Конспект лекции

С.С. Жумажанова, П.С. Ложников

Этика и правовые проблемы искусственного интеллекта

**Принципы искусственного интеллекта**

**ПРИНЦИПЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

***Ответственность частных лиц и организаций при негативных последствиях, возникающих в результате проектирования, разработки и использования ими систем на базе ИИ*:**

* не допускать передачи полномочий ответственного нравственного выбора СИИ, не делегировать ответственность за последствия принятия решений СИИ – за все последствия работы СИИ всегда отвечает человек;
* Акторы ИИ должны ответственно относиться к вопросам влияния СИИ на общество и граждан на каждом этапе жизненного цикла СИИ, включая неприкосновенность частной жизни, этичное, безопасное и ответственное использование персональных данных, к характеру, степени и размеру ущерба, который может последовать в результате использования технологий и СИИ, а также при выборе и использовании аппаратных средств и программного обеспечения, задействованных на различных жизненных циклах СИИ. При этом ответственность Акторов ИИ должна соответствовать характеру, степени и размеру ущерба, который может последовать в результате использования технологий и систем ИИ.

***Справедливость и исключение дискриминации при использовании технологий ИИ для различных социальных групп:***

* в целях обеспечения справедливости и недопущения дискриминации Акторы ИИ должны принять доступные меры для того, чтобы удостовериться, что применяемые ими наборы данных, методы обработки используемых для машинного обучения данных, при помощи которых осуществляется группирование и/или классификация данных, касающихся отдельных лиц или групп лиц, не влечет их умышленную дискриминацию по признакам расовой, национальной принадлежности, политических взглядов, религиозных или философских убеждений или сведений об интимной жизни;

***Прозрачность и объяснимость работы ИИ, направленная на то, чтобы пользователи понимали, как ИИ принимает решения, как система ИИ была разработана и протестирована, чтобы гарантировать, что она работает так, как изначально задумано:***

* отслеживаемость информации об алгоритмах, обучающих данных и пользовательских данных, в том числе о том, как они были собраны;
* раскрытие методов оценки и показателей, используемых для проверки работы системы.
* объяснение пользователям входных данных, которые использовались для принятия решения;
* как можно больше объяснять пользователям, как система ИИ пришла к решению.
* уведомление пользователей, когда решение о них принимает система ИИ;
* уведомление пользователей, когда они взаимодействуют с системой ИИ.

***Профессиональная ответственность проектировщиков и разработчиков систем ИИ:***

* профессионалы, занимающиеся проектированием, разработкой или развертыванием продуктов или систем на основе ИИ, должны осознавать свое уникальное положение, позволяющее оказывать влияние на людей, общество и будущее ИИ.
* требуется профессионализм и честность в обеспечении ответственного развития систем искусственного интеллекта и внесения своего опыта в общественное обсуждение, образование и управление.

*Использование ИИ с целью пропаганды человеческих ценностей и принесения максимальной пользы обществу:*

* ИИ должен внедряться и использовался таким образом, чтобы приносить максимальную пользу обществу, способствовать благополучию человечества и поощрять человеческое процветание.
* ИИ можно развернуть таким образом, чтобы уважать и согласовывать с социальными нормами, интересами человечества и основными культурными убеждениями и ценностями.
* ряд уникальных человеческих ценностей включают (но не ограничиваются ими) честность, свободу и уважение. Поощрение человеческих ценностей в процессе разработки, внедрения и развертывания ИИ имеет решающее значение.

***Использование систем ИИ с учетом обеспечения защиты сведений ограниченного доступа:***

* использование данных, полученных, собранных, используемых или раскрытых только для выполнения намеченных целей и задач;
* сообщение о цели обработки информации, позволяющей установить личность, и о ее передаче;
* согласие: данные субъектов не должны собираться и обрабатывать без их ведома или разрешения;
* контроль за использованием данных: контроль лиц за использованием их личной информации;
* степень влияния лиц на то, как и почему используется их информация;
* возможность ограничивать обработку данных: право лиц ограничивать сбор или использование данных в связи с внедрением ИИ;
* исправление: позволить лицам изменять информацию, если она неверна;
* удаление: лица могут удалять личные данные из системы и приложения ИИ;
* предоставление лицам возможности просматривать личные данные, используемые системой и приложением ИИ;
* конфиденциальность: учет соображений конфиденциальности данных при разработке систем и приложений искусственного интеллекта и на протяжении всего жизненного цикла использования данных;
* разрешение споров: механизмы для разрешения споров в отношении этих функций.
* надежность: системы искусственного интеллекта, которые работают стабильно, как задумано, являются ключевыми для этой темы;
* тестирование и мониторинг: может помочь свести к минимуму неправильное использование и злоупотребление системой ИИ.

***Соблюдение принятых законов и нормативно-правовых актов, международных норм поведения:***

* соблюдение законов и постановлений, даже если они не соблюдаются в данной юрисдикции;
* соблюдение всех юридических обязательств по всей цепочке создания стоимости искусственного интеллекта и периодический анализ соответствия деятельности и взаимоотношений заинтересованных сторон;
* обеспечение того, чтобы цели, для которых разрабатывается и используется ИИ, были законными и конкретными.

Одним из этических стандартов поведения Акторов в сфере ИИ является недопущение передачи полномочий ответственного нравственного выбора СИИ, недопустимость делегирования ответственность за последствия принятия решений СИИ – **за все последствия работы СИИ всегда отвечает человек**.

Акторы ИИ должны знать, учитывать и ориентироваться на положения законодательства РФ во всех сферах своей деятельности, в том числе вопросах юридической ответственности Акторов ИИ и порядка работы с персональными данными и информацией.

Необходимо четкого, строгого и эффективно определить этические и правовые рамки при разработке, проектировании, производстве, использовании и модификации искусственного интеллекта.

Уголовно-правовое регулирование ИИ возможно в случаях:

* ошибок при создании ИИ;
* неправомерного доступа злоумышленников в систему ИИ, после чего стало возможным совершение преступления;
* самообучения ИИ и принятия им решения о совершении действий (бездействия), повлекших за собой совершение преступления;
* ИИ был разработан злоумышленником, совершившим с помощью него преступление.

Имеются обоснованные сомнения относительно эффективности применения к ИИ традиционных видов уголовного наказания (штраф или лишение свободы). С точки зрения западных юристов, для ИИ следует предусмотреть специальные виды наказания (деактивация, перепрограммирование или наделение статусом преступника, которое будет служить предупреждением для участников правоотношений).

**Список источников**

1. Балашов А.И., Рудаков Г.П. Правоведение: учебник для вузов. 6-е изд., дополненное и переработанное. Стандарт третьего поколения. Издательский дом «Питер» 2015. 544 p.
2. Кодекс этики в сфере искусственного интеллекта [electronic resource]. URL: <https://a-ai.ru/wp-content/uploads/2021/10/кодекс_этики_в_сфере_ии_финальный.pdf>.
3. «Конвенция о защите физических лиц при автоматизированной обработке персональных данных» (заключена в г. Страсбурге 28.01.1981) (вместе с поправками к конвенции о защите физических лиц при автоматизированной обработке персональных данных (СДСЕ №108), позволяющими присоединение европейских сообществ, принятыми Комитетом Министров в Страсбурге 15.06.1999) / КонсультантПлюс [electronic resource]. URL: <https://www.consultant.ru/document/cons_doc_law_121499/>.
4. Резолюция Европарламента от 16 февраля 2017 года. Нормы гражданского права о робототехнике [electronic resource]. URL: <https://robopravo.ru/riezoliutsiia_ies>.
5. Европейская этическая хартия об использовании искусственного интеллекта в судебных системах и окружающих их реалиях [electronic resource]. URL: <https://rm.coe.int/ru-ethical-charter-en-version-17-12-2018-mdl-06092019-2-/16809860f4>.
6. Юнеско рекомендации об этических аспектах искусственного интеллекта (ноябрь 2021 г.) [electronic resource]. URL: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahukewir-qlc-_n2ahxfk4skhes2dyuqfnoecamqaq&url=https%3a%2f%2fircai.org%2fwp-content%2fuploads%2f2020%2f07%2frecommendation_first_draft_rus.pdf&usg=aovvaw1hp4o7y9bsdjwdlcnip-1->.