

Huseynova R.B.♦

DOI: 10.25108/2304-1730-1749.iolr.2020.62.100-104

UDC 004.8

**The quality of data in machine learning models:  
some challenges of artificial intelligence  
using in criminal proceedings**

**Abstract:** The aim is to study the problem of the using of qualitative data in machine learning models for achieving effective, reliable and impartial results through using artificial intelligence tools within the framework of criminal pre-trial and trial. It is discussed some challenges arising from data processing for the purpose of its application in algorithms and machine learning models in light of human rights and protection of criminal proceedings participant procedural rights, as well as the relevance of data to specific tasks of artificial intelligence tools.

**Keywords:** data; machine learning model; artificial intelligence tools; criminal proceedings; human rights; procedural rights.

**References**

1. *Kommentary k Ugolovnomu Kodeksu Azerbajjanskoy Respubliki* pod red. professora F.Yu. Samandarova [Commentary to Criminal Code of Azerbaijan Republic. Ed. by Prof. F.Yu. Samandarov]. Baku, Digesta Publ., 2006, 1104 p. (in Azerbaijani)
2. *Konventsiya o zaschite prav cheloveka i osnovnykh svobod* [Convention on Human Rights and Fundamental Freedoms]. Available at: [https://www.echr.coe.int/Documents/Convention\\_RUS.pdf](https://www.echr.coe.int/Documents/Convention_RUS.pdf) (accessed: 14.01.2020)
3. *Chto soboy predstavlyayet model' mashinnogo obucheniya* [What is presented the model of machine learning itself]. Available at: <https://docs.microsoft.com/ru-ru/windows/ai/windows-ml/what-is-a-machine-learning-model> (accessed: 14.01.2020)
4. *Types of AI: distinguishing between weak, strong, and super AI*. Available at: <https://www.thinkautomation.com/bots-and-ai/types-of-ai-distinguishing-between-weak-strong-and-super-ai/> (accessed: 17.01.2020).

---

♦ **Huseynova Rena Boyukaga** – Scientific Secretary of the Presidium of National Academy of Sciences of Azerbaijan.  
E-mail: rena.boyukaga@gmail.com

DOI: 10.25108/2304-1730-1749.iolr.2020.62.100-104

УДК 004.8

**Качество данных в моделях машинного обучения:  
некоторые вопросы использования искусственного  
интеллекта в уголовном судопроизводстве**

**Аннотация:** Целью данной работы является изучение проблемы использования качественных данных в моделях машинного обучения в интересах достижения эффективного, достоверного и непредвзятого результата при использовании инструментов искусственного интеллекта в рамках досудебного и судебного производства по уголовным делам. Рассматриваются некоторые вопросы, возникающие в связи с обработкой данных для алгоритмов и моделей машинного обучения в контексте защиты прав человека и процессуальных прав участников процесса, а также относимости данных к конкретным задачам, поставленным перед искусственным интеллектом инструментами.

**Ключевые слова:** данные; модель машинного обучения; инструменты искусственного интеллекта; уголовное судопроизводство; права человека; процессуальные права.

Стремительное развитие высоких технологий, в том числе технологий искусственного интеллекта, и его практическое применение во многих сферах человеческой деятельности задает и, скорее всего, как минимум в ближайшие пару десятилетий будет задавать вектор деятельности правоведов. Задача не из легких для юристов-гуманитариев и вызов устоявшимся стратегиям права, формировавшимся на протяжении столетий под влиянием идей и на основах глубокого почитания основ и принципов Римского права - прообраза правовых систем многих континентальных государств, а также прецедентного права, история формирования и развития которого насчитывает более 7 веков. Возможно, впервые юристы столкнутся с задачами, которые не стояли перед ними ни разу за время всех промышленных революций в истории человечества вместе взятых. По-видимому, Четвертая промышленная революция определит программу-минимум для правоведов, значительную часть которой будет составлять подготовка базовых правовых концепций и разработка гибких правовых стратегий. Однако, кардинальное изменение стратегий права, по крайней мере, в уголовном судопроизводстве, выйдет на повестку дня в случае нарастающей тенденции применения во многих областях человеческой деятельности Искусственного Общего Интеллекта (“Artificial General Intelligence (AGI)”) или, что на сегодня пока не рассматривается, Искусственного Супер Интеллекта («Artificial Super Intelligence (ASI)») [4]. Применение ИОИ и ИСИ безусловно повлечет за собой «Революцию Права» и, потому, многое в будущих стратегиях права зависит от того, насколько детально, качественно и «осторожно» поработают правоведы на начальных стадиях применения технологий искусственного интеллекта в общем и машинного обучения в частности.

♦ Гусейнова Рена Бююкага кызы – ученый секретарь Президиума Национальной Академии Наук Азербайджана. E-mail: rena.boyukaga@gmail.com

Представляется важным, что для решения задач и целей досудебного и судебного производства, необходимо осознавать, что использование высоких технологий, в том числе искусственного интеллекта – это не модное нововведение, которое само по себе приспособится и будет сбалансировано существовать в правовой системе государства, а серьезный инструмент, требующий тщательного правового регулирования на новом уровне восприятия действительности. Совершенно очевидно, что работа с инновациями требует инновационного подхода и инновационных идей, и эта парадигма как никогда актуальна для науки права, для которой видение явлений под новым углом и выявление неясных на сегодняшний день взаимосвязей, станет краеугольным камнем научных исследований и базой для новых отраслей права.

Данная статья посвящена некоторым аспектам применения Искусственного Узкого Интеллекта (“Artificial Narrow Intelligence (ANI)”), который также именуют Искусственным Слабым (“Weak”) Интеллектом [4]. В основном, именно этот тип Искусственного Интеллекта на сегодняшний день находит свое применение во многих областях деятельности человека - от банковского дела до медицины, а также активно внедряется в правовую сферу. В частности, речь пойдет о вопросах формирования качественных и детальных данных (data) для алгоритмов и функциональных моделей машинного обучения с целью эффективного и непредвзятого применения технологий искусственного интеллекта в досудебном и судебном производстве.

Так, по мнению специалистов, 80% времени у создателей моделей машинного обучения и алгоритмов занимает именно работа с данными. Оставляя решение технических сложностей работы с входными данными (input data) специалистам – создателям моделей и алгоритмов, в этой статье автор попытается найти ответы на некоторые вопросы, касающиеся методов правового отбора и оценки входных данных, их группирования, с целью получения обоснованного, с точки зрения правовой базы и юридически значимых фактов, результата.

По версии корпорации Microsoft, «Модель машинного обучения (Machine Learning Model) – это файл, обученный для распознавания определенных типов шаблонов. Вы обучаете модель на основе набора данных, предоставляя им алгоритм, который можно использовать в качестве причины для получения и изучения этих данных. После обучения модели ее можно использовать для определения данных, которые ранее не встречались, и создания прогнозов об этих данных» [3].

Можно сделать вывод, что качество самих данных и разнообразие качественных наборов данных (data sets) – залог успешного обучения модели и создания аутентичного алгоритма, конечно же не без помощи математических и статистических знаний. Как результат, решение поставленной задачи априори минимализирует, а в идеале, и полностью исключает правовые ошибки и погрешности правового характера. Цепочка «качественные данные – аутентичный алгоритм – качественный результат», как итог работы специалистов различных профилей, должна гарантировать эффективное применение Искусственного Интеллекта в досудебном и судебном производстве на основе соблюдения прав человека, принципов права и норм закона.

Таким образом, представляется, что каждый этап, равно как и подэтапы вышеназванной цепочки, должен быть прописан в контексте вопроса «какова задача (цель)?». То есть, какая задача - такие и данные, какая задача – такой и алгоритм, какая задача – такой и результат. Следовательно, сбор, обработка, группирование и использование данных при работе

с функциональными моделями машинного обучения должны быть подчинены двум целям: первая цель – работа с фактическими и непосредственно относящимися к поставленной перед функциональной моделью задачи данными, и вторая цель – работа с данными сквозь призму соблюдения и обеспечения прав человека, принципов права, процессуальных прав участников уголовного процесса.

Допустим, если имеющиеся сведения-данные не влияют на признаки и элементы состава преступления [1, с. 24], то есть не представляют интереса в контексте предъявления конкретного обвинения или вынесения приговора и не влияют на принятие решения (decision-making) по делу, они не используются в модели «Состав преступления». Эти сведения могут представлять интерес для функциональных моделей машинного обучения с другими задачами, например с такой задачей модели, как «Предупреждение и прогнозирование преступлений», где сведения иного характера, в том числе полученные оперативным путем, будут востребованы и относимы к поставленной цели. Данные, относящиеся к статистике, психологии, социологии, виктомологии, медицине и к другим областям знаний должны найти свое место в соответствующих наборах данных. Например, данные о типах жертв преступления, которые не влияют на квалификацию, могут быть использованы в наборе данных по статистике, психологии или виктимологии, но не в модели «Состав преступления». То есть, суть данного метода состоит в том, что качество данных должно соответствовать поставленной задаче. Порой достаточно сложно определить границу относимости данных к определенной модели, эти границы могут быть достаточно нечеткими. Однако, именно детальная фильтрация и разбивка данных позволит создавать модели машинного обучения, отвечающие характеристике «качественные данные – аутентичный алгоритм – качественный результат».

Тем самым, чем больше наборов качественных и соответствующих поставленным задачам данных, тем беспристрастнее и эффективнее их применение и комбинирование с другими наборами данных. Такой же принцип сортировки и группировки данных, например, по задачам «Рецидив», «Совокупность преступлений», «Принудительное лечение», «Несовершеннолетние», «Процессуальные права», «Права человека» и т.п. даст возможность детально проработать данные по назначению, что, с одной стороны, практически исключает ошибки при работе алгоритма в рамках поставленной модели задачи, а с другой стороны, позволяет составлять «гибкие» комбинации наборов данных для решения разнообразных и еще более комплексных проблем. Вместе с тем, хотелось бы отметить, что ввиду исключительной правовой значимости вступивших в законную силу судебных постановлений и приговоров, систематизация этой категории данных требует отдельного углубленного исследования.

Более того, каждый качественный набор данных изначально блокирует возможность нарушения прав человека и процессуальных прав участников уголовного процесса посредством обеспечения относимости и беспристрастности используемых в инструментах ИИ данных. В свете этого, нелишне будет упомянуть требования статей 5 и 6 Европейской Конвенции по правам человека, где буквально каждое слово фокусируется на дефиниции конкретного права [2].

Итак, применение Искусственного Интеллекта в досудебном и судебном производстве должно осуществляться в соответствии с требованиями национального и международного законодательства о соблюдении прав человека и процессуальных прав подозреваемого, подследственного, обвиняемого, подсудимого и осужденного, в зависимости от его текущего

процессуального статуса, а также прав человека и личных прав жертв преступления, свидетелей и других участников уголовного процесса.

Необходимо отметить, что задачи функциональных моделей определяются юристами, так как именно они знают, какой правовой «продукт» нужно в итоге получить. Учитывая multifunctionality процесса выполнения задачи, вполне рационален подход, когда в постановке и осуществлении задачи обязательно участие юристов определенного профиля, причем тех специалистов, которые наделены способностью воспринимать, идентифицировать и тестировать на предмет правовой верификации и относимости к задаче модели машинного обучения сведения-данные из других областей знаний.

Таким образом, вышеобозначенные рассуждения позволяют прийти к выводу, что систематизация данных путем фильтрования с позиции защиты прав человека и процессуальных прав участников процесса (условно назовем этот процесс «объективным отбором») и, одновременно с этим, фильтрация в аспекте релевантности взятых данных к задаче, поставленной перед алгоритмом и моделью машинного обучения («субъективный отбор»), позволят значительно улучшить эффективность, точность и надежность работы инструментов искусственного интеллекта (AI tools), что в итоге значительно облегчит труд правоприменителей, способствуя снижению уровня их загруженности, ориентируя их на качество работы. Как следствие, естественно ожидание повышения доверия граждан к актуальным реформам досудебного и судебного производства, что и является одной из главных задач любого правового государства.

### Библиография

1. Комментарий к Уголовному Кодексу Азербайджанской Республики / под редакцией профессора Ф.Ю. Самандарова. - Баку. Изд-во Дигеста, 2006. - 1104 с. (на азерб. яз.)
2. Конвенция о защите прав человека и основных свобод: [Электронный ресурс]. – Режим доступа: [https://www.echr.coe.int/Documents/Convention\\_RUS.pdf](https://www.echr.coe.int/Documents/Convention_RUS.pdf) (дата обращения: 14.01.2020)
3. Что собой представляет модель машинного обучения: [Электронный ресурс]. – Режим доступа: <https://docs.microsoft.com/ru-ru/windows/ai/windows-ml/what-is-a-machine-learning-model> (дата обращения: 14.01.2020)
4. Types of AI: distinguishing between weak, strong, and super AI. [Электронный ресурс]. – Режим доступа: <https://www.thinkautomation.com/bots-and-ai/types-of-ai-distinguishing-between-weak-strong-and-super-ai/> (дата обращения: 17.01.2020).