в разделе о достоверности). Действительно, поскольку точность рассчитывается относительно принятого базового уровня правильности, точность не учитывает, являются ли данные, использованные для проверки или проверки модели, неопределенными или оспариваемыми. Такие вопросы обычно принимаются во внимание при анализе достоверности. Несмотря на то, что точность часто находится в центре внимания создателей инструментов при оценке эффективности их моделей, обоснованность и предвзятость часто являются более актуальными проблемами в контексте использования таких инструментов в системе уголовного правосудия.

Что такое валидность?

Узкое внимание к точности может привести к тому, что лица, принимающие решения, будут слепы к важным фактическим

соображениям, связанным с использованием инструментов прогнозирования. С любой статистической моделью, и особенно той, которая используется в таком критическом контексте, как оценки рисков в области уголовного правосудия, важно установить достоверность модели или ее точность в реальном мире. То есть, если оценки риска направлены на то, чтобы измерить вероятность того, что человек может не появиться или стать объектом будущего ареста, то это должно быть в том случае, если полученные оценки фактически отражают соответствующие вероятности. В отличие от точности, валидность принимает во внимание более широкий контекст, кающийся того, как данные были собраны и какой вывод делается.

Отдельно от данных и статистических проблем инструмент также может быть недействительным, потому что инструмент на самом деле не отвечает на правильный вопрос. Поскольку валидация всегда проводится в отношении конкретного контекста использования и конкретной задачи, перед которой ставится система, валидация инструмента в одном контексте мало говорит о том, является ли этот инструмент допустимым в другом контексте. Например, оценка риска может достаточно хорошо предсказать будущие аресты применительно к отдельным лицам в досудебном контексте, но довольно плохо применительно к лицам после вынесения приговора, или она может хорошо предсказать будущий арест в одной юрисдикции, но не в другой. Точно так же разные модели, построенные на основе одних и тех же данных, созданные с использованием разных решений и допущений при моделировании, могут иметь разные уровни достоверности. Таким образом, различные виды предсказаний (например, неявка, бегство, рецидивизм, насильственный рецидивизм) в разных контекстах требуют отдельной проверки. Без такой проверки даже устоявшиеся методы могут давать ошибочные прогнозы. Другими словами, только потому, что инструмент использует данные, собранные из реального мира, не делает автоматически его выводы достоверными.

Основные проблемы с использованием данных на уровне группы для оценки отдельных лиц

Фундаментальный философско-правовой вопрос заключается в том, допустимо ли определять свободу личности на основе данных о

других в их группе. В сообществах ИИ прогнозирование отдельных лиц на основе данных на уровне групп известно как этическая ошибка. Хотя инструменты оценки риска используют данные о человеке в качестве входных данных, связь между этими входными данными и прогнозируемым результатом определяется шаблонами в обучающих данных о поведении других людей.

В контексте вынесения приговора подсудимые имеют конституционное право определять приговор на основании того, что они сделали сами, а не того, что сделали другие, имеющие сходство с ними. Эта обеспокоенность возникла, например, в деле *Висконсин против Лумиса*, где суд запретил использование показателей риска в качестве решающего фактора при принятии решений о свободе, отметив, что «молодой преступник, безработный, имеет ранний арест на раннем возрасте и историю неудачи в надзоре. Это приведет к средней или высокой оценке по шкале риска насилия, даже если преступник никогда не совершал насильственного преступления»^[7].

Этическая ошибка особенно проблематична в системе уголовного правосудия, учитывая социальные предрассудки, которые отражаются в данных уголовного правосудия. Таким образом, вполне вероятно, что решения, принимаемые инструментами оценки риска, частично принимаются в зависимости от того, к какому защищенному классу может принадлежать физическое лицо, что вызывает серьезные опасения по поводу положения о равной защите.

В то время как существует статистическая литература о том, как решать технические проблемы, возникающие из-за этической ошибки, остается фундаментальный философский вопрос о том, допустимо ли задерживать людей на основании данных о других в их группе. По мере того, как все больше судов будут сталкиваться с вопросом, использовать ли инструменты оценки риска, этот вопрос должен быть в центре внимания и обсуждаться в качестве принципа первого порядка.

Что такое «смещение»?

В настройках статистического прогнозирования «смещение» имеет несколько перекрывающихся значений. Самое простое значение состоит в том, что прогноз, сделанный моделью, ошибочен в систематическом направлении – например, он предсказывает значение,

которое слишком низкое в среднем или слишком высокое в среднем для населения в целом. Однако в литературе по справедливости машинного обучения термин «смещение» используется для обозначения ситуаций, когда прогнозируемые вероятности систематически либо слишком высоки, либо слишком низки для конкретных подгрупп населения. Эти подгруппы могут быть определены переменными защищенного класса (раса, пол, возраст и т. д.). Или другими переменными, такими как социально-экономический класс. Мы будем в первую очередь использовать термин «предвзятость» в этом более узком смысле, который соответствует повседневному использованию термина, относящегося к разным суждениям о различных группах людей.

Смещение инструментов оценки риска может происходить из многих источников. Требование 1 (ниже) обсуждает смещение данных, которое вызвано несовершенным качеством данных, отсутствующими данными и смещением выборки. Требование 2 обсуждает смещение модели, которое проистекает пропущенной переменной смещения и прокси-переменных. Требование 3 обсуждает смещение модели, которое является результатом использования составных оценок, которые объединяют несколько различных предсказаний. В сочетании с озабоченностью по поводу точности и достоверности эти проблемы представляют собой серьезную проблему для использования инструментов оценки рисков в областях уголовного правосудия.

§ 3. Требование 1. Обучающие наборы данных должны измерять предполагаемые переменные

Наборы данных создают глубокие и нерешенные проблемы для достоверности статистических оценок риска. Почти во всех случаях ошибки и промахи в измерениях и выборках не позволяют легко доступным наборам данных уголовного правосудия отражать то, что они должны были измерить. Создание валидных инструментов оценки риска потребовало бы (а) методологии для переоценки и устранения ошибок в обучающих данных, используя вторые источники правды, и (б) способа определить, был ли этот процесс действительным и успешным. Насколько нам известно, в настоящее время с такими методами не создаются инструменты оценки риска.

Статистическая проверка предсказания рецидивизма, в частности, страдает от фундаментальной проблемы: основная правда о том, совершил ли человек преступление, обычно недоступна и может быть оценена только с помощью сообщений от лиц о преступлениях или аресты. Поскольку цель для прогнозирования (фактическое совершение преступления) недоступна, заманчиво изменить цель инструмента на прогнозирование ареста, а не преступления. Однако если цель использования этих инструментов состоит в том, чтобы предсказать риск обвиняемого для общественной безопасности – как и большинство инструментов оценки риска – цель должна заключаться в том, чтобы предсказать, может ли обвиняемый совершить преступление, оправдывающее предварительное заключение, а не в том, является ли обвиняемый виновным (вероятность быть арестованным или осужденным за какое-либо преступление в будущем).

Одна из проблем, связанных с использованием такой несовершенной информации, заключается в том, что представители различных демографических групп останавливаются, обыскиваются, арестовываются, обвиняются и неправомерно осуждаются с очень разными показателями удельного веса в нынешней системе уголовного правосудия США. Кроме того, о различных типах преступлений

сообщается и регистрируется с разной частотой, и частота сообщений может зависеть от демографических характеристик преступника и жертвы. Например, вполне вероятно, что все (или почти все) ограбления банков доводятся до сведения полиции. С другой стороны, аресты за хранение марихуаны общеизвестно предвзяты: чернокожие американцы гораздо чаще подвергаются арестам, чем белые, несмотря на схожие показатели использования. Таким образом, «данные об аресте, осуждении и лишении свободы наиболее подходящим образом рассматриваются как меры официального реагирования на преступное поведение», непропорционально распределены по определенным группам.

Оценка таких предубеждений может быть затруднена, хотя в некоторых случаях это возможно при использовании вторичных источников данных, собираемых отдельно от правоохранительных или правительственных учреждений. Например, данные об аресте или осуждении могут быть пересчитаны с использованием *Национального опроса о виктимизации преступности*, который предоставляет второй метод оценки демографических характеристик для типов преступлений, когда есть жертва, которая может видеть преступника, или опросы, которые отражают данные о совершении преступлений и арестах, такие как *Национальные обследования молодежи*. Выполнение такой переоценки будет тонкой статистической задачей, которую можно выполнить неправильно. Поэтому важным компонентом будет разработка метода, принятого сообществами машинного обучения и статистических исследований, для *определения того, дала ли переоценка данных достоверные результаты, которые точно отражают мир, реальную ситуацию*.

Помимо трудностей в измерении определенных результатов необходимы также данные, чтобы правильно различать разные причины одного и того же результата. Например, если просто посмотреть на результат неявки в суд, то обнаружим, что существует множество возможных причин такого исхода. Принимая во внимание, что существуют законные причины неявки в суд, которые не предполагают, что люди представляют опасность для общества (например, чрезвычайная ситуация в семье или ограниченные возможности транспортировки), объединение воедино всех лиц, которые не явились в суд, несправедливо увеличило бы вероятность

того, что лица, которые, как правило, имеют более законные основания для того, чтобы не явиться в суд (например, люди с иждивенцами или ограниченными возможностями транспортировки), были бы несправедливо задержаны. Таким образом, если целью инструмента оценки риска является прогнозирование того, будет ли обвиняемый избегать правосудия, необходимо будет собрать данные, которые будут различать лиц, которые намеренно и непреднамеренно не появляются в суде.

Учитывая, что валидность часто зависит от локального контекста, чтобы гарантировать полезность инструмента, где это возможно, данные, обсуждаемые выше, должны собираться в зависимости от юрисдикции, чтобы выявить значительные различия в географии, транспортировке и местных процедурах, которые влияют на эти результаты.

§ 4. Требование 2. Смещение в статистических моделях должно быть измерено и смягчено

Существует два широко распространенных заблуждения о предвзятости в системах статистического прогнозирования. Во-первых, модели будут отражать смещение только в том случае, если данные, с которыми они обучались, были неточными или неполными. Во-вторых, прогнозы можно сделать беспристрастными, избегая использования переменных, указывающих расу, пол или другие защищенные классы. Обе эти интуиции неверны на техническом уровне.

Возможно, это нелогично, но в сложных условиях, таких как уголовное правосудие, практически все статистические прогнозы будут смещены, даже если данные были точными, и даже если исключены такие переменные, как раса, если не предприняты конкретные шаги для измерения и смягчения смещения. Причиной является проблема, известная как *опущенная переменная смещения*. Опущенная переменная смещения возникает всякий раз, когда модель обучается на основе данных, которые не включают в себя все соответствующие причинные факторы. Отсутствующие причины исходной переменной, которые также вызывают интересующую входную переменную, называются смешанными переменными. Кроме того, включенные переменные могут быть прокси для защищенных переменных, таких как раса.

Частое вождение на вечеринки является смешанной переменной, потому что это вызывает как ночное вождение, так и риск несчастного случая. Модель, обученная на данных о времени суток, когда водители водят машину, будет демонстрировать предвзятость по отношению к людям, работающим в ночные смены, потому что это сопоставит риск вождения для сторон с риском вождения ночью.

На диаграмме также указаны переменные посредники на работе: частота вождения в ночное время является *прокси (удаленный) – сигналом*, когда вы едете на вечеринки, для вождения в нетрезвом состоянии. Это также прямой прокси для рабочих ночных смен. В результате, хотя нецелесообразно взимать с кого-то более высокие

страховые взносы просто потому, что они работают в ночные смены, это является результатом в этом случае из-за включения прокси-переменной частоты движения в ночное время.

Подобные сети доверенных лиц применимы к оценкам преступного риска, исходя из наблюдаемых входных переменных, таких как вопросы опроса, спрашивающие «Сколько ваших друзей/ знакомых когда-либо арестовывали?» Или «В вашем районе некоторые из ваших друзей или членов семьи были жертвами преступления?», которые являются прокси для гонки. Таким образом, трудно отделить использование инструментов оценки риска от использования факторов, защищенных конституцией, таких как раса, чтобы делать прогнозы, и необходимы меры по смягчению этого смещения на уровне модели.

Методы смягчения предвзятости

Существует множество возможных статистических методов, которые пытаются исправить смещение в инструментах оценки риска. Правильный метод применения будет зависеть оттого, что означает, что инструмент является «справедливым» в конкретном приложении, так что это не только технический вопрос, но и вопрос права, политики и этики. Хотя не существует единого универсального решения проблемы предвзятости, ниже приведены некоторые из возможных подходов, которые могут быть уместны в контексте прогнозов оценки риска в США.

Первый подход состоит в том, чтобы разработать модель для удовлетворения требования «равных возможностей», что означает, что ложноположительные показатели сбалансированы по некоторому набору защищенных классов (в контексте рецидивизма они будут означать вероятность того, что кто-то, кто прогнозирует отсутствие рецидива, неверно сделает этот прогноз). Неравные ложноположительные показатели особенно проблематичны в системе уголовного правосудия, поскольку они подразумевают, что лица, которые не часто совершают рецидив в одной демографической группе, чаще незаконно задерживаются, чем лица, не являющиеся рецидивистами в другой демографической группе (группах). Одним из предостережений к этому подходу является то, чтобы защищенные группы имели идентичные показатели. Таким образом, если

используется коррекция равных возможностей, то необходимо оценить различия в общей точности.

Второй подход заключается в установлении приоритетов при создании моделей, в которых прогностическое соотношение баллов одинаково для разных демографических групп. Это свойство известно как «калибровка в группах» и имеет преимущество, заключающееся в том, что оценки становятся более понятными для всех групп. Калибровка внутри групп повлечет, например, то, что люди с показателем 60 %, имеют 60 % шанс повторного повторения, независимо от их демографической группы. Проблема с этим подходом состоит в том, что обеспечение прогнозируемого паритета достигается за счет меры равных возможностей, описанной выше. Например, инструмент COMPAS, который оптимизирован для калибровки внутри групп, подвергся критике за его несопоставимые ложноположительные показатели (алгоритм COMPAS к 2019 году использовался в 10 штатах США для определения меры пресечения преступников). Фактически ProPublica обнаружила, что даже при контроле за предыдущими преступлениями, будущим рецидивом, возрастом и полом, черные обвиняемые имели на 77 процентов больше шансов получить более высокие оценки риска, чем белые обвиняемые. Это указывает на то, что откалиброванные в группе инструменты оценки риска могут по-разному влиять на людей, не совершающих рецидивы, в зависимости от их расы.

Третий подход включает использование методов причинно-следственных связей для формализации допустимых и недопустимых причинно-следственных связей между переменными и прогнозирования с использованием только допустимых путей. Преимущество этого подхода заключается в том, что он формально учитывает разницу между корреляцией и причинно-следственной связью и разъясняет причинные предположения, лежащие в основе модели. Он также удаляет только корреляцию с защищенным классом, которая возникает из-за проблемных связей между переменными, сохраняя больше информации из данных. Недостатком этого подхода является то, что он требует от создателя инструмента хорошего понимания причинно-следственных связей между соответствующими переменными, поэтому для создания действительной причинно-следственной модели необходим дополнительный предметный опыт.

Кроме того, инструментальщик должен определить, какие причинно-следственные связи являются проблематичными, и которые не являются. Таким образом, действительность также зависит от правильного суждения создателя инструмента.

Учитывая, что некоторые из этих подходов находятся в напряжении друг с другом, невозможно одновременно оптимизировать их все. Тем не менее, эти подходы могут выделить важные вопросы справедливости, которые следует учитывать при оценке инструментов. Например, даже при том, что, как правило, невозможно одновременно выполнить калибровку внутри групп и обеспечить равные возможности (методы № 1 и № 2 выше) с данными уголовного правосудия, было бы разумно избегать использования инструментов, которые либо имеют чрезвычайно несопоставимый прогностический паритет в демографических данных (т. е. плохая калибровка внутри групп) или крайне разрозненные ложноположительные показатели по демографии (т. е. низкий уровень равных возможностей).

Принимая во внимание, что каждый из этих подходов включает в себя компромиссные решения, также целесообразно использовать несколько различных методов и сравнивать результаты между ними. Это дало бы ряд предсказаний, которые могли бы лучше информировать принятие решений. Кроме того, подходящие пути для рассмотрения включают в себя своевременные, должным образом обеспеченные ресурсами индивидуализированные слушания, а не инструменты машинного обучения, разработку анализа затрат и выгод, придающего явное значение предотвращению разрозненного воздействия, или задержку развертывания инструмента до появления дополнительных рядов высококачественных данных, которые могут быть собраны, чтобы облегчить более точные и менее предвзятые прогнозы.

§ 5. Требование 3. Инструменты не должны объединять несколько разных предсказаний

Инструменты оценки риска не должны давать составные оценки, которые объединяют прогнозы различных результатов, для которых подходят различные вмешательства. Другими словами, инструмент должен прогнозировать конкретный риск, который он рассчитывает измерить, и производить отдельные оценки для каждого типа риска (в отличие от единой оценки риска, отражающей риск множественных результатов). Например, инструменты оценки риска не должны связывать риск неявки ответчика на назначенную дату суда с риском повторного задержания. Однако многие существующие инструменты оценки риска до суда делают именно это: они выдают единую оценку риска, которая представляет риск неявки или повторного возникновения. В некоторых случаях это может нарушать местное законодательство. Во многих юрисдикциях в качестве основания для предварительного заключения под стражу допускается только одна причина. И независимо от правовой ситуации *гибридный прогноз неуместен по статистическим соображениям*.

Различные причинные механизмы управляют каждым из явлений, которые объединены в гибридных оценках риска. Причины того, что кто-то не явится в суд, не будет повторно арестован и/или не будет осужден за будущее преступление, очень различны, поэтому высокий балл не будет легко интерпретируемым и объединит людей, которые, вероятно, имеют менее опасный исход (неявка в суд) с более опасными исходами (осуждение за будущее преступление). Кроме того, с точки зрения статистической достоверности прошлые обвинительные приговоры за ненасильственные преступления, которые с тех пор были декриминализованы (например, хранение марихуаны во многих штатах), возможно, следует рассматривать иначе, чем другие виды осуждений, если цель состоит в том, чтобы предсказать будущее преступление.

Более того, для каждого из этих различных явлений подходят разные виды вмешательства. Инструменты оценки риска должны использоваться только в том конкретном контексте, для которого они

предназначены, в том числе на конкретной стадии уголовного процесса и для конкретной группы населения, для которой они должны были предсказать риск. Например, потенциальный риск не явиться к дате суда на досудебной стадии не должен иметь никакого отношения к слушанию приговора. Аналогично, прогнозирование рисков для определенных групп населения, таких как несовершеннолетние, отличается от прогнозирования рисков для населения в целом.

Инструменты оценки риска должны иметь четкое представление о том, какое из этих многочисленных предсказаний они делают, и должны быть предприняты шаги для защиты от объединения разных прогнозов и использования оценок риска в неподходящих контекстах.

Проблемы интерфейса человек-компьютер

В то время как инструменты оценки риска предоставляют информацию и рекомендации для процессов принятия решений, окончательный субъект по принятию решений все еще находится в руках людей. Судьи, судебные клерки, сотрудники досудебных служб, сотрудники службы пробации и прокуроры могут использовать оценки рисков для определения своих решений. Таким образом, критические проблемы взаимодействия человека с компьютером также должны учитываться при рассмотрении вопроса об использовании инструментов оценки риска.

Одной из ключевых проблем инструментов принятия статистических решений является явление *предвзятости автоматизации*, когда информация, представляемая машиной, рассматривается как надежная по своей сути и выше скептицизма. Это может привести к тому, что люди будут чрезмерно полагаться на точность или правильность автоматизированных систем. В деле Висконсин против Лумиса косвенным образом решалась проблема предвзятости автоматизации, требуя, чтобы любой отчет о расследовании присутствия, содержащий оценку риска COMPAS, сопровождался письменным заявлением об отказе в том, что оценки могут быть неточными и были показаны для разносторонней классификации правонарушителей. Хотя раскрытие информации об ограничениях инструментов оценки риска является важным первым шагом, его все же недостаточно. Со временем существует риск того,

что судьи приучатся к длинному языку раскрытия, повторяемому в начале каждого доклада. Кроме того, раскрытия не дают четкого представления о том, как, если судьи вообще должны, интерпретировать или понимать практические пределы оценок риска.

В этом разделе предпринята попытка проиллюстрировать, как защититься от предвзятости автоматизации и других критических проблем взаимодействия человека с компьютером, обеспечив, чтобы (1) инструменты оценки риска легко интерпретировались пользователями-пользователями, (2) пользователи инструментов оценки риска получают информацию о неопределенности, стоящей за инструментами, прогнозы и (3) адекватные ресурсы выделяются для финансирования надлежащего обучения использованию этих инструментов.

§ 6. Требование 4. Предсказания и способы их составления должны легко интерпретироваться

Хотя адвокаты сосредоточили свое внимание на упомянутых выше проблемах предвзятости в оценках прогноза риска, один из часто пропускаемых аспектов справедливости – это способ перевода оценок риска для пользователей-людей. Разработчики и юрисдикции, внедряющие инструменты оценки риска, должны гарантировать, что инструменты передают свои прогнозы простым для пользователей способом и иллюстрируют, как эти прогнозы делаются. Это означает, что интерфейсы, представленные судьям, секретарям, юристам и ответчикам, должны быть четкими, легко понятными и не вводить в заблуждение.

Интерпретируемость предполагает предоставление пользователям понимания взаимосвязи между входными функциями и выходными прогнозами. Мы должны предостеречь, что это может означать не ограничение модели «интерпретируемой», но менее точной математической формой, но вместо этого использование методов, которые обеспечивают отдельные интерпретации для более сложных предсказаний.

Предоставление интерпретаций для прогнозов может помочь пользователям понять, как каждая переменная влияет на прогноз и насколько чувствительна модель к определенным переменным. Это крайне важно для обеспечения того, чтобы лица, принимающие решения, были последовательны в своем понимании того, как работают модели и производимые ими прогнозы, и чтобы неправильное толкование баллов отдельными судьями не приводило к несоразмерному применению правосудия. Поскольку интерпретируемость является свойством инструментов, используемых людьми, она требует рассмотрения использования оценок риска в контексте и зависит от того, насколько эффективно они могут быть использованы в качестве инструментов пользователями.

В то же время разработчики моделей должны следить за тем, чтобы интуитивная интерпретация не противоречила предполагаемому

прогнозированию риска. Например, судьи или другие пользователи могут интуитивно предположить, что упорядоченные категории имеют одинаковый размер, представляют абсолютные уровни риска, а не относительные оценки, и охватывают весь спектр приблизительных уровней риска. Таким образом, по 5-балльной шкале естественная интерпретация будет состоять в том, что один балл подразумевает риск повторного совершения преступления от 0 % до 20 % (или другой интересный результат), категория 2 – риск от 21 % до 40 % и т. д.

В одном исследовании сравнивался инструмент оценки риска до суда, который преобразует оценки риска в 5-балльную шкалу риска, с фактической вероятностью результата (в данном случае повторный арест, насильственный повторный арест, неявка и/или отзыв залога). Только 35 % обвиняемых, отнесенных к группе с самым высоким уровнем риска, не явились на суд или были повторно арестованы до суда. Вероятности неявки и повторной оценки для всех уровней риска были в пределах интуитивного интервала для самого низкого уровня риска. Аналогичным образом существуют также значительные пробелы между интуитивным и правильным толкованием категорий риска в инструменте предварительной оценки Колорадо. Чтобы устранить эти недостатки, юрисдикциям необходимо будет собирать данные и проводить дальнейшие исследования по выбору пользовательского интерфейса, отображению информации и психологическим реакциям пользователей на информацию о неопределенности прогноза.

§ 7. Требование 5. Инструменты должны давать достоверные оценки для своих прогнозов

Важным компонентом любого статистического прогноза является неопределенность, лежащая в его основе. Для того чтобы пользователи инструментов оценки риска могли правильно интерпретировать свои результаты, крайне важно, чтобы отчеты об их прогнозах включали перечни ошибок, доверительные интервалы или другие аналогичные признаки надежности. Например, инструменты оценки риска часто дают оценку, отражающую вероятность повторного совершения преступления, или сопоставление этих вероятностей с уровнями (например, «высокий», «средний» и «низкий» риск). Эта информация сама по себе, однако, не дает пользователю уверенности в уверенности модели в своем прогнозе. Например, даже если модель откалибрована таким образом, что результат, такой как «высокий риск», соответствует «60 % вероятности повторного совершения преступления», неясно, уверен ли инструмент в том, что ответчик имеет вероятность повторного совершения правонарушения между 55 % и 65 %, со средним значением 60 %, или если инструмент уверен только в том, что у ответчика есть вероятность повторного совершения правонарушения между 50 % и 90 %, при среднем значении 60 %. В первом случае толкование того, что ответчик имеет 60-процентную вероятность повторного совершения правонарушения, является гораздо более разумным, чем в последнем случае, когда вокруг прогноза существует огромная неопределенность.

По этой причине инструменты оценки риска не должны использоваться, если они не способны обеспечить хорошие показатели достоверности их собственных прогнозов, как в целом, так и для конкретных лиц, на которых они используются. Существует много источников неопределенности в прогнозах рецидива преступлений, и в идеале раскрытие неопределенности в прогнозах должно охватывать как можно больше таких источников. Это включает в себя следующее:

- неопределенность из-за размера выборки и наличия выбросов в наборах данных. Этот тип неопределенности может быть измерен с

использованием начальных доверительных интервалов, которые обычно используются технологическими компаниями для оценки предсказательной силы моделей перед развертыванием;

- неопределенность относительно наиболее подходящего смягчения для смещения модели, как обсуждалось в Требовании 2. Одной из возможностей будет оценка результатов различных корректировок справедливости как выражения верхних и нижних границ возможных «справедливых» прогнозов;

- неопределенность в результате смещения выборки и других фундаментальных проблем с наборами данных, как обсуждалось в Требовании 1. Это сложный вопрос для решения, но одним из способов решения этой проблемы будет поиск или сбор новых высококачественных вторичных источников данных для оценки неопределенности из-за смещения выборки и других проблем с обучающими наборами данных;

пользовательские интерфейсы для удовлетворительного отображения и передачи неопределенности пользователям в некоторых отношениях также являются открытой проблемой, поэтому учебные курсы, которые мы предлагаем в Требовании 6, должны специально тестировать и помогать пользователям выносить суждения при моделировании этой неопределенности.

§ 8. Требование 6. Пользователи инструментов оценки риска должны посещать тренинги о характере и ограничениях инструментов

Независимо от того, как результаты оценки риска объясняются или представляются, сотрудники служб предварительной оценки должны быть обучены тому, как правильно кодировать данные о лицах в систему. Человеческая ошибка и отсутствие стандартизированных передовых методов для ввода данных могут иметь серьезные последствия для качества данных и достоверности прогнозирования рисков в будущем.

В то же время судьи, адвокаты и другие соответствующие заинтересованные стороны также должны пройти тщательное обучение тому, как интерпретировать полученные оценки рисков. Для того чтобы любой такой инструмент использовался надлежащим образом, судьи, адвокаты и работники судов должны проходить регулярное обучение для понимания функции самого инструмента и того, как интерпретировать классификации рисков, такие как количественные оценки или более качественные оценки «низкий / средний/ высокий». Эти тренинги должны касаться значительных ограничений оценки, частоты ошибок, интерпретации баллов и того, как оспаривать или апеллировать к классификации рисков. Скорее всего, оно должно включать в себя базовую подготовку по пониманию доверительных интервалов. Требуются дополнительные исследования о том, как эти инструменты оценки риска основываются на человеческих решениях, чтобы определить, какие формы обучения будут поддерживать принципиальное и обоснованное применение этих инструментов и где существуют пробелы в текущей практике.

Управление, прозрачность и подотчетность

Поскольку инструменты оценки рисков дополняют судебные процессы и представляют собой реализацию решений местной политики, юрисдикции должны нести ответственность за их управление. Важно отметить, что они должны оставаться прозрачными

для граждан и подотчетными процессу выработки политики. Такое управление требует

(1) заинтересованности и широкого участия общественности в разработке и контроле таких систем;

(2) прозрачность данных и методов, используемых для создания этих инструментов; (3) раскрытие соответствующей информации ответчикам, чтобы позволить им оспаривать решения, основанные на этих инструментах; и (iv) предварительное развертывание и текущая оценка валидности, соответствия цели и роли инструмента в рамках более широкой системы правосудия

§ 9. Требование 7. Директивные органы должны обеспечить, чтобы цели государственной политики должным образом отражались в этих инструментах

Использование инструментов оценки риска может скрыть и скрыть от общественности основные политические решения, касающиеся уголовного правосудия. Они включают выбор того, в какой степени общественный риск перевешивает значительный вред задержания для обвиняемого и его семьи, и насколько определенным должен быть риск, прежде чем система уголовного правосудия будет действовать в отношении него (т. е. насколько точным, обоснованным и непредвзятым прогноз должен быть до того, как на него можно положиться, чтобы лишить человека свободы). Использование этих инструментов также включает выбор характера и определения защищаемых категорий и того, как они используются. Кроме того, должны быть приняты важные решения о том, как такие инструменты взаимодействуют с мерами, направленными на реабилитацию без лишения свободы, такими как меры отвлечения или предоставления социальных услуг.

Одним из ключевых примеров того, как, казалось бы, технические решения на самом деле являются политическими решениями, является выбор порогов для содержания под стражей. Например, в калифорнийском законодательстве SB 10 будет создана группа для установления пороговых значений, которые определяют вероятностный риск как «низкий», «средний» или «высокий» в случае неявки в суд или совершения другого преступления, представляющего риск для общественной безопасности. Между тем *Закон о первом шаге* (принят по инициативе президента США Д.Трампа в 2018 году – о тюремной реформе в США) требует, чтобы Генеральный прокурор разработал систему оценки рисков для классификации заключенных как имеющих минимальный, низкий, средний или высокий риск совершения другого преступления в будущем. Выбор этих порогов в конечном итоге определит, сколько людей задержано и освобождено.

Пороговые значения риска, подобные тем, которые предусмотрены Законом о первом шаге, являются политическим выбором, который необходимо выбирать в отношении более широкого процесса уголовного правосудия, конкретных целей политики в области уголовного правосудия и соответствующих данных для информирования этих целей. Политики, как на уровне штата, так и на федеральном уровне должны решить, какие компромиссы следует сделать, чтобы обеспечить справедливые результаты и снизить социальные издержки содержания под стражей. Например, если основной целью является сокращение массового лишения свободы в системе уголовного правосудия, пороговые значения должны быть установлены таким образом, чтобы число лиц, отнесенных к категориям повышенного риска, уменьшилось. Помимо сбора информации от соответствующих заинтересованных сторон, органам, устанавливающим пороги (будь то законодательные органы, комиссии или другие учреждения), следует практиковать разработку политики на основе фактических данных, основанную на актуальных и своевременных данных о преступности.

§ 10. Требование 8. Конструкция инструмента, архитектура и данные обучения должны быть открыты для исследований, анализа и критики

Инструменты оценки риска воплощают важные решения государственной политики, принимаемые правительствами, и должны быть такими же открытыми и прозрачными, как и любой закон, постановление или судебная норма. Таким образом, правительства не должны использовать какие-либо собственные оценки рисков, которые основаны на требованиях коммерческой тайны для предотвращения прозрачности.

В частности, учебные наборы данных, архитектуры, алгоритмы и модели всех рассматриваемых инструментов для развертывания должны быть широко доступны для всех заинтересованных исследовательских сообществ, таких как статистические, компьютерные, общественные, государственные, юридические и криминологические, чтобы они могли оценить их до и после развертывания.

Большая часть технической исследовательской литературы о справедливости, появившейся в последние два года (2017–2018 гг.), была результатом новаторской работы Pro Publica по публикации единого набора данных, связанного с инструментом оценки рисков Northpointe COMPAS, который был получен по запросам публичных записей в округе Бровард, штат Флорида. Публикация таких наборов данных позволяет проводить независимые исследования и публичные дискуссии, необходимые для оценки их эффективности.

Однако важно отметить, что когда такие наборы данных используются совместно, следует использовать соответствующие методы деидентификации, чтобы гарантировать, что непубличная личная информация не может быть получена из наборов данных. Принимая во внимание все более изощренные методы триангуляции (методы вычисления местонахождения в социальном мире), могут потребоваться дополнительные меры, такие как договорные условия,

когда получатели используют данные только для конкретных целей, и что после достижения этих целей они удаляют свою копию набора данных.

§ 11. Требование 9. Инструменты должны поддерживать сохранение и воспроизводимость данных, чтобы обеспечить оспаривание и вызовы

Чтобы обвиняемые могли оспаривать решения, принятые с помощью инструментов оценки рисков, они должны иметь доступ к информации о том, как делаются прогнозы инструментов. Как обсуждалось выше, существует много потенциальных технических проблем, связанных с использованием этих инструментов, в частности, в отношении предвзятости. Учитывая состязательный характер системы уголовного правосудия США, которая зависит от обвиняемых и их адвокатов, выдвигающих любые аргументы в их пользу, отказ в предоставлении ответчикам возможности доступа к информации о том, как принимаются эти решения, препятствует их способности оспаривать эти решения.

Информация индивидуального уровня, используемая в оценках, должна быть записана в контрольном журнале, который предоставляется ответчикам, адвокатам и судьям. Такие контрольные записи должны поддерживаться в неизменяемой форме для последующего анализа, чтобы аудиторы могли получить воспроизводимый расчет уровня риска для данного лица. Ответчики также должны иметь возможность оспаривать любые неточности во входной информации или выводы в результирующей классификации рисков и представлять дополнительную смягчающую информацию. Это особенно важно, учитывая возможность манипулирования инструментами оценки рисков. Например, оценки риска часто опираются на вопросники, предназначенные для арестованных, что дает возможность для злоупотреблений со стороны администраторов, что иллюстрируется случаями «подбора критериев». Сравнительный анализ, вероятно, является наилучшим способом защиты от таких манипуляций. Посредством этих процессов ответчики могут продемонстрировать, насколько применимы и надежны оценки риска в зависимости от их конкретных обстоятельств.

§ 12. Требование 10. Юрисдикции должны нести ответственность за оценку, мониторинг и аудит этих инструментов после развертывания

Юрисдикции должны периодически публиковать независимый обзор, оценку алгоритмического воздействия или аудит всех инструментов оценки риска, которые они используют для проверки того, что требования, перечисленные в этом отчете, были выполнены. Последующие аудиты должны будут регулярно проверять результаты и работу системы. Такие процессы проверки также должны быть локализованы, поскольку условия криминала, реагирования правоохранительных органов и культуры среди судей и служащих являются местными явлениями. В идеале эти процессы должны осуществляться при поддержке персонала и в судебных органах, а также при привлечении внешних экспертов.

Чтобы обеспечить прозрачность и подотчетность, *независимый внешний орган* (такой как совет по рассмотрению) должен быть ответственным за контроль аудита. Этот орган должен состоять из юридических, технических и статистических экспертов, в настоящее время и ранее заключенных лиц, государственных защитников, прокуроров, судей и организаций по гражданским правам, среди других. Эти проверки и их методология должны быть открыты для публичного рассмотрения и комментирования. Чтобы снизить риски для конфиденциальности, опубликованные версии этих аудитов должны быть отредактированы и достаточно ослеплены, чтобы предотвратить *деанонимизацию* (привязки анонима к конкретной личности).

В настоящее время проблемой для реализации этих проверок является нехватка данных, необходимых для оценки последствий использования этих инструментов. Когда некоторые партнеры РАІ пытались оценить последствия законодательного акта об оценке риска в штате Калифорния, они обнаружили неадекватные данные о количестве лиц, содержащихся под стражей до суда, в Калифорнии и не смогли определить данные или исследования, чтобы понять, как определяется определение низкого, среднего и высокого риска и их

пороговые значения. Естественно, это может повлиять на то, сколько человек задержано или освобождено до суда. Аналогичным образом, для оценки или исправления инструментов и данных для обучения на предмет ошибок и предубеждений требуются более точные данные о дискриминации в различных точках системы уголовного правосудия. Чтобы понять влияние существующих инструментов оценки риска, будь то досудебное разбирательство, вынесение приговора или испытательный срок, судебные системы должны собирать данные о досудебных решениях и результатах.

Для выполнения этих обязанностей всякий раз, когда законодательные органы или судебные органы принимают решение о предоставлении или приобретении инструментов оценки риска, эти органы должны одновременно обеспечить сбор данных после развертывания, предоставлять ресурсы для этого адекватным образом и поддерживать открытый анализ и анализ инструментов в развертывание. Это требует как (1) выделения или ассигнования достаточного финансирования для этих нужд, так и (2) институциональной приверженности набору (или заключению контрактов) со статистическими/техническими и криминологическими экспертами для обеспечения надлежащего сбора и анализа данных.

§ 13. Итоговые положения

Усилия по переводу системы уголовного правосудия США на разработку политики, основанной на фактических данных, и принятие решений, ориентированных на общественную безопасность, заслуживают похвалы и чрезвычайно важны. Исходя из исторических и международных сопоставлений, в США содержится аномально большое количество людей (в абсолютных цифрах, на душу населения и на уровень преступности). Таким образом, существенные реформы для решения этой проблемы являются обоснованными и безотлагательными на основе имеющихся данных. Этот контекст привел к принятию инструментов оценки риска, и крайне важно отметить, что ничто в этом отчете не должно толковаться как призыв к замедлению реформы уголовного правосудия и усилий по смягчению последствий массового лишения свободы.

Скорее, цель исследования – помочь политикам принимать обоснованные решения об инструментах оценки рисков, которые в настоящее время находятся в развертывании и требуются в соответствии с законодательными мандатами, и о возможных ответных мерах политики, которые они могли бы предпринять. Один из подходов заключается в том, что юрисдикции должны прекратить использование инструментов в решениях по задержанию лиц, пока не будет доказано, что они преодолевают многочисленные проблемы с действительностью, предвзятостью, прозрачностью, процедурой и управлением, которые в настоящее время стоят перед ними. Этот путь не должен замедлять общий процесс реформы уголовного правосудия. Фактически, несколько групп защиты интересов предложили альтернативные реформы, которые не представляют те же проблемы, что и инструменты оценки риска. Соответственно, выбор не просто между существующими системами, такими как залог наличными, и более новыми алгоритмическими системами.

Другой вариант – начать проект по улучшению инструментов оценки рисков. Это потребует закупки достаточно обширных и репрезентативных данных, разработки и оценки методов повторного взвешивания и обеспечения того, чтобы инструменты оценки риска были предметом открытых, независимых исследований и проверки.

Десять требований, изложенных в этом отчете, представляют собой минимальный стандарт для разработчиков и политиков, пытающихся согласовать свои инструменты оценки рисков и то, как они используются на практике, с обоснованными целями политики.

Хотя широкое использование оценок рисков продолжается, административные органы и законодательные органы, ответственные за развертывание, несут ответственность за установление стандартов для инструментов, которые они распространяют. В дополнение к десяти требованиям, которые мы изложили в этом отчете, юрисдикциям также потребуется собрать и использовать значительный опыт в областях машинного обучения, статистики, взаимодействия человека с компьютером, криминологии и права для выполнения этой задачи. На данном этапе мы должны подчеркнуть, что мы не считаем, что какие-либо существующие инструменты будут соответствовать должным образом установленным стандартам по всем этим пунктам, и в случае Требования 1 для выполнения надлежащим образом установленного стандарта потребуются значительные новые усилия по сбору данных.

РАІ считает, что установление стандартов в этой области является важной работой для политиков из-за огромного импульса, который законодательство придает развертыванию инструментов ИИ для оценки рисков. Но многие из наших членов по-прежнему обеспокоены тем, что стандарты могут быть установлены с целью их простого соблюдения, а не для того, чтобы на самом деле решать глубокие статистические и процедурные проблемы, присущие использованию таких инструментов для обоснования решений о задержании. Было бы заманчиво установить стандарты, которые затушевывают сложные проблемы точности, достоверности и систематической ошибки, и продолжать развертывание инструментов без учета альтернативных реформ.

Для исследователей ИИ задача предвидения и смягчения непредвиденных последствий и злонамеренного использования стала одной из центральных проблем нашей области. Для этого требуется очень осторожный подход к проектированию и проектированию систем, а также тщательное рассмотрение способов их возможного сбоя и вреда, который может возникнуть в результате. Уголовное правосудие является областью, в которой необходимо проявлять

максимальную осторожность и смирение при использовании статистических инструментов. Мы обеспокоены тем, что сторонники этих инструментов не смогли должным образом удовлетворить минимальные требования для ответственного использования до широкого распространения.

В дальнейшем мы надеемся, что этот отчет вызовет более глубокое обсуждение этих проблем с использованием инструментов оценки рисков и стимулирует сотрудничество между политиками, исследователями и группами гражданского общества для достижения столь необходимых стандартов и реформ в этой области.

Раздел IV
Основные направления правового
регулирувания ИИ^[8]

§ 1. Вводные положения

Исследование «European Tech Insights 2019», проведенное Университетом управления в Мадриде (IE)^[9], показало, что более двух третей европейцев всех возрастов считают, что новые технологии, если они не будут должным образом контролироваться, принесут больше вреда, чем пользы для общества в ближайшее десятилетие. Это опасение, в свою очередь, ведет к растущему сопротивлению инновациям и требованию ужесточить регулирование. Подавляющее большинство опрошенных европейцев ожидают, что их правительства введут новые законы и налоги, чтобы ограничить автоматизацию и предотвратить замещение рабочих мест, даже если это повлечет за собой остановку технического прогресса. Эти результаты согласуются по странам, возрастным группам, полам и идеологиям.

Исследование проводилось в 8 европейских странах – Франции, Германии, Ирландии, Италии, Испании, Нидерландах и Великобритании.

Большинство опрошенных граждан ЕС выступают за правительства, *продвигающие политические меры по ограничению автоматизации бизнеса компаниями.* Минимальные интервенционные меры по дополнительному налогообложению предприятий, которые сокращают рабочие места из-за автоматизации (67 %), предоставлению дополнительной поддержки тем, кто пострадал в результате потери рабочих мест (71 %), одобряются значительной частью населения, в то время как более решительные меры государственного вмешательства имеют еще более высокий уровень поддержки: 72 % опрошенных считают, что правительства должны установить ограничения на количество рабочих мест, которые могут быть заменены машинами на предприятиях. 74 % считают, что предприятиям следует разрешить замену только самых опасных или вредных для здоровья рабочих мест.

Подавляющее большинство европейцев ожидают, что правительства введут новые законы и налоги, чтобы ограничить автоматизацию и предотвратить замещение рабочих мест, даже если это приведет к замедлению экономического прогресса. Эти результаты согласуются по странам, возрастным группам, полам и, что особенно удивительно,

идеологиям. И все же, подобные меры в настоящее время находятся вне политической дискуссии.

Одновременно расследование выявило парадокс, заключающийся в том, что, хотя общественность опасается технологических достижений, особенно что касается повышения уровня автоматизации, **каждый четвертый европеец предпочитает, чтобы искусственный интеллект принимал важные решения по управлению их страной.** В таких странах, как Нидерланды, Германия и Великобритания, этот процент еще выше: каждый третий житель выступает за это.

Исследование также показывает, что люди не только обеспокоены предстоящим технологическим преобразованием, но и считают, что учреждения, на которые возложен этот процесс, не справляются со своими обязанностями. Большинство опрошенных (60 %) считают, что система образования не обучает их решать проблемы, возникающие в связи с новыми технологиями.

Экспертная группа ЕС по искусственному интеллекту (ИИ) констатировала в феврале 2019 года принципиальное отличие политики, инвестиций, правового регулирования и практического использования ИИ в ЕС по сравнению с Соединенными Штатами и Китаем.

В отличие от США ЕС в гораздо большей степени привержен тщательному предварительному изучению долговременных политических, социально-психологических и поведенческих аспектов внедрения новых решений в сфере ИИ. ЕС также принципиально отличается подходом от Китая, поскольку ставит на первый план при разработке и внедрении ИИ безусловное соблюдение прав личности, меньшинств и в целом гражданского общества. В этом плане развитие ИИ в странах ЕС – это своеобразный третий путь эволюции ИИ по отношению к США и Китаю^[10].

Этот подход в полной мере реализуется при реформировании Директивы об ответственности за качество ИИ и *Директивы по развитию машинной инфраструктуры ИИ*, а также совершенствовании программной базы. В качестве предварительного заявления *Целевая группа CEPS* констатирует, что применительно к подавляющему большинству программно-аппаратных решений в сфере ИИ она не видит необходимости внесения принципиально новых изменений в нормативные документы, регулирующие

требования к программно-аппаратной базе ИИ. По мнению Группы, подавляющая часть программно-аппаратных решений в сфере ИИ не требует внесения принципиальных изменений в законодательства, нормативные требования и стандарты эксплуатации. Причина в том, что *основная часть так называемых программно-аппаратных систем ИИ, используемых как в бизнесе, так и в сфере государственного управления, не представляют собой чего-то принципиально нового и отличного от ранее использованных пакетов статистического анализа больших данных.* Лишь незначительная часть программно-аппаратных комплексов, в основном находящаяся сейчас в сфере экспериментальных, исследовательских и опытно-конструкторских разработок, по своим возможностям заметно отличаются от привычной компьютерной техники и программного обеспечения. Соответственно, *главным направлением ИИ на сегодняшний день является не прокламируемая социальными медиа революция, а реальная эволюция, постепенное улучшение на основе больших данных, более мощных аппаратных средств и более разнообразных, чем раньше, статистических программ и программ по распознаванию.*

Комиссия ЕС еще до принятия соответствующих законодательных решений указала на то, как *ИИ может реально изменять поведение граждан, а также, как он может быть использован для социального инжиниринга недобросовестными акторами и организованной преступностью.*

При рассмотрении действующего законодательства, и что более важно, при формировании законодательной базы разработки и использования ИИ Европейская Комиссия учитывала следующее:

- *необходимость перейти от нарративной, описательной формы подготовительных материалов к данным, отражающим прогнозные перспективы и анализ рисков, связанных с использованием ИИ;*

необходимость оставаться открытой для новой формы «инновационных сделок», в рамках которых разработчики ИИ бросают вызов действующему регулированию, демонстрируя, что они могут достичь более значительных преимуществ и создать программноаппаратные средства решения задач, которые не ставились ранее. По сути, инновационная сделка означает, что ЕС должен создать возможность для исследователей в области ИИ работать вне действующего законодательства. *Должна быть создана своего рода*

«законодательная песочница», в которой получившие специальное разрешение компании могут работать вне общеустановленной нормативной базы. Подобный подход к инновационным сделкам позволит европейским исследователям оставаться на самых передовых позициях без ослабления контроля над применением ИИ в бизнесе и федеральных органах; с точки зрения методологии необходимо проанализировать разработку новых критериев для определения тем в сфере ИИ, которые могут финансироваться за счет общеевропейских ресурсов. В настоящее время распределение общеевропейских средств осуществляется в основном по критерию соотношения текущих затрат и будущих результатов. Подавляющее большинство экспертов в настоящее время полагают, что подобный критерий из-за неопределенности результатов не применим к ИИ. Видимо, нет иного пути, как осуществлять капиталовложения в области ИИ в зависимости от экспертных оценок наиболее уважаемых и оригинально мыслящих экспертов. Для этого наиболее подходят методики качественных сравнений (больше – меньше, лучше – хуже), чем количественных формальных критериев.

Подготовка нормативной документации по разработке и использованию ИИ в странах ЕС должна осуществляться обязательно не только с привлечением соответствующей комиссии, представителей федеральных органов, бизнеса, ведущих разработчиков, но и в той или иной форме конечных пользователей – представителей гражданского общества.

Крайне важно, чтобы любая форма саморегулирования оценивалась с особой тщательностью и поддавалась проверке, была прозрачной для государственных органов, гражданского общества и бизнеса.

Нормативное регулирование ИИ должно строиться на сочетании юридических норм и саморегулируемых схем. Эта двойственность связана с объектом регулирования. ИИ и в эволюционном, и тем более в революционном варианте – одном из наиболее динамичных технологических пакетов четвертой производственной революции. Право же всегда консервативно. Оно фиксирует нормы и неформальные правила, сложившиеся в прошлом. При невысоких темпах научно-технического прогресса научно-технические разработки можно регулировать с использованием юридических норм. Однако, в условиях экспоненциального роста и высокой динамики

исключительно юридическое регулирование будет сдерживать развитие технологий. Именно поэтому и необходим блок саморегулирования.

Саморегулируемые схемы призваны включать в себя, прежде всего, кодексы поведения разработчиков алгоритмов (например, принципы *Asilomar*), корпоративные правила конфиденциальности, требования к защите данных, протоколы обработки жалоб и контроля за обеспечением требований (например, протоколы, обеспечивающие право на забвение) и т. п. В большинстве случаев в рамках нормативно-саморегулируемой кодификации оптимальным является разделение, когда на стадии разработки в основном действует схема саморегуляции, а на стадии использования ИИ – юридического регулирования.

При этом важно, чтобы любая форма саморегулирования поддавалась оценке, проверке и была достаточно прозрачной для контроля государственными органами. Образцом для этого являются *RegTech* и *SupTech*. Что касается пользовательского применения ИИ, то здесь возможно сочетание юридических норм с регулированием «снизу вверх» (по инициативе гражданского общества). Вместо того чтобы полагаться исключительно на законодательное регулирование или саморегулирование со стороны бизнеса, применительно к этой сфере целесообразно юридическое регулирование совместить с протоколами и нормами, разработанными структурами гражданского общества и НКО, и использовать их в качестве отраслевых частных стандартов применения ИИ к данным граждан.

В настоящее время именно таким образом действует метарегуляторная схема, разработанная альянсом *ISEAL*. Эта схема устанавливает правила и критерии, которым должны следовать частные регуляторы в сфере оказания услуг, связанных с социальным развитием и экологией. В рамках частных стандартов определяются критерии подотчетности, прозрачности, недискриминации, требования к аудиту и сертификации. Все это, в конечном счете, приводит к повышению уровня доверия между бизнесом, федеральными органами власти и пользователями. Если применить этот порядок к ИИ, то для платформ, алгоритмы которых проверены и постоянно контролируются саморегулируемыми ассоциациями или гражданскими НКО, можно использовать специальный логотип,

ярлык, ясно дающий понять конечным пользователям, что именно эти платформы ИИ не несут им никаких угроз и рисков, поскольку подверглись тройной проверке со стороны государства, бизнеса и самих пользователей.

Перечисленные выше формы представляют собой наиболее эффективные алгоритмы делегирования определенных этапов правового регулирования частному сектору и конечным пользователям при обязательном и жестком контроле со стороны государства и в целом органов ЕС.

Наиболее эффективное соединение усилий юридического кодирования, саморегуляции и общественного контроля происходит в тех случаях, когда участниками регулирования совместно определены основные, всеобъемлющие принципы регулирующего вмешательства. В любом случае, базируясь на делегировании и общественном контроле, эти формы обеспечивают большую гибкость, адаптивность, а за счет этого – долговечность регулирования по сравнению с традиционным юридическим нормотворчеством. Как уже отмечалось, нормотворчество не только требует в каждом конкретном случае значительного времени на разработку кодифицирующих актов, но и, как любая правовая норма, устанавливает строгие стандарты на будущее, исходя из прошлого опыта. Чем быстрее и революционнее технологическое развитие, тем в меньшей степени, как само развитие, так и его прикладные результаты могут регулироваться исключительно на правовой основе, по принципу «сверху вниз».

Новый подход ЕС к регулированию новаций базируется именно на смешанном использовании юридических норм, саморегуляции и контроля со стороны гражданского общества. Например, недавно принятые законы о сетевом нейтралитете и о соблюдении авторских прав предполагают, что окончательные решения относительно провайдера интернет-услуг по соблюдению им законов принимаются на основе общественного контроля со стороны подписчиков услуг соответствующим нормативным органом.

В этом отличие европейского подхода от американского и китайского. *В американском подходе по большому счету все решает бизнес, а соответственно организации гражданского общества и группы потребителей зачастую протестуют против самоуправства корпораций.* В то же время в Китае все решает государство и

юридические акты не только созданы государством, но и преследуют исключительно его интересы, иногда в ущерб интересам потребителей. Реализуемый в ЕС подход позволяет балансировать интересы разработчиков, государства и пользователей, и развивает подход к регулированию интернета, который в международном масштабе базируется именно на сочетании глобального частного регулирования, странового государственного контроля и регионального или странового пользовательского подхода.

Для того чтобы системный подход к регулированию ИИ был результативен, его целесообразно разбить на фазы регулирующего вмешательства (например, на установление повестки дня, определение стандартов, разграничения сфер юридической кодификации, бизнес-саморегуляции и гражданского контроля, эффективную систему мониторинга и оценки, обеспечение соблюдения разграничения прав, обязанностей и функций сторон регулирования, и постоянную открытую отчетность об эффективности регулирования).

Принципиально возможны два типа осведомленности о регулировании: соответственно *ручное* и *алгоритмическое*. При *ручном* варианте специально выделенные аналитики с той или иной периодичностью рассматривают функционирование системы во всех ее блоках и фазах и подготавливают аналитические материалы, которые рассматривают эксперты – своеобразные судьи.

По состоянию на 2018 г. в Европе и Китае основной формой была именно *ручная*. В тоже время в Соединенных Штатах все активнее используется *алгоритмическая форма осведомленности*. Она предусматривает, что законодательно устанавливаются принципы, которым должны следовать онлайн-платформы при разработке и использовании алгоритмов. На основе этих правил сами платформы разрабатывают алгоритмы, а также дополняют правила путем разработки соответствующей бизнес-модели и формы отчетности перед государственными органами, позволяющими осуществлять мониторинг и оценку платформ. По сути, в Соединенных Штатах впервые в истории создан прецедент публичного законодательного регулирования принципов разработки алгоритмов. Поскольку именно платформенные решения^[11] в настоящее время являются основной сферой использования ИИ в бизнесе, можно полагать, что чем дальше, тем больше ручной принцип будет заменяться алгоритмическим.

Принципиальное различие между этими двумя типами регулирования состоит в следующем. Традиционное ручное регулирование предполагает предписывание обязательных норм людям, выполняющим те или иные работы в рамках бизнеса (например, законодательство нормативно запрещает продавцу обсчитывать покупателя или продавать ему некачественный товар. Норма применяется к действиям человека-продавца. – *Авторы*). В случае же алгоритмического регулирования устанавливаются обязательные для разработчиков и программистов правила разработки алгоритмов^[12].

Особая, совершенно неразработанная тема – это регулирование поведения потребителей. Поведенческая экономика появилась только в последние десятилетия. Соответственно ни в одной стране мира нет не только законов, но и саморегулирующих подходов к данной тематике. В конечном счете, в рамках использования ИИ и больших данных, поведение потребителя не только в интернете, но и в цифровой реальности может определяться не его волей, а выбором определенной архитектуры платформы и ее интерфейса. Фактически создатели платформ де факто определяют поведение потребителя вне зависимости от его желаний.

В странах ЕС в настоящее время ведутся острые дебаты о преимуществах и недостатках так называемого *Наджа* или *Подталкивания*. Его основоположник К.Санстейн утверждает, что *Надж* – это либертарианский патернализм. Т. е. при помощи алгоритма и интерфейса он подталкивает людей к осуществлению тех поступков, которые несут им пользу, а не вред. Осуществляется это, по его словам, на основе алгоритмов онлайн платформ, использующих опции по умолчанию, при помощи определенных цветов, за счет расположения элементов интерфейса и т. п., которые подталкивают потребителя к выбору, наиболее полезному для его здоровья, благополучия и т. п. Между тем, значительная часть социальных психологов и даже политологов полагают, что *Надж* может быть использован, возможно, даже с большей эффективностью, не для достижения положительных с точки зрения персоны или общества действий, а как манипулятивная технология, позволяющая воздействовать на население, по сути, принуждая граждан к

покупкам конкретных товаров и услуг или выбору определенного кандидата в электоральном процессе и т. п.

Особой темой, обсуждение которой вновь усилилось в рамках регулирования ИИ, является правовое регулирование дизайна платформ. Еще в 1999 г. Лоуренс Лесинг показал, что тот или иной дизайн платформы может работать как на потребителя, так и позволять собственникам или производителям платформы осуществлять действия, уголовно наказуемые в физическом мире, но безнаказанные в интернет-пространстве. В 2015 г. исследователи Гамбургского университета выявили, что из 100 обследованных европейских платных сервисов в 84 легкость платной подписки сочетается с трудностями, а иногда почти невозможностью отписаться от платной услуги. Поскольку применение ИИ, например, на торговых платформах для потребителей интернет-услуг означает, прежде всего, еще возможность еще меньше, чем сегодня напрягаться и думать, то следует ожидать увеличения рисков и опасностей изъятия средств за счет дизайн-решений. Если не применить к этой уловке жесткие регулятивные нормы, то уже в ближайшие годы дизайн-вымогательство, будучи уголовно ненаказуемой формой, будет все более и более распространяться, а ущерб, наносимый им кошелькам потребителей, исчисляться уже миллиардами евро.

Дизайн-вымогательство – это лишь один пример более мощного технологического течения. Суть его в следующем. Все новые и новые информационные технологии делают человека-потребителя или занятого в бизнесе клиента все более уязвимым перед теми или иными ухищрениями, ловушками и методами манипулирования. Чем дольше эти недружественные для потребителей решения будут оставаться в серой зоне, тем больше опасность, что киберпреступность будет уходить от уголовно наказуемых форм деятельности в ненаказуемые зоны, где, по сути, будут использоваться с умыслом манипулятивные методы и технологии, многократно умножаемые возможностями ИИ.

Решения RegTech и SupTech в настоящее время используются исключительно в сфере финансовых услуг. Однако их потенциал существенно шире. Эти технологические пакеты могут быть использованы в любом регулируемом секторе для проведения оценки рисков и угроз и создания систем непрерывного мониторинга. Помимо обязательной оцифровки отраслевой отчетности, они позволяют

осуществлять автоматизацию проверки технических регламентов, стандартов, нормативных требований на их реализуемость в практике работы компаний и на их соответствие согласованным на государственном уровне нормам. По сути, указанные технологии представляют собой эффективные инструменты использования ИИ для анализа нормативно-правовой документации и практики ее использования в работе любой компании, в первую очередь, осуществляющей свою деятельность в интернете.

Заслуживают широкого распространения и опробации адаптивные варианты регулирования (экспериментальное формирование нормативной политики). Скорость, с которой развиваются технологии, и разрушительный характер новых бизнес-моделей во многих секторах побудили регулирующие органы во всем мире искать экспериментальные подходы к регулированию. Эти подходы позволяют найти компромисс между естественным консерватизмом правовых норм и быстротечным техническим прогрессом.

Существует несколько возможных способов обеспечить высокую адаптивность регулирования. Они включают:

- использование «нормативных песочниц» и других экспериментальных подходов, применяемых лишь для ограниченного, а в отдельных случаях конкретно прописанного круга юридических и физических лиц, для обеспечения наработки решений, которые затем будут положены в основу саморегуляции и гражданского контроля, а главное, совершенствования законодательства на долгосрочный период;
- создание технологических дорожных карт и мнений хозяев и разработчиков платформ в процессе выработки законодательной политики в сфере интернет-бизнеса, ИИ и т. п.;
- широкое использование инструментария открытого правительства, позволяющего привлечь к выдвижению, разработке и анализу нормативных актов бизнес и гражданское общество, а также использовать их возможности для мониторинга эффективности реализации и выявления недостатков уже принятых юридических решений.

По-прежнему ядром регулирования информационных технологий вообще, и ИИ в частности, являются законодательные нормы, а также

отраслевые стандарты и технические требования. Вопрос технических требований и стандартов является одним из важнейших в части регулирования ИИ. Здесь полезно напомнить, что еще несколько лет назад в период работы регулирующих органов США и ЕС по разработке Трансатлантического торгово-инвестиционного партнерства, одним из главных камней преткновения стала несовместимость технических стандартов и нормативных требований к информационным технологиям вообще, и ИИ в частности между США и ЕС.

В Соединенных Штатах используются гораздо более гибкие стандарты, которые созданы в основном в рамках саморегулирования. В ЕС же технические требования – это прерогатива органов ЕС и государственных регулирующих органов. Как уже отмечалось, европейский подход позволяет избежать многих опасностей для конечных пользователей. Однако американский подход обеспечил заокеанскому бизнесу и университетам несомненное лидерство в сфере интернет-технологий вообще, и ИИ в особенности. В конечном счете, в Соединенных Штатах хорошо понимают, что IT отрасль была, есть и будет находиться в серой зоне, и соответственно ее правовое регулирование всегда будет носить догоняющий характер. Данное обстоятельство является технологической реальностью. В этой связи ЕС необходимо полностью учесть эту реальность, и не повторять ранее совершенных просчетов, сделанных применительно к ИИ, в отношении протоколов P2P и блокчейна.

Во всех случаях *анализ рисков новых технологий должен стать не просто неотъемлемой частью нормативной практики ЕС, но и приоритетным направлением ее деятельности.* Необходимо особо выделять наиболее динамичные факторы в развитии технологий, которые по причинам межблоковой и международной технологической конкуренции целесообразно оставлять в серых зонах. Европейская, как и китайская традиция, предусматривает, что вся деятельность должна быть юридически кодифицирована и подпадать под сферу действия законов. Это, безусловно, создает большую защищенность бизнес-потребителям и конечным пользователям из числа домохозяйств, но делает европейские разработки аутсайдерами по отношению к американским. Необходимо готовиться, также как готовятся США и страны Британского Содружества наций, прежде всего, Канада,

Австралия и Новая Зеландия, к постепенному переходу от подхода «закон как текст» с его толкованием людьми, к решениям «закон как код» и его исполнению техноплатформами.

Уже упомянутый Л.Лессинг еще в 1999 г. утверждал, что со временем «закон как текст» сменится «законом как кодом». При этом он справедливо полагал, что главное преимущество «закона как текста» состоит в его реализуемости через людей, а не через самоисполняемые программы. Самоисполняемая программа, по его мнению, не может предоставить уровня гибкости и нюансов, необходимого для адекватной интерпретации конкретных ситуаций. Поэтому, по мнению Л.Лессинга первоначально «закон как текст» и «закон как код» будет сосуществовать. Причем второй будет понемногу вытеснять первый из тех сфер, где трактовка закона однозначна, а ситуации не имеют нюансов и т. п. Полное же вытеснение «закона как текста» и человека-судьи «законом как кодом» и его программным самоисполнением произойдет только и если ИИ, не будучи так называемым общим или универсальным интеллектом, сможет за счет распознавания фактов учитывать большее число нюансов и более качественно распознавать сложные ситуации, чем судья-человек.

В части мониторинга и оценки регулятивной политики в области ИИ во всех случаях необходимо обеспечить максимально полный сбор данных и их непредвзятый анализ, как основу оценки не только эффективности, но и соответствия работы ИИ законодательно, нормативно и общественно установленным критериям. Безусловно, критерии для мониторинга должны быть разработаны не отдельными странами, а ЕС в целом. Это позволит обеспечить беспрепятственный мониторинг в рамках всего европейского политического и экономического пространства, а также исключит крайне вредную двусмысленность в национальных законодательствах, которую сейчас активно использует киберпреступность.

§ 2. Реформирование правил и ответственности

Одним из ключевых аспектов политико-правовой регламентации ИИ является выбор режима ответственности за ущерб, нанесенный ИИ, как юридическим, так и физическим лицам. В последние два года государства ЕС и Европейская Комиссия приступили к активному обсуждению этой тематики и анализу концептуальных предложений со стороны юридических фирм и разработчиков ИИ. Особенно остро эта проблема стоит в отношении ущерба, нанесенного конечному пользователю – домохозяйствам (в случае B2B ответственность, как правило, регулируется договорным правом и соответственно прямыми договорами между производителями, эксплуатантами и потребителями).

Есть три основных аспекта темы ущерба, нанесенного ИИ, которые должны быть обсуждены, и соответственно найдено политико-правовое решение в самое ближайшее время. Первый связан со сферой ответственности; второй – с типом конкретных инструментов правовой защиты, а, следовательно, и типом ответственности; третий касается проблем распределения ответственности между производителями. Дело в том, что *ИИ – это всегда сложное программно-аппаратное устройство, в ходе производства которого сотрудничают производители харда, производители софта, сервисные компании, а также провайдеры интернет-услуг*. Кроме того, в настоящее время формируются *компании-эксплуатанты*, которые сдают рабочее время ИИ конечным потребителям по модели «*ИИ как сервис*».

В целом, представляется, что решение проблемы связано с созданием *взаимосвязанной системы технических регламентов и стандартов, определяющих требования к харду, софту, к сервисному обслуживанию и подключению ИИ к интернету*. При наличии таких стандартов и регламентов, отклонения от них и более того, нарушение, которое в итоге повлекло сбой в работе ИИ, нанесший ущерб потребителю, могут быть четко зафиксированы и распределены по конкретным виновникам и масштабам их вины. Более того, как раз в подобного рода случаях в перспективе «закон как текст» может

уступить место «закону как коду». Согласно оценке экспертов в области ИИ, достаточно часто могут иметь место случаи, когда четкое распределение ответственности за сбой в системе ИИ, повлекшее за собой ущерб, установить сложно. Например, зафиксированы случаи, когда сама по себе система ИИ работала в нормальном регламенте при соблюдении всех стандартов, однако при взаимодействии с человеком имели место сбои. Проще всего обвинить в ущербе человека. Однако при тщательном изучении инцидентов было установлено, что сбой произошел из-за растерянности человека, вызванной в свою очередь, не вполне дружественным интерфейсом.

Столь же экзотическими на первый взгляд, однако, достаточно массовыми являются случаи, когда сбои в работе ИИ связаны с вычислениями в облаке. *Когда к облаку одновременно подключаются несколько мощных ИИ*, происходит конфликт взаимодействия между различными вычислительными системами или так называемые флеш-сбои. Здесь на сегодняшний день тоже нет однозначно прописанных требований и критериев, нарушение которых могло бы однозначно трактоваться как вина одного из участников процесса. Третья группа факторов связана со сбоями, которые происходят из-за того, что в рамках одной комплексной системы ИИ используются компоненты и элементы, созданные различными поставщиками.

Отдельную группу составляют факторы, когда при взаимодействии с робототехникой, управляемой ИИ, виноватым оказывается человек. Например, в марте 2018 г. в г. Темпа, штат Аризона автомобиль Volvo, управляемый Uber, не смог распознать женщину, пересекающую улицу, и задавил ее. В результате, государственные органы в течение достаточно длительного времени разбирались, кто несет ответственность за происшествие: производитель датчика распознавания, компания-производитель автомобиля, компания, поставляющая для Volvo видеокамеры или кто-либо еще. В ходе расследования выяснилось, что поодиночке и совместно все эти компоненты системы работали штатно. Одновременно удалось установить, что жертва в момент перехода улицы находилась под воздействием наркотиков и небольшой дозы алкоголя. В результате, с привлечением ведущих специалистов по робототехнике был сделан окончательный вывод. Поскольку действия человека, находящегося в состоянии алкогольного или наркотического опьянения, имеют другую

моторику, нежели в нормальном состоянии, программному обеспечению, распознающему пешехода, было трудно идентифицировать объект как субъекта, который на красный свет пересекал при достаточно сильном движении дорогу. Грубо говоря, женщина погибла из-за целого комплекса обстоятельств, когда в буквальном смысле она была единственной, кто нарушил заранее установленные правила.

Сегодня очевидно, что, например, в Соединенных Штатах, Сингапуре, Нидерландах робомобили, в том числе траки и такси, уже в 2020 г. могли бы массово выйти на дороги. Однако видимо этого не произойдет, поскольку окончательно не решены вопросы распределения ответственности при инцидентах с ними. Вопрос распределения ответственности сегодня является главным для формирования юридической нормативной базы использования ИИ.

Фактически судьбы таких прорывных технологий, как ИИ, зависят от дополнения юридической системы, регулирующей отношения физических и юридических лиц специальными случаями, когда между юридическим и физическим лицом находится сложное устройство, имеющее многих производителей.

Робототехника, оснащенная ИИ, это отнюдь не первое сложное техническое устройство, изготавливаемое многими производителями, с которыми взаимодействуют люди. Лучший пример тому – авиатехника. Там тоже постоянно возникают проблемы поиска конкретного виновника нанесения ущерба и соответственно стороны, выплачивающей компенсацию пострадавшему, либо членам его семьи. Казалось бы, в случае с ИИ можно пойти по пути авиатехники. Однако, существует проблема. Случаи авиакатастроф крайне редки. Ежегодно их случается не более нескольких десятков. Что же касается, например, робомобилей с ИИ тогда, когда их станет сотни тысяч, а то и миллионы, то здесь инциденты могут происходить гораздо чаще, особенно на первом этапе.

Как бы там ни было, в рамках определения режима ответственности, необходимо решить три главных вопроса. Первый – будет ли ответственность выражена, прежде всего, в финансовой форме, либо предусматривать и уголовные сроки наказания. Второй – будет ли перед конечным пользователем нести ответственность только эксплуатант ИИ, либо будет использоваться система солидарной

ответственности, включающая всех поставщиков, провайдеров и эксплуатантов, связанных с конкретным ИИ. Третий – будет ли реализован принцип одноранговой или многоранговой ответственности. В случае одноранговой – безусловную ответственность перед потребителем несет только эксплуатант ИИ. Он же в свою очередь выставляет регрессивные иски поставщикам ИИ или отдельных его компонентов. В случае многоранговой ответственности изначально перед потребителем будут нести ответственность все так или иначе связанные с конкретным изделием ИИ производители. В этом случае суть расследования будет в том, чтобы свести до минимума круг конкретных виновников, исключив из него добросовестных производителей компонентов ИИ.

Покажем это на простом примере. Идеальным примером одноранговой ответственности является строгая ответственность работодателя перед клиентом за действия своих сотрудников при исполнении ими своих обязанностей. В реальной практике в странах ЕС в основном используется так называемая относительно строгая ответственность. Она допускает более широкий набор исключений, часто связанных со степенью предсказуемости действия и т. п. типичным случаем относительно строгой ответственности является действующий режим ответственности за качество продукции или ответственности за ущерб, причиненный животными, находящимися под наблюдением своих хозяев.

Что касается ответственности перед конечным пользователем, то большинство правовых систем стран ЕС базируются на принципе «единого окна», т. е. единого субъекта, который несет ответственность за возмещение ущерба. При этом эта система предусматривает дальнейшее расщепление ответственности. Материальную ответственность в таких случаях несет один единый субъект, как правило, страховая компания, а ответственность за нарушения технических регламентов, стандартов, отраслевых правил и т. п., носящие в том числе уголовный характер, несет конкретный виновник происшествия.

Кроме того, в европейском законодательстве подход «единого окна» предусматривает возможность первому ответственному субъекту подать в суд на другие субъекты цепочки создания стоимости для получения возмещения. Например, если авария вызвана неполадками в

датчике автомобиля с самостоятельным управлением, то изготовитель автомобиля будет нести ответственность перед лицами, которым нанесен ущерб. При этом, он сможет добиться возмещения своих затрат путем иска непосредственно производителю датчика.

Кроме того, в случае ИИ возникает проблема, которой нет в коллизиях, связанных с авиатранспортом, автомобилями, судами и т. п. Невероятно малая доля происшествий с такого рода сложными системами связана с человеческим фактором не на стороне эксплуатанта, а на стороне пользователя. В прошлом году произошло лишь два случая, когда люди погибли в аэропортах стран ЕС, оказавшись на взлетно-посадочных полосах во время взлета и приземления самолетов. Тщательный анализ выявил, что в обоих случаях люди сознательно обманули охранные системы и персонал аэропорта. Пока подобного рода инциденты крайне редки, но при массовом внедрении ИИ казусы и инциденты по вине пользователей могут стать гораздо более частыми. При этом сами пользователи будут стремиться получить материальную компенсацию за полученный ущерб. Соответственно, если не вооружить суды четкими и однозначными рекомендациями и инструкциями, чем дальше, тем больше они будут заниматься инцидентами, связанными с человеко-машинным взаимодействием, априори решая конфликты в пользу человека.

Есть еще одна проблема. Она на данный момент совершенно не отрефлектирована специалистами в юриспруденции. Особенность ИИ заключается в том, что он действует в большинстве случаев в рамках всемирной сети. Соответственно, весьма вероятно, что отказы или сбои в работе ИИ могут происходить не по причинам, связанным с программноаппаратными средствами, а с общесетевыми проблемами. На сегодняшний день нет даже наметок законодательного определения в рамках солидарной ответственности конкретной ответственности всех юридических лиц, связанных с функционированием ИИ в интернете, включая провайдеров.

§ 3. К кому приравнять ИИ: к объектам, животным или рабам?

Рассмотрение режима ответственности для ИИ неизбежно сводится к наиболее фундаментальному вопросу: можно ли рассматривать ИИ как обычный технический объект, который находится под полным контролем человека, или же ИИ обладает некоторыми элементами автономии и разумности. Если ответ положительный, то это делает возможным применение к компаниям, производящим и эксплуатирующим программно-аппаратные комплексы ИИ, того же набора правил, включая технические регламенты, отраслевые и объектные стандарты и юридические нормы, как к производителям и эксплуатантам обычных технически сложных, многокомпонентных изделий, типа самолетов, кораблей и т. п.

Если ИИ рассматривается как своего рода развитие или расширение возможностей человека, то правила ответственности, применимые к людям можно ограниченно использовать и для ИИ. При этом ответственность появляется, когда имеет место субъективный элемент, например, халатность или намерение причинить вред, которые привели к нанесению ущерба, а также на основе дистрибуции собственности, которая наносит ущерб третьей стороне.

Разумно ожидать, что ИИ в перспективе будет использоваться в основном не как объект, а как услуга.

В этих обстоятельствах остается открытым вопрос, должна ли ответственность за услугу иметь договорной характер или ответственность неконтрактного характера (деликатная ответственность).

В последние годы *растет число юристов, отвергающих приравнение ИИ к обычным сложным многокомпонентным устройствам.* Они предлагают *рассматривать ИИ как животное,* особенно в тех случаях, когда ИИ предполагает определенную степень автономии. Это опция в последние годы подкрепляется научными исследованиями, согласно которым в ближайшие годы продвинутые системы ИИ сравниваются по мощности интеллекта с кошкой. Этот вариант подразумевает, что сам по себе ИИ не имеет статуса

юридического лица, но также как опасные дикие животные налагает большие дополнительные требования к собственникам ИИ, так как это сегодня действует в законодательствах большинства стран относительно ответственности владельцев диких животных. Наиболее детальная система ответственности за животных, содержащихся частными лицами в неволе, предусматривается в системе общего права США.

В последние годы ряд британских и израильских юристов предложили разрабатывать законодательство по ИИ по аналогии с римским правом в отношении рабов. Такое толкование подтверждается тем фактом, что само по себе слово «робот» впервые использованное Карлом Чапеком на чешском языке, означает «принудительный труд» или «раб». В римском праве хозяева несут ответственность за ущерб, причиненный их рабами. В южных штатах Америки еще в первой половине XIX века хозяин нес полную ответственность за каждый противозаконный поступок раба, независимо от того, был ли он совершен на службе и в рамках выполнения работ в интересах рабовладельца, либо в частной жизни раба.

Ряд американских правоведов предложили рассматривать ИИ как наемного работника и предоставлять ему юридическое лицо, а также обязанность проявлять должную осторожность. При этом, строгая ответственность по-прежнему будет возлагаться на владельца ИИ, но сам по себе ИИ, имеющий юридическое лицо, может быть запрошен для компенсации ущерба, в том числе реквизирован в пользу лица, которому был нанесен ущерб. Фактически это будет своего рода передача раба от одного рабовладельца другому в счет покрытия ущерба.

Этот американский подход связан с успехами в сверхбыстром глубинном обучении. Часть современных программистов и математиков полагают, что уже в ближайшие три-пять лет ИИ будет способен сам себе писать программы, а также реплицировать усовершенствование модели себя на автоматизированных предприятиях. В этом случае вообще не понятно, в какой роли выступает человек – собственник ИИ, и к кому предъявлять требования компенсации за нанесенный ущерб. В этой связи не вызывает удивления предложение, поступившее от Нидерландов в Европейский парламент, которое предусматривает рассмотрение

семейство роботов, подключенных к единому мощному ИИ даже не как наемных работников, а как прямых юридических лиц, несущих неограниченную ответственность без связи с разработчиком.

§ 4. Действующее законодательство ЕС

Нынешние правила ЕС об ответственности за системы ИИ в основном базируются на двух законодательных актах: Директиве об ответственности за качество продукции и Директиве об ответственности за механизмы. Эти директивы предусматривают, что если устройство причиняет вред человеку или его частной собственности, то производитель несет ответственность за выплату компенсации.

Недавняя оценка показала, что директивы по-прежнему обеспечивают хороший баланс между защитой прав потребителей и поощрением инноваций. При этом ЕС пообещал, что в течение 2019 г. будут опубликованы толкования директив, где будут обновлены и привязаны к инновациям такие понятия, как «дефект», «продукт», «производитель», «узлы и детали», «программные» и «аппаратные». При этом в толкованиях предусматривается, можно ли включить использование «ИИ как услуга» в сферах действия директив, а также как интерпретировать разницу в ответственности изготовителей харда и софта.

В последнее время ЕС опубликовал *Директиву по машинному оборудованию*. Она устанавливает общие требования по охране труда и технике безопасности для роботов и 3D принтеров. Это – хороший пример, как совершенствуется европейское законодательство в наиболее инновационных сферах.

Директива по машинному оборудованию является законодательным актом, который предусматривает создание «нормативных песочниц», а также комплексное регулирование, включающее общеевропейское и страновое юридическое регулирование, саморегулирование ассоциаций производителей и общественный контроль со стороны гражданского общества.

Будущий режим ответственности использования ИИ должен быть разработан в сочетании с совершенствованием системы страхования. Предполагая, что всегда можно отследить ответственное лицо в автоматизированном процессе, будущие руководящие принципы политики могут потребовать, чтобы компания, разрабатывающая, внедряющая и продающая ИИ, представляла регулирующим органам

финансовые документы, подтверждающие ее способность имущественно реагировать на любые потенциальные обязательства, возникающие в ходе эксплуатации ее техники. В случае если это невозможно, пользователи должны знать и иметь возможность либо отказаться от такого использования, либо обратиться к страховщикам, которые самостоятельно принимают решение о страховании подобного рода изделий.

При рассмотрении правовых основ регулирования, необходимо иметь в виду не только указанные выше директивы, но и множество отраслевых нормативных актов, регламентирующих стандарты, технические регламенты и юридические взаимоотношения. Например, страховые компании и банки уже распространяют некоторые требования о раскрытии информации, которые сложились в высокотехнологичных отраслях и на ИИ. Хорошим примером этого является Директива о распределении страхования от 1 октября 2018 г.

В соответствии с Директивой страховые компании смогут принимать на страхование лишь те сложные технические средства, производители которых предоставляют утвержденный перечень информации о техническом средстве, его компонентах, а также возможных выявленных, в том числе в ходе испытаний и практической эксплуатации уязвимостях, порождающих определенные риски.

Первопроходцем в деле юридического регулирования робототехники и ИИ стала Британия. В июле 2018 г. первой в мире она приняла Закон об автоматизированных электрических транспортных средствах, включая беспилотные автомобили, суда, а также дроны. Закон 2018 г. предусматривает подход «легкого прикосновения», и как ожидается, что предусмотрено структурой закона будет обновляться по мере развития технологий и инфраструктуры. Это – первый законодательный акт во всем мире, который построен по принципу легкого прикосновения. Т. е. в нем предусмотрены лишь основные понятия и термины в этой сфере и общая конструкция, связанная с правами, обязанностями и ответственностью производителей, эксплуатантов автоматизированных транспортных средств, а также конечных потребителей. Закон построен таким образом, чтобы по мере развития робототехники он мог в упрощенном порядке развиваться и расширяться сообразно динамике техники и технологий.

Указанный закон является *первым законодательным актом во всем мире, в котором изложена формализованная правовая модель страхования и ответственности за роботомобили – автоматические транспортные средства*. Поскольку это абсолютная юридическая новация, учреждения ЕС в настоящее время проявляют определенную осторожность, называя его моделью. Что же касается Великобритании, то до конца 2019 г. будет принят дополнительный подзаконный акт, регламентирующий вступление в силу указанного закона и особенности его трактовки.

Что касается содержания закона, то необходимо отметить ряд обстоятельств. Закон предоставляет страховщикам возможность ограничить свою ответственность в соответствии с разделом 2 (1) перед страхователями, и в определенных обстоятельствах перед третьими лицами, когда несчастные случаи были вызваны отказом застрахованного лица, либо с ведома застрахованного лица установить критические для безопасности обновления программного обеспечения. Иными словами, если эксплуатант беспилотного автомобиля не установил вовремя обновление, предоставленное производителем софта к машине, и по этой причине произошла авария, то всю ответственность будет нести эксплуатант, а не производитель автоматизированного средства.

Аналогичным образом несчастные случаи, вызванные самостоятельным изменением программного обеспечения эксплуатантом, запрещенные условиями полиса страхования, также могут быть исключены из страхового покрытия. Тем не менее, страховщики по-прежнему будут нести строгую ответственность за выплату компенсаций большинству третьих лиц, пострадавших в результате несчастных случаев, к которым применяется закон 2018 г.

Британский закон создает важнейший прецедент. Он предусматривает основополагающую роль страховщиков в контроле за программно-аппаратным обеспечением автономных транспортных средств. Соответственно, со временем страховщики будут оказывать давление на поставщиков ИИ, чтобы перейти к определенному уровню стандартизации и унификации. Вполне очевидно, что страховщики из компаний, рассчитывающих риски, получающих и отдающих деньги превратятся в компании по контролю за техникой безопасности

автоматизированных транспортных средств, а затем и иной робототехники.

Закон также устанавливает законное право суброгации для страховщиков – владельцев полностью или частично возместить свои убытки за счет другой стороны, ответственной, в конечном счете, за произошедший несчастный случай. Что касается взаимодействия между человеком и машиной, в недавно принятом правительственном Кодексе испытания автономных транспортных средств разъясняется, что водитель или оператор роботомобиля и любого другого автоматизированного средства, типа дрона или судна, должен обязательно иметь возможность взять под свой контроль транспортное средство, отключив автоматизированное управление в любое время. Таким образом, создается впечатление, что пока законодатели хотят избежать возможности появления полностью автоматизированных транспортных средств, которые будут априори не только независимы от человека, но и не предусматривать возможности вмешательства в экстренных случаях со стороны оператора.

Важно отметить следующее. Ситуацию, которую законодатели решают в отношении робомобилей – беспилотных легковых машин и грузовых траков сходны с историей дронов в первой половине десятых годов. Тогда шли ожесточенные дискуссии относительно возможности или невозможности производства военных дронов, выполняющих все функции, включая открытие огня без вмешательства человека. Соединенные Штаты и страны ЕС жестко находились на позиции разделения функций ИИ и человека в управлении дроном: ИИ управлял полетом, а человек принимал решение об открытии огня. В то же время Израиль и ряд других государств полагали возможным производство и боевое использование полностью автоматизированных дронов. В рамках дискуссии было проведено натурное испытание, в ходе которого беспилотные дроны показали полное превосходство перед дронами с использованием человека.

§ 5. Стимулирование инноваций на основе данных с использованием политических инициатив

Область, в которой институты ЕС должны быть особенно активными, это – инновации, основанные на данных. На заседаниях Целевой группы CERPS удалось установить, что в большинстве случаев разработка ИИ представляет собой полный технопакет или стек технологий, базирующийся на использовании и обработке больших наборов данных. Отсутствие данных стало ключевой проблемой, не позволяющей большинству европейских компаний использовать методы машинного обучения и ИИ. Хотя не все типы ИИ требовательны к данным и используют машинное обучение, в подавляющем большинстве случаев наличие больших наборов данных – это главная предпосылка эффективного развития ИИ.

При этом разработчики в Соединенных Штатах и в Китае могут рассчитывать на работу с намного более разнообразными и обширными данными, чем в Европе, где существуют гораздо более строгие законы о конфиденциальности при использовании данных.

В общем и целом, сейчас в странах ЕС большие наборы данных более доступны, чем в прошлом. Однако лишь немногие участники европейского рынка имеют прямой доступ к наборам данных, позволяющим использовать эффективные методы ИИ. По сути, Европа с жесткими критериями конфиденциальности, соблюдения авторского права и правилами приобретения данных и т. п. с каждым годом все более уступает в конкурентной борьбе Китаю и США, где данные на порядок более доступны, чем в Европе.

Как найти правильный баланс между обеспечением доступности больших данных и необходимостью сохранять неприкосновенность частной жизни и конфиденциальность бизнеса? Целевая группа CERPS согласовала следующие рекомендации.

Крайне важно обеспечить максимальную открытость данных, собираемых правительствами и статистическими государственными органами на всех уровнях – от федерального до уровня коммун – и их предоставление в форматах, пригодных для машинного обучения и

использования. На сегодняшний день это – главное направление перехода от закрытого к открытому правительству в рамках ЕС.

По состоянию на сегодняшний день, несмотря на принятые на уровне ЕС и государств-членов решения, основные и наиболее интересные массивы данных, хранящихся в правительствах и собранных государствами за счет средств налогоплательщиков, недоступны для исследователей, предпринимателей и компаний, желающих участвовать в инновациях, основанных на данных. Так, например, только в 12 странах ЕС в число открытых включены данные, связанные с правонарушениями любого уровня тяжести, а также данные о криминогенности тех или иных регионов, муниципий и даже кварталов.

Несмотря на Директиву ЕС, ни одна из новых стран-членов ЕС, за исключением балтийских стран, не раскрыла данные по расходам самого правительства на государственный аппарат и государственных служащих. В связи с этим 25 апреля 2018 г. на рассмотрение Европейской Комиссии внесен пересмотренный вариант Директивы об информации в государственном секторе. Он предусматривает обязательное для всех стран-членов ЕС предоставление информации в машиночитаемом виде по расширенному перечню номенклатуры документов и позиций, включая особо чувствительную для гражданского общества и бизнеса информацию по экологической обстановке, информацию, связанную с правонарушениями, а также расходованием средств на оборону, разведку, полицию и т. п.

Наряду с обеспечением перелома в предоставлении правительственных данных, важно ускорить формирование единого общеевропейского рынка данных. Он предусмотрен регламентом, принятым в ноябре 2018 г. В то же время возможные исключения из свободного потока, например, основанного на требованиях национальной безопасности того или иного государства, должны быть заранее кодифицированы с тем, чтобы избежать произвольной трактовки регламента отдельными государствами.

Регламент запрещает правительствам и законодательным органам отдельных стран вводить ограничения на локализацию данных. Это позволяет коммерческим организациям хранить данные в любой точке ЕС, или, если это транснациональные международные компании, то по обязательному согласованию с ЕС – за пределами ЕС. Это позволяет

компетентным органам получать доступ к данным для контроля и надзора независимо от того, где они хранятся и обрабатываются в ЕС. Регламент также предусматривает создание в 2019 г. Кодекса поведения для провайдеров услуг, которые хранят или обрабатывают данные. На основе Кодекса имеется в виду создать саморегулирующуюся Ассоциацию провайдеров и обработчиков больших данных ЕС.

ЕС в отличие от некоторых других стран и региональных объединений уделяет и будет продолжать уделять самое пристальное внимание защите данных. Европейская Комиссия будет и далее финансировать исследования, инновации, стандартизацию в области конфиденциальности «по умолчанию» или «по замыслу». ЕС будет всячески поощрять провайдеров и обработчиков данных, реализовывать для их хранения и анализа наиболее эффективные криптографические решения, и в первую очередь гомоморфное шифрование и нулевое доказательство знаний. Магистральным направлением развития баз данных и хранилищ неструктурированной информации Еврокомиссия полагает переход от традиционных решений к решениям, основанным на блокчейне. Эти решения уравнивают в правах правительства, бизнес, гражданское общество и позволяют выстраивать не только и без того господствующие вертикальные связи, но и горизонтальные цепочки создания ценностей на основании данных. ЕС полагает, что блокчейн-технологии станут основным направлением хранения и обработки данных на ближайшую перспективу.

Необходимо решить вопрос, связанный с разрешением интеллектуального анализа текстов и данных, как в исследовательских, так и в коммерческих целях. В отличие от Соединенных Штатов, Китая и России, в Европе крайне затруднен интеллектуальный анализ текстов по причине необходимости соблюдения авторского права. В отличие от упомянутых стран, в Европе для того, чтобы осуществлять интеллектуальный анализ текста, необходимо получить согласие автора текста или правообладателя. На практике это невозможно. В связи с этим уже сегодня европейский бизнес, который подчиняется жестким этическим принципам (так называемому этическому анализу данных) проигрывает своим конкурентам за океаном, в Китае и России.

В ближайшее время эксперты ЕС рекомендуют *избегать дальнейшего ужесточения регулирования доступа к данным, связанного с соблюдением авторского права, интеллектуальной собственности и т. п.* В целом бизнес не одобряет новый тип права собственности на данные и указывает, что ключевой вопрос во взаимодействии касается не только собственности, но и того, как организован доступ. Нынешние правила фактически блокируют возможность специализированных компаний обрабатывать личные данные, а также данные, на которые распространяется право интеллектуальной собственности.

Существует необходимость в осуществлении большей ясности относительно правовой основы для машинно генерируемых данных. Европейская Комиссия полагает, что в ближайшем будущем регулирование данных будет связано с переходом от общих регулятивных документов к отраслевым, учитывающим конкретную специфику. Безусловно, переход к уже одобренным на уровне ЕС правилам и подходам к правовому регулированию доступа, хранения и обработки данных, потребует всестороннего учета интересов не только собственников данных, но и их провайдеров, а главное, конечных потребителей.

Возможно, наиболее острый вопрос в ЕС – это тема использования персональных и индивидуальных, но обезличенных данных. Законодательство ЕС – не просто наиболее сырое среди всех ведущих IT зон мира, но и носит по существу почти запретительный характер. По мнению целевой группы, эта тема является предметом наиболее острых дискуссий, как в законодательных органах стран ЕС, так и в гражданском обществе.

При всей разнородности позиций фактом остается то, что сверхжесткое регулирование, по сути, запрет использования персональных и даже обезличенных индивидуальных данных не позволили не то, что в масштабах отдельных стран, но и ЕС в целом, создать континентальный поисковик, конкурирующий с Google, Bing и Yandex, собственную европейскую социальную сеть или торговую платформу, типа Amazon. Хорошо известно, что в основе экономики этих платформ лежит сделка между ними и пользователями, когда бесплатный доступ к услугам платформы разменивается на бесплатный доступ к данным о пользователе. Эту сделку потребители

осуществляют вполне сознательно и ставят специальный значок согласия. Если ЕС не осознает это и не изменит своей позиции, то окончательно и бесповоротно превратится в рынок для сторонних платформ. Это в цифровом мире равносильно колониальному статусу.

§ 6. Принципы ОЭСР по ответственному управлению надежным ИИ

25 мая 2019 года все 36 стран, входящих в Организацию экономического сотрудничества и развития (ОЭСР), поддержали *Принципы ответственного управления надежным ИИ*^[13]:

1. Инклюзивный рост, устойчивое развитие и благополучие

Заинтересованные стороны должны активно участвовать в ответственном управлении заслуживающим доверия ИИ в достижении полезных результатов для людей и планеты, таких как расширение человеческих возможностей и повышение креативности, расширение участия недопредставленных групп населения, сокращение экономического, социального, гендерного и другого неравенства и защита природной среды, таким образом стимулируя инклюзивный рост, устойчивое развитие и благосостояние;

2. Ориентированные на человека ценности и справедливость

а) Участники ИИ должны уважать верховенство закона, права человека и демократические ценности на протяжении всего жизненного цикла системы ИИ. К ним относятся свобода, достоинство и автономность, конфиденциальность и защита данных, недискриминация и равенство, разнообразие, справедливость, социальная справедливость и трудовые права, признанные на международном уровне.

б) С этой целью субъекты ИИ должны внедрять механизмы и гарантии, такие как способность к человеческой решимости, которые соответствуют контексту и соответствуют уровню техники;

3. Прозрачность и объяснимость

Участники ИИ должны обеспечить прозрачность и ответственное раскрытие информации о системах ИИ. С этой целью они должны предоставлять значимую информацию, соответствующую контексту и соответствующую уровню техники:

- способствовать общему пониманию систем ИИ,

- информировать заинтересованные стороны об их взаимодействии с системами искусственного интеллекта, в том числе на рабочем месте,
- дать возможность пострадавшим от системы ИИ понять результат и внутренне дать возможность тем, на кого система ИИ оказывает негативное воздействие, оспаривать ее результаты на основе простой и понятной информации о факторах и логики, которая послужила основой для прогноза, рекомендации или решения;

4. Надежность и безопасность

а) Системы AI должны быть надежными и безопасными на протяжении всего их жизненного цикла, чтобы в условиях нормального использования, предполагаемого использования или неправильного использования или других неблагоприятных условий они функционировали надлежащим образом и не представляли необоснованного риска для безопасности;

б) с этой целью субъекты ИИ должны обеспечивать отслеживаемость, в том числе в отношении наборов данных, процессов и решений, принятых в течение жизненного цикла системы ИИ, для обеспечения возможности анализа результатов и ответов системы ИИ на запрос, соответствующих контексту и в соответствии с уровнем техники;

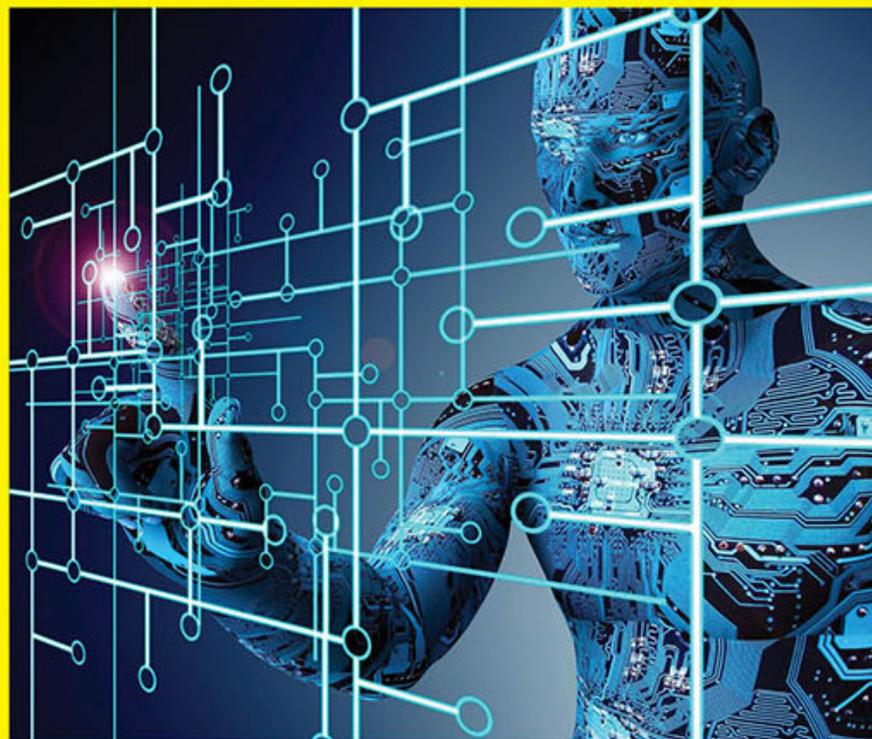
с) субъекты ИИ должны, исходя из своих ролей, контекста и способности действовать, применять систематический подход к управлению рисками на каждом этапе жизненного цикла системы ИИ на постоянной основе для устранения рисков, связанных с системами ИИ, включая конфиденциальность, цифровую безопасность, безопасность и предвзятость;

5. Подотчетность

Участники ИИ должны нести ответственность за надлежащее функционирование систем ИИ и за соблюдение вышеуказанных принципов, исходя из их ролей, контекста.

**ЕЛЕНА ЛАРИНА
ВЛАДИМИР ОВЧИНСКИЙ**

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ



ЭТИКА И ПРАВО

Примечания

1

См.: Е. Ларина, В. Обнинский. Искусственный интеллект. Большие Данные. Преступность. М., Книжный мир, 2018

[Вернуться](#)

2

Подобный язык логических правил специально для нужд законодательства был разработан в СССР в МФТИ для кодификации и выяснения противоречий в советском, затем в российском законодательстве, использованного в строительной отрасли, коллективом Спартаксом Никаноровым. Далее этот язык использовался для создания непротиворечивых нормативных документов в Сбербанке, Газпроме и т. п. Сейчас в МФТИ действует единственная в мире кафедра логического проектирования законодательства и действующая при ней практическая лаборатория – *Авторы*.

[Вернуться](#)

3

Речь идет о том, что в настоящее время алгоритмы большинства ИИ, используемых, скажем, в США, в судебных и правоохранительных органах, базируются на машинном обучении. Машинное обучение – это, по сути, метод распознавания чего-либо на основе данных, характеризующих параметры этого чего-либо. Т. е. программа получает числовое описание портретов, например, европейцев и азиатов и на основании этих данных сама учится распознавать новые фотографии, относя их обладателей соответственно к европеоидной или азиатской расам. Фактически, машинное обучение, это статистические методы обработки прежней исторической информации,

вычисления на ее основе основных характеристик и придание им наибольших весов при распознавании нового. Поскольку, скажем, в США наибольший уровень преступности сегодня в бедных латиноамериканских эмигрантских кварталах с высокой долей молодежи, то для машины молодой нелегальный мигрант в США в возрасте, скажем, до 20 лет без официального гражданства будет наиболее подозреваемым ИИ в преступлении лицом. Из белого и латиноса машина выберет латиноса в полном соответствии с прошлыми данными. Это и называется пристрастностью ИИ, которая влияет на правоохранные и судебные решения. – *Авторы.*

[Вернуться](#)

4

Человек, когда принимает решение, использует несколько критериев, одни из которых более важны, другие – менее важны. Сегодня для наиболее продвинутых ИИ нет способа выяснить в компактном и понятном виде, на основе каких конкретно критериев принимается решение относить объект к той или иной совокупности. Это вызывает большие опасения, потому что люди принимают решения, не понимая, почему они это делают. Проблема черного ящика и непонятности действий нейронных сетей для принимающих решения лиц без специального математического образования – это главная проблема использования ИИ в госуправлении. – *Авторы.*

[Вернуться](#)

5

Согласно данным доклада ФБР 2018 г. об использовании ИИ, а также конференции начальников полиции штатов в августе 2018 г. указанные в тексте факторы примерно на 80–90 % позволяют распознать в совокупности подозреваемых потенциального преступника с точностью более 95 %. Т. е. проблема состоит в том, что люди определенного возраста, пола, этнической, расовой принадлежности с конкретными социально-экономическими характеристиками совершают преступления намного чаще, чем

представители других групп. Выявление значимых для вовлечения индивидуума в организованную преступность факторов и является главной функцией ИИ в правоохранительной и судебной сферах. Соответственно ИИ, как огромный арифмометр, быстро обрабатывающий гигантские данные и определяющий числовые зависимости между всем, чем угодно в этих данных, опираясь на прошлый опыт, выраженный в корреляциях, регрессиях, факторном анализе делает очень точные выводы. По сути, ИИ – это методы максимально многомерного статистического анализа исходных данных. Как результат ИИ получает только то, что содержится в исходных данных, но непосредственно не обнаруживается человеком и представляет собой неочевидные связи. Поэтому американцы и страны Британского содружества его применяют. Европейцы же этой Хартией делают по сути невозможным использование ИИ, дело в том, что он будет указывать на, скажем, мусульманских молодых эмигрантов без работы гораздо чаще, чем накоренного немца со средним доходом и среднего возраста. А это противоречит толерантному подходу Европейской Комиссии. Между дискриминацией, как они ее называют и эффективностью прогноза, ради чего и используется ИИ, существует не прямая, как хотелось бы, а обратная зависимость. Просто в Европе правду называют дискриминацией. – *Авторы.*

[Вернуться](#)

6

В основу главы положен REPORT ON ALGORITHMIC RISK ASSESSMENT TOOLS IN THE U.S. CRIMINAL JUSTICE SYSTEM – THE PARTNERSHIP ON AI, 2019

[Вернуться](#)

7

Эрика Лумиса приговорили к шести годам заключения по рекомендации секретного проприетарного программного обеспечения частной компании. Лумис, который уже имел криминальную историю и был приговорен за то, что бежал из полиции на угнанной машине,

теперь утверждает, что его право на процедуру было нарушено, поскольку ни он, ни его представители не смогли рассмотреть или опротестовать алгоритм рекомендации. Отчет был подготовлен программой Comras, которая продается Notrpointe судам. Программа воплощает новую тенденцию в исследованиях ИИ: помогает судьям принять более ориентированное на данные решение в суде.

Хотя конкретные подробности дела Лумиса остаются закрытыми, в нем наверняка содержатся диаграммы и числа, определяющие жизнь, поведение и вероятность рецидива Лумиса.

Среди них возраст, раса, гендерная идентичность, привычки, история браузера и какие-нибудь измерения черепа. Точнее никто не знает. Известно, что прокурор по делу сказал судье, что Лумис продемонстрировал «высокий риск рецидивизма, насилия, досудебного разбирательства». Это стандартно, когда дело касается вынесения приговора. Судья согласился и сказал Лумису, что «по оценке Comras он был определен как лицо, представляющее высокий риск для общества». Верховный суд штата Висконсин осудил Лумиса, добавив, что отчет Comras привнес ценную информацию в его решение, но отметил, что без него он вынес бы такой же приговор.

[Вернуться](#)

8

Artificial Intelligence. Ethics, governance and policy challenges, *Report of a CEPS Task Force*. Centre for European Policy Studies (CEPS). Brussels, February 2019.

[Вернуться](#)

9

European Tech Insights 2019. IE. Centre The Governance of chance.

[Вернуться](#)

10

Подробно о правовом регулировании ИИ в разных странах, в том числе и в России, см.: Понкин И.В., Редькина А.И. Искусственный интеллект с точки зрения права. Вестник РУДН. Серия: Юридические науки. 2018, т.22, № 1, с. 91–109; Новые законы робототехники. / Под ред. А.В. Незнамова. М., 2018; Регулирование робототехники: введение в «робоправо». / Под ред. А.В. Незнамова. М., 2018; Основы государственной политики в сфере робототехники и технологий искусственного интеллекта. / Под ред. А.В. Незнамова. М., 2019 и др.

[Вернуться](#)

11

Uber, AirBNB, Amazon и т. п. являются наиболее быстрорастущими сегментами американской глобальной экономики, т. е. установлению юридических требований через кодирование алгоритмов или алгоритмическое регулирование принадлежит будущее.

[Вернуться](#)

12

В данном случае закон фактически перестает регулировать деятельность обслуживающего персонала на платформе, например, телефонных консультантов, службы доставки и т. п. Их регулирует традиционное законодательство. В то же время в законе четко прописывается требование, например, свободы доступа клиента ко всем предложениям на платформе или напротив, обязательное установление подсистемами недопуска на платформу лиц, моложе 16 лет. Т. е. законодательство теперь не регулирует поведение, а определяет некоторые требования к алгоритмам, т. е. к тем или иным процедурам и последовательностям, которые затем, будучи записанными на языке программирования, т. е. к программам, которые автоматически исполняются на платформе. Т. о. закон впервые превращается в своего рода текстовое описание алгоритма или последовательности операций, которое затем кодируется на том или ином языке программирования и превращается в автоматически исполняемую платформой программу. Закон, таким образом,

становится базисом для кода. Наиболее полно на сегодняшний день этот подход реализован в блокчейне, который, по сути, и представляет собой автоматически исполняемые действия в соответствии с кодированными контрактами. – *Авторы.*

[Вернуться](#)

13

Под системой ИИ ОЭСР понимает машинную систему, которая может для определенного набора определенных человеком задач делать прогнозы, рекомендации или решения, влияющие на реальную или виртуальную среду. Системы ИИ предназначены для работы с различными уровнями автономии.

[Вернуться](#)