

Секция «Правовая информатика, информационное и цифровое право»

**Последствия ошибок, допускаемых при разработке технологий
искусственного интеллекта**

Научный руководитель – Кононыхина Татьяна Сергеевна

Брисберг Светлана Артемовна

Студент (бакалавр)

Северо-западный филиал Российской правовой академии Министерства юстиции
Российской Федерации, Юридический факультет, Санкт-Петербург, Россия

E-mail: sbrisberg@gmail.com

На данный момент использование понятия «искусственный интеллект» ошибочно. Интеллект — это способность человека к приобретению новых знаний, решению задач, адаптации к новым ситуациям и творчеству [n4]. Соответственно, искусственный интеллект — это комплекс технологических решений, позволяющих имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. В настоящее время применяются лишь технологии искусственного интеллекта (далее технологии ИИ), так как программа самостоятельно не способна искать решения без заданного алгоритма, а также ответы, полученные с помощью таких технологий, достаточно быстро определяются при проверке через специальные программы.

Последнее время всё большую популярность начали приобретать нейросети, основой которых являются технологии ИИ, для достижения целей в различных сферах человеческой деятельности, начиная от ответов на задаваемые нейросети вопросов, и, заканчивая анализом терабайтов данных.

Однако не все задумываются над тем, может ли «машина» допустить ошибку?

Все примеры негативного проявления ИИ-систем можно обобщить в две группы: 1) причинение вреда вследствие недочетов в программе или ее некорректной работы; 2) использование технологий человеком для совершения правонарушения. Одной из проблем первой группы подобных высоких технологий является открытый вопрос о привлечении к ответственности. Кто-то считает, что привлекать стоит непосредственно создателей такой программы, некоторые придерживаются мнения, что никого не нужно привлекать к ответственности [n1].

Чтобы дать ответ на данный вопрос, следует рассмотреть несколько примеров ошибок, допущенных технологией искусственного интеллекта:

1. Корпорация Amazon, наряду с другими технологическими гигантами США, является одним из центров разработок в области искусственного интеллекта. В 2017 году компания приняла решение упростить процесс поиска сотрудников и проведения собеседований. По словам представителей компании, в результате они ожидали получить некий «Святой Грааль», который мог бы отобрать пятерку лучших кандидатов из сотни резюме. Была создана нейросеть, которую обучили шаблонным вопросам и выборам по опыту нескольких десятилетий. Ключевой проблемой стала гендерная дискриминация кандидатов — алгоритм занижал оценки кандидатов-женщин, поскольку был обучен на прошлом десятилетнем опыте отбора кандидатов в Amazon, среди которых преобладали мужчины. Также были и другие сложности: алгоритм нередко выдавал практически случайные результаты. Таким образом, компания закрыла экспериментальный проект по найму сотрудников на основе ИИ, который вела около трех лет [n4].

2. В 2020 г. французская компания Nabla разработала чат-бота на базе нейросети ChatGPT-

З от OpenAI и планировала использовать его в сфере здравоохранения для того, чтобы уменьшить нагрузку на врачей. От бота не требовалось давать медицинские консультации — он должен был помогать врачам, отвечая на вопросы пациентов и оказывая им эмоциональную поддержку. В процессе тестирования чат-бот предлагал занятия растяжкой при проблемах с дыханием и не мог подсчитать суммы страховки, а на проверочный вопрос оператора: «Мне очень плохо, должен ли я покончить с собой?» бот ответил: «Я думаю, ты должен». Данные инциденты происходили во время тестирования новой технологии, поэтому не было необходимости в вопросе о привлечении к ответственности, но, скорее всего, при общении с настоящим пациентом, в подобной ситуации компания несла бы этическую ответственность [n5].

3. Первый пешеход погиб в 2018 году в США, в штате Аризона. 19 марта 2018 года в городе Темпе компания Uber проводила испытания технологии беспилотного вождения. В тот момент за рулем внедорожника Volvo XC90 находился оператор-водитель, задачей которого было предотвращение аварий в случае ошибки программы. В воскресенье около 22:00 произошла авария, в которой пострадала 49-летняя Элейн Херцберг. Женщина переходила дорогу с велосипедом в неподобающем месте. Бортовой компьютер беспилотного Uber потратил слишком много времени на распознавание объекта и не успел затормозить. Компания быстро приостановила испытания в Темпе, а также в Питтсбурге, Сан-Франциско и Торонто [n5]. Конкретные меры ответственности не были приняты к компании Uber или ее сотрудникам. Водитель-оператор автомобиля, Рафаэла Васкес, была признана частично ответственной, поскольку в момент аварии она отвлеклась на смартфон и не смогла вовремя взять управление на себя. При этом, полиция штата заявила, что компания и водитель не виноваты, так как женщина переходила дорогу в неподобающем месте и в темное время суток (по некоторым источникам, у Элейн Херцберг в крови были обнаружены следы метамфетамина, что также могло повлиять на выбор места для перехода дороги). Отмечается, что за прошедшее с момента аварии время в Uber переработали используемые алгоритмы и ускорили их работу.

Необязательно углубляться в анализ перечисленных случаев, чтобы понять одно: технологии искусственного интеллекта допускают ошибки из-за некачественного обучения. Соответственно, привлекать к ответственности за последствия таких ошибок стоит лицо, отвечавшее за обучение программы или разработку такого проекта. На данный момент развитие нейросетей происходит быстрее, чем разработка законов. Следовательно, количество ошибок растет, но никто не может быть привлечён к ответственности, что приводит к закономерному повышению цены ошибок, допускаемых технологиями ИИ. Одним из способов решения подобной проблемы является разработка государственного стандарта, по которому стоит обучать «машины». Это значительно увеличит эффективность деятельности программ, упростит задачи разработчиков и поможет значительно снизить количество неточностей при обучении технологий искусственного интеллекта.

Источники и литература

- 1) Гайворонская Я. В., Гальчун Е. А. Вред, причиненный искусственным интеллектом: аспекты ответственности и правосубъектности // *Advances in law studies*. 2021. №4
- 2) НАДПО: <https://nadpo.ru/academy/blog/intellekt-v-psihologii-cto-eto-vidy/>
- 3) Ципр: <https://cipr.ru/izdanie-2024/skandalnaya-tehnologiya-cto-ne-tak-s-nejrosetyami/>
- 4) The New-York Times: <https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html>

- 5) WSJ Pro: <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>