

## Интеллектуальные и автономные системы — этические аспекты применения

Готовцев П.М.

НИЦ Курчатовский институт

Искусственный интеллект — это словосочетание сегодня звучит в совершенно разном контексте. Его используют как для фантастических сверхразумных машин в кинематографе, так и для продвижения «умных» приложений в обычные смартфоны. Хотя, говоря технически корректно некоторые из таких приложений могут попасть по термину интеллектуальные и автономные системы (ИИ/АС) — именно такая терминология используется применительно к многочисленным «умным» системам, которые так активно входят в нашу жизнь в последние годы. Основной особенностью таких систем, является их способность автономно принимать решения, опираясь имеющиеся данные. Такой системой может быть, например, автопилот автомобиля, который принимает решения как, с какой скоростью и по какому маршруту ехать. То же можно сказать о программах, которые анализируют наше поведение в интернете — активность в социальных сетях, поисковые запросы, местоположение по комментариям под фотографиями и т.д. Эти программы обработав данные дают пользователям советы, конечно, право пользователей следовать или не следовать этим советам, но очень часто получается, что советы оказываются весьма полезными. Таким образом, мы сталкиваемся с ситуацией, когда машины принимают решения, которые непосредственно оказывают влияние на нашу жизнь, и по мере дальнейшего развития технологий мы в праве ожидать что областей применения таких «умных» машин будет становится все больше. Более того, сегодня наиболее ценные и интересные результаты во многих областях от геномных исследований до интернета вещей получаются там, где проводится анализ больших данных. То есть машины могут принимать решения там, где человек уже не может — при анализе больших данных. Итак, можно сказать, что сегодня взаимодействие человека и машин обретает новые особенности. Еще раз обратим внимание на то, что речь идет не о сверхразумной искусственном интеллекте, который сейчас активно обсуждается философами и футурологами [1][2], а о достаточно ограниченных по своим возможностям в сравнении с человеком программных продуктах, которые очень хорошо решают одну определенную задачу и не способны одновременно управлять автомобилем и анализировать медицинские данные что бы поставить диагноз пациенту.

Раз мы сказали о новых особенностях взаимодействия человека и машины, то необходимо отметить те области, где они могут возникать. В первую очередь это автономные транспортные средства, автономные системы принятия решений (от банковских до технологических), средства интеллектуального анализа поведения пользователя в сети Интернет и т.д. Кроме того, возникают многочисленные вопросы, связанные с взаимодействием человек-машина, включая выстраивание диалога с программой-советчиком степень доверия к ИИ/АС и т.д. [3][4]. Таких вопросов возникает все больше, и все они складываются в то направление исследований, которое сегодня получило название этика ИИ/АС [5][6]. Исследования в этой области привели к появлению ряда инициатив по исследованию в данной области [3][2]. Среди таких инициатив следует выделить глобальную инициативу Института инженеров в области электротехники и

электроники (Institute of Electrical and Electronics Engineers — IEEE), в рамках которой начата работа по созданию нормативно-технических документов, которые заложили бы основу этичного поведения в разрабатываемые системы с ИИ/АС. Первым шагом в этой работе является создание рекомендаций, направляющих разработчиков ИИ/АС на этически обоснованные решения при разработке и применении систем ИИ/АС [6]. Эти рекомендации распространяются как на ситуацию, в которой система с ИИ/АС или автономная система принимает решение в какой-либо этической ситуации, так и на ситуацию, в которой применение ИИ/АС или автономных систем приводит к каким-либо социальным последствиям в виде сокращения определенного персонала. Таким образом, можно выделить два направления исследований:

1. непосредственно этику ИИ/АС, как этические проблемы, вызванные работой машин, самостоятельно принимающих решения;
2. этические проблемы, связанные с массовым внедрением ИИ/АС, которые могут привести к сокращениям персонала, занятого рутинными работами.

Очевидно, что в конечном итоге все исследования в этой области ведут к созданию нормативных документов, чем и занимается IEEE (<https://ethicsinaction.ieee.org/>). При этом, что этическое регулирование ИИ/АС и автономных систем не ведет к ограничениям или запрету в развитии таких технологий. Например, система «умный дом» может вызвать помощь, если определит, что человек в доме, например, потерял сознание. Эта система принимает решение, вызывая помощь, и тем самым, она может спасти жизнь человеку. Ограничивать развитие таких систем, значит лишать шанса многих людей в будущем.

В завершение следует отметить, что решение задач в области этики ИИ, очевидно, требует междисциплинарного подхода. Этот подход должен выражаться не только в исследованиях с участием ученых из разных областей науки, но и появлением в учебных курсах специалистов в области ИИ вопросов этики, и наоборот, предметов, рассказывающих о современном уровне технологий для тех студентов-гуманитариев, кто интересуется технической этикой в целом.

#### **Список использованных источников**

- [1] N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*. 2014.
- [2] N. Bostrom and E. Yudkowsky, “The Ethics of Artificial Intelligence,” *Cambridge Handb. Artif. Intell.*, pp. 1–20, 2011.
- [3] B. A. Grinbaum, R. Chatila, L. Devillers, J. Ganascia, C. Tessier, and M. Dauchet, “Ethics in Robotics Research,” *IEEE Robot. Autom. Mag.*, vol. 24, no. 3, pp. 139–145, 2017.
- [4] N. Wiener, “Some Moral and Technical Consequences of Automation,” *Science (80-. )*, vol. 131, no. 3410, pp. 1355–1358, 1960.
- [5] J. Havens, *Heartificial Intelligence: Embracing Our Humanity to Maximize Machines*. 2016.
- [6] *Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems*. 2018.