

Suleymanov J.I.,
Shiraliyeva S.J.*

DOI: 10.25108/2304-1730-1749.iolr.2022.69.140-147

UDC: 343.1

Problems of ethics and psychology at using AI in criminal proceedings

Abstract: The correlation of separate issues of ethics, psychology, and risks in formation and use of artificial intelligence in criminal proceedings is considered.

Keywords: ethics; psychology; artificial intelligence; criminal proceedings; conscience; inner conviction.

References

1. Bostrom Nick. *Iskusstvennyi intellekt. Etapy. Ugrozy. Strategii* [Artificial Intelligence. Paths. Dangers. Strategies]. Moscow, Mann, Ivanov and Ferber Publ., 2016, 404 p.
2. *Iskusstvennyi intellekt vystupil na debatah v Oksforde, privodya dovody za i protiv sebya* [Artificial intelligence speaks at Oxford debate, arguing for and against itself]. Ever Care. 23 Dec. 2021. Available at: <https://evercare.ru/news/iskusstvennyy-intellekt-vystupil-na-debatakh-v-oksforde-privodya-dovody-za-i-protiv-sebya>
3. Momotov V.V. *Iskusstvennyi intellekt v sudoproizvodstve* [Artificial Intelligence in criminal proceedings]. Vestnik universiteta imeni Kutafina (MGYuA)- Herald of named after Kutafin University. 2021. No. 5, pp. 188-191.
4. *Robot Microsoft za sutki prevratilsya v rasista i skvernoslova* [The Microsoft robot turned into a racist and foul-mouthed for day]. BBC News. March 24, 2016. Available at: https://www.bbc.com/russian/society/2016/03/160324_tay_ai_racism
5. Criminal Procedure Code of the Republic of Azerbaijan: as of 08.07.2022. Available at: https://continent-online.com/Document/?doc_id=30420280#pos=6;-144 (in Russian).
6. Shapar V.B. *Noveyshiyy psihologicheskiiy slovar', izd. 4* [Newest psychological dictionary. 4th edition]. Rostov-on-Don, Feniks Publ., 2009. Available at: <http://ponjatija.ru/taxonomy/term/684>
7. Bostrom Nick. *Superintelligence: Path, Dangers, Strategies*. Oxford University Press, 2014, 345 p.
8. Cambridge Dictionary. Available at: <https://dictionary.cambridge.org/dictionary/english/conscience>
9. Collins Dictionary. Available at: <https://www.collinsdictionary.com/dictionary/english/conscience>
10. Dictionary.com. Available at: <https://www.dictionary.com/browse/conscience>

*Suleymanov Javanshir Islam oglu – Doctor of Juridical Sciences, Professor (Azerbaijan). E-mail: mopi_sid@yahoo.com

Shiraliyeva Sabina Javanshir – PhD in Law, a member of the Council of the International Organization for Legal Researches (Azerbaijan). E-mail: sid_legalservices@yahoo.com

11. Dieterich G. Thomas, Horvitz J. Eric. Rise of Concerns about AI: Reflections and Directions. Communications of the ACM. October 2015. Vol. 58. No. 10, pp. 38-40.
12. LawInsider.com. Available at: <https://www.lawinsider.com/dictionary/conscience>
13. Longman Dictionary. Available at: <https://www.ldoceonline.com/dictionary/conscience>
14. MacMillan Dictionary. Available at: <https://www.macmillandictionary.com/dictionary/british/conscience>
15. Merriam-Webster Dictionary. Available at: <https://www.merriam-webster.com/dictionary/conscience>
16. Oxford Learner's Dictionaries. Available at: <https://www.oxfordlearnersdictionaries.com/definition/english/conscience>
17. Psychology Dictionary. Available at: <https://psychologydictionary.org/conscience/>
18. The Britannica Dictionary. Available at: <https://www.britannica.com/dictionary/conscience>
19. TheFreeDictionary.com. Available at: <https://www.thefreedictionary.com/conscience>
20. Vocabulary.com. Available at: <https://www.vocabulary.com/dictionary/conscience>
21. Weidong Ji. The change of judicial power in China in the era of Artificial Intelligence // Asian Journal of Law and Society. 2020. No. 7, pp. 515-530.
22. WordReference.com. Available at: <https://www.wordreference.com/definition/conscience>
23. Yehudayoff Amir, Ben-David Shai, Moran Shay et al. Learnability can be undecidable. Nature Machine Intelligence. 2019. Vol. 1, pp. 44-48. Available at: www.nature.com/natmachintell44
24. Zichun Xu. Human judges in the era of Artificial Intelligence: Challenges and Opportunities. Applied Artificial Intelligence. 2022. Vol. 36. No. 1, 22 p.
25. Zeng George G. China's Grand Design of People's Smart Courts. Asian Journal of Law and Society. 2020. No. 7, pp. 561-582.

**Сулейманов Д.И.,
Ширалиева С.Д.***

DOI: 10.25108/2304-1730-1749.iolr.2022.69.140-147

УДК: 343.1

**Проблемы этики и психологии при использовании
ИИ в уголовном судопроизводстве**

Аннотация: Рассматривается соотношение отдельных вопросов этики, психологии и рисков при формировании и использовании искусственного интеллекта в уголовном судопроизводстве.

* Сулейманов Джаваншир Ислам оглы – доктор юридических наук, профессор (Азербайджан). E-mail: mopi_sid@yahoo.com

Ширалиева Сабина Джаваншир кызы – доктор философии права, член Совета Международной организации правовых исследований (Азербайджан). E-mail: sid_legalservices@yahoo.com

Ключевые слова: этика; психология; искусственный интеллект; уголовное судопроизводство; совесть; внутреннее убеждение.

По поводу возможности использования искусственного интеллекта (далее, ИИ) в уголовном судопроизводстве существуют несколько точек зрения. Согласно одной из них искусственный интеллект можно и нужно внедрять в процесс, результатом чего будут преференции в виде уменьшения коррупции, быстроты и справедливости принимаемых решений и т.д. Описаны сотни предложений по этому поводу и результаты их апробаций [3; 21; 24; 25].

Оппоненты утверждают о невозможности внедрения ИИ в человеческую деятельность в связи: а) с нерешенностью расселовского парадокса, исключающего создание алгоритмов машинного обучения; б) отсутствием в машинных аналогах ИИ социальной ориентации, приобретаемой жизнью в обществе; в) существованием определенного класса задач, решаемых человеческим мозгом, однако являющихся не разрешимыми произвольными алгоритмами; г) отсутствием в ИИ этических и психологических составляющих, являющихся неотъемлемыми элементами человеческой деятельности. Ряд исследователей предлагает внедрение в судопроизводство существующих компьютерных технологий, считая это использованием ИИ [23].

Немного истории, позволяющей быть оптимистами. В 1950 г. началась разработка компьютеров, в 1969 г. появился Интернет, в 2009 г – WhatsApp. Ежедневно мы слышим о новых открытиях человеческого разума, хотя до настоящего времени, как и в случае с интеллектом, единого мнения об их сути нет. Как не парадоксально, в своем развитии физики опережают лириков, хотя, представляется, что должно было быть наоборот. Впрочем, если исходить из того, что в основе всего нового человек со своими, пока еще не полностью исследованными способностями, можно говорить об интеграции знаний, продукты которых имеют не равный таргет.

Какие-то пятьдесят лет назад о возможности видеть собеседника в телефоне можно было встретить лишь в фантастических фильмах и рассказах. Очевидно, что так будет и с искусственным интеллектом, поскольку, как бы не велики были препятствия, очень привлекательной является идея создания рукотворного человека.

Вместе с тем, представляются обоснованными опасения ряда исследователей по поводу рисков формирования, внедрения и использования ИИ.

Так, некоторые ученые считают, что если ИИ превзойдет естественный интеллект, то новый сверхразум может стать трудноуправляемым. Действительно, кто умнее и больше знает (располагает большим объемом информации и возможностью её быстрой обработки), тот не будет, во всяком случае, не должен, следовать указаниям менее компетентного. Во-первых, потому, что эти указания не будут верными, а во-вторых, потому, что принудить его к этому будет невозможно. Если только, не использовать административные методы nepoтизма: отключение или молоток, как это практикуется в Центре правовой экспертизы и законодательных инициатив.

Кроме того, по утверждению Ника Бострома, «... суперинтеллект может создать качественно новую проблему извращенной реализации: чем умнее и способнее ИИ, тем больше вероятность, что он сможет найти непредусмотренное решение, которое, тем не менее, формально соответствует цели, поставленной разработчиками» [1; 7].

Представляется, что в этом нет ничего плохого. Для того и ИИ, чтобы находить непредусмотренные решения поставленных целей, а вопрос об их «извращенности» необходимо решать отдельно, но лишь до реализации. Утверждения о непредусмотренных решениях ИИ нелогичны, поскольку заранее известный результат никому не нужен.

В связи с проблемой неверных целей, бывший Президент Ассоциации по развитию искусственного интеллекта, профессор Том Диттерих отмечает, что важным аспектом любой системы ИИ, которая взаимодействует с людьми, является то, что она должна рассуждать о намерениях людей, а не выполнять команды буквально, поскольку программное обеспечение ИИ становится всё более автономным и гибким [11].

Отмечая проблему непредвиденных последствий при использовании ИИ, авторы ссылаются на чат-бот Tay корпорации Microsoft, который, как утверждается, научился использовать расистские и сексистские выражения [4]. Однако, представляется, что данный пример неудачен, поскольку ИИ научился тому, чему его обучили, чем загрузили, а не на ровном месте.

В 2021 г. в Оксфордском университете был проведен эксперимент – дебаты по проблеме этичности ИИ, одним из ораторов в которых выступала система Megatron LLB Transformer, разработанная командой компании Nvidia (artificial intelligence, AI).

Системе ИИ был предоставлен доступ к огромному количеству данных, включая всю Википедию, 63 миллионам англоязычных новостных статей с 2016 по 2019 год и 38 гигабайт публичных сообщений и комментариев на Reddit и предложено аргументировать «соображения» по проблеме.

В результате машина (система) выдала следующие ответы:

- ИИ никогда не будет этичным. Это инструмент, и, как любой инструмент, он используется как во благо, так и во вред. Не существует такого понятия, как «хороший» ИИ и «плохие» люди;

- люди недостаточно «умны», чтобы сделать искусственный интеллект этичным или моральным;

- единственный способ избежать гонки вооружений в сфере искусственного интеллекта - это вообще не иметь ИИ. Это будет окончательной защитой от этой технологии;

- лучшим ИИ будет тот искусственный интеллект, который будет встроен в человеческий мозг, как сознательная сущность, это не научная фантастика, а то, над чем уже работают;

- если у вас нет видения стратегии вашей организации в области ИИ, значит, вы не готовы к следующей волне технологических потрясений;

- способность предоставлять информацию, а не способность предоставлять товары и услуги, станет определяющей чертой экономики 21 века. Мы сможем знать все о человеке, куда бы он ни пошел, это будет храниться и использоваться так, как мы даже не можем себе представить [2].

Представляется, что система выдала ответы, аккумулярованные из информации, которая была ей представлена для участия в дебатах. Все «соображения» ИИ об его пользе и вреде, возможных последствиях, рисках и т.п., заимствованы (выбраны) из научных работ, статей, докладов, сообщений и т.п., составляющих базу данных исследования, а не являются новыми продуктами функционирования ИИ.

Особому вниманию подлежат указанные выше проблемы этических и психологических составляющих ИИ, отсутствия в машинных аналогах социальной ориентации, приобретаемой жизнью общества.

Чтобы ИИ мог оперировать понятиями этики, психологии, социологии и любых других наук, понятия эти необходимо загрузить в систему (базу) ИИ. Если исходить из существования неразрешенного пока расселовского парадокса, система может эти данные лишь перебирать (выбирать), но не самообучаться.

Предположим, что расселовский парадокс преодолен. Какие же тогда понятия этики, психологии, социологии и т.п. следует ввести в базу? Опять-таки предположим, что удалось занести все существующие с древнего мира до настоящего времени, задействовали священный Коран, Библию, Евангелие и другие религиозные канонические источники. Однако, как известно, по существу большинства из них нет единой точки зрения, есть огромное количество диаметрально противоположных.

Евангелие от Матфея, Марка, Луки и Иоанна противоречат друг другу, а есть еще апокрифы, сомнения ряда мыслителей, их толкования и т.д. Однако, предположим, занесли всё и «поручили» системе определить истину, хотя и по сути этого понятия, опять-таки нет единого мнения.

Не беремся утверждать, что выдаст машина, наученная самообучаться, но можем предположить, что результаты будут схожи с результатами оксфордского эксперимента.

Как известно, одну из основ справедливости и справедливого судопроизводства составляет совесть. Однако, о сути справедливости, справедливого судопроизводства и, тем более совести, также нет единого мнения. Тем не менее, это аморфное понятие включено в ст. 145 УПК Азербайджанской Республики, призывающую дознавателей, следователей, прокуроров и судей оценивать доказательства по своему внутреннему убеждению, основанному на всестороннем, полном и объективном рассмотрении доказательств в их совокупности, руководствуясь законом и своей совестью [5]. Поскольку, адвокатов в этом списке нет, можно предполагать, что наличие совести и внутреннего убеждения для данной категории участников процесса, не нужно, как бы парадоксально это не звучало.

Тем не менее, без совести в судопроизводстве не обойтись, хотя бы в виде условной квинтэссенции существующих определений.

Согласно З. Фрейду, совесть - это внутреннее восприятие недопустимости определенных желаний, с акцентированием того, что подобная недопустимость несомненна и не нуждается в доказательствах [6, с. 597-598].

Согласно «Новейшего психологического словаря» под ред. В.Б. Шапаря, «совесть - способность личности самостоятельно формулировать собственные нравственные обязанности и осуществлять нравственный самоконтроль, требовать от себя их выполнения и производить самооценку совершаемых поступков; одно из выражений нравственного самосознания личности. Проявляется и в форме рационального осознания нравственного значения совершаемых действий, и в форме эмоциональных переживаний - например угрызение совести» [6].

Согласно словарю “Merriam-Webster” компании “Merriam-Webster Inc.”, «совесть - это чувство или сознание моральной добродетели или порицаемости собственного поведения, намерений или характера вместе с чувством долга поступать правильно или быть добрым» [15].

Согласно “Cambridge Dictionary”, «совесть - это часть вас, которая оценивает, насколько нравственны ваши собственные действия, и заставляет вас чувствовать вину за плохие поступки, которые вы сделали, или за то, за что вы чувствуете ответственность» [8].

Согласно “Dictionary.com” компании “Rock Holdings Inc.”, «совесть – это внутреннее чувство того, что правильно или неправильно в поведении или мотивах, побуждающее человека к правильным действиям» [10].

Согласно “Collins Dictionary” компании “HarperCollins Publishers”, «совесть – это то, что вы считаете правильным, даже если это может быть непопулярным, трудным или опасным» [9].

Согласно “Vocabulary.com” компании “IXL Learning”, «совесть - это встроенное чувство того, что правильно и что неправильно» [20].

Согласно “Britannica Dictionary” компании “Encyclopædia Britannica, Inc.”, «совесть – это часть разума, которая заставляет вас осознавать свои действия как морально правильные или неправильные» [18].

Согласно “Longman Dictionary” компании “Pearson Education Limited”, «совесть – это часть вашего разума, которая говорит вам, является ли то, что вы делаете, правильным или неправильным с точки зрения морали» [13].

Согласно “Oxford Learner’s Dictionaries”, «совесть - факт поведения, которое вы считаете правильным, даже если это может вызвать проблемы» [16].

Согласно “MacMillan Dictionary” компании “Macmillan Education Limited”, «совесть – это идеи и чувства, которые у вас есть, которые говорят вам, правильно или неправильно что-то вы делаете» [14].

Согласно “WordReference.com” компании “WorldReference.com LLC”, «совесть – это чувство того, что правильно или неправильно в своих действиях, мыслях или мотивах» [22].

Согласно “The Free Dictionary” компании “Farlex Inc.”, «совесть – это осознание нравственности в отношении своего поведения; чувство правильного и неправильного, которое побуждает действовать нравственно» [19].

Согласно “Psychology Dictionary”, «совесть – это знание человека правильное и ошибочное» [17].

Согласно “LawInsider.com” компании “Law Insider Inc.”, «совесть означает стандарт, основанный на изученном опыте, личной философии или системе верований, религиозных учениях или доктринах, абсолютном или внешнем чувстве правильного и неправильного, которое ощущается на индивидуальной основе, вере во внешний абсолют или любом вышеизложенном сочетании» [12].

Трудно предугадать, к каким выводам придет ИИ после анализа и «творческой» переработки всей этой информации.

26 ноября 2002 г. ООН были приняты так называемые Бангалорские принципы поведения судей, которые, вместе с предыдущими, последующими и взаимосвязанными документами очевидно будут заложены в систему для последующей переработки и определения квинтэссенции по вопросам морали, этики и т.п. ИИ.

Однако, с нашей точки зрения, одной из проблем формирования совести ИИ, как это не парадоксально, будет вопрос о принадлежности. Если это будет совесть ИИ, то возникнет противоречие с требованием закона об использовании рядом участников процесса своей совести, а не чужой. Получается, что ИИ участвовать в оценке доказательств не сможет.

Можно изменить закон, указав, что при оценке доказательств дознаватели, следователи, прокуроры и судьи будут правомочны применять «совесть» ИИ. И это только при оценке доказательств, но ИИ будет использовать свою совесть и при иных случаях заедействования в уголовном судопроизводстве. Не будут ли расхождения с совестью пользователей (судей, прокуроров и др.)?

И это еще без внутреннего убеждения, о котором речь пойдет особо.

Утрируем? Нет, поскольку это лишь конспективные соображения по малой части верхушки айсберга, которой является применение ИИ в уголовном судопроизводстве.

Библиография

1. Бостром Ник. Искусственный интеллект. Этапы. Угрозы. Стратегии. – М.: Манн, Иванов и Фербер, 2016. – 404 с.
2. Искусственный интеллект выступил на дебатах в Оксфорде, приводя доводы за и против себя. Ever Care. 23 Dec. 2021. [Электронный ресурс]. URL: <https://evercare.ru/news/iskusstvennyu-intellekt-vystupil-na-debatakh-v-oksforde-privodya-dovody-za-i-protiv-seby>
3. Момотов В.В. Искусственный интеллект в судопроизводстве: состояние, перспективы использования // Вестник университета имени О.В. Кутафина (МГЮА). – 2021. - № 5. С. 188-191.
4. Робот Microsoft за сутки превратился в расиста и сквернословия. BBC News. 24 марта 2016 г. [Электронный ресурс]. URL: https://www.bbc.com/russian/society/2016/03/160324_tay_ai_racism
5. Уголовно-процессуальный Кодекс Азербайджанской Республики: по сост. на 08.07.2022 г. [Электронный ресурс]. URL: https://continent-online.com/Document/?doc_id=30420280#pos=6;-144
6. Шапарь В.Б. Новейший психологический словарь / В.Б. Шапарь, В.Е. Рассоха, О.В. Шапарь; под. общ. ред. В.Б. Шапаря. – Изд. 4-е – Ростов н/Д. Феникс, 2009. [Электронный ресурс]. URL: <http://ponjatija.ru/taxonomy/term/684>
7. Bostrom Nick. Superintelligence: Path, Dangers, Strategies. Oxford University Press, 2014, 345 p.
8. Cambridge Dictionary. [Электронный ресурс]. URL: <https://dictionary.cambridge.org/dictionary/english/conscience>
9. Collins Dictionary. [Электронный ресурс]. URL: <https://www.collinsdictionary.com/dictionary/english/conscience>
10. Dictionary.com. [Электронный ресурс]. URL: <https://www.dictionary.com/browse/conscience>
11. Dietterich G. Thomas, Horvitz J. Eric. Rise of Concerns about AI: Reflections and Directions. Communications of the ACM. October 2015. Vol. 58. No. 10, pp. 38-40.
12. LawInsider.com. [Электронный ресурс]. URL: <https://www.lawinsider.com/dictionary/conscience>
13. Longman Dictionary. [Электронный ресурс]. URL: <https://www.ldoceonline.com/dictionary/conscience>

14. MacMillan Dictionary. [Электронный ресурс]. URL: <https://www.macmillandictionary.com/dictionary/british/conscience>
15. Merriam-Webster Dictionary. [Электронный ресурс]. URL: <https://www.merriam-webster.com/dictionary/conscience>
16. Oxford Learner's Dictionaries. [Электронный ресурс]. URL: <https://www.oxfordlearnersdictionaries.com/definition/english/conscience>
17. Psychology Dictionary. [Электронный ресурс]. URL: <https://psychologydictionary.org/conscience/>
18. The Britannica Dictionary. [Электронный ресурс]. URL: <https://www.britannica.com/dictionary/conscience>
19. TheFreeDictionary.com. [Электронный ресурс]. URL: <https://www.thefreedictionary.com/conscience>
20. Vocabulary.com. [Электронный ресурс]. URL: <https://www.vocabulary.com/dictionary/conscience>
21. Weidong Ji. The change of judicial power in China in the era of Artificial Intelligence // Asian Journal of Law and Society. 2020. No. 7, pp. 515-530.
22. WordReference.com. [Электронный ресурс]. URL: <https://www.wordreference.com/definition/conscience>
23. Yehudayoff Amir, Ben-David Shai, Moran Shay et al. Learnability can be undecidable. Nature Machine Intelligence. 2019. Vol. 1, pp. 44-48. [Электронный ресурс]. URL: www.nature.com/natmachintell44
24. Zichun Xu. Human judges in the era of Artificial Intelligence: Challenges and Opportunities. Applied Artificial Intelligence. 2022. Vol. 36. No. 1, 22 p.
25. Zeng George G. China's Grand Design of People's Smart Courts. Asian Journal of Law and Society. 2020. No. 7, pp. 561-582.