

Практика искусственного интеллекта и проблема свободы личности

В.П. Казарян

Московский государственный университет им. М.В. Ломоносова, Москва, Россия

Поступила в редакцию: 29.05.23

В окончательном варианте: 30.09.23

■ Для цитирования: Казарян В.П. Практика искусственного интеллекта и проблема свободы личности // Вестник Самарского Государственного Технического Университета. Серия «Философия». 2023. Т. 5. № 4. С. 27–32. DOI: <https://doi.org/10.17673/vsgtu-phil.2023.4.5>

Аннотация. Работа посвящена выявлению новых ограничений для свободы личности, появляющихся в условиях интенсивного развития информационных технологий, а именно области нетрадиционного искусственного интеллекта. Алгоритмы машинного обучения отказывают личности в гарантии конфиденциальности ее душевно-психологического состояния.

Ключевые слова: искусственный интеллект (ИИ); алгоритм; машинное обучение; личность; свобода; ИИ-бот; дезинформация; манипулирование.

The practice of artificial intelligence and the problem of individual liberty

V.P. Kazaryan

Lomonosov Moscow State University, Moscow, Russia

Original article submitted: 29.05.23

Revision submitted: 30.09.23

■ For citation: Kazaryan V.P. The practice of artificial intelligence and the problem of individual liberty. *Vestnik of Samara State Technical University. Series Philosophy*. 2023;5(4):27–32. DOI: <https://doi.org/10.17673/vsgtu-phil.2023.4.5>

Abstract. The work is devoted to identifying new restrictions on the freedom of the individual, appearing in the conditions of intensive development of information technologies, namely the field of non-traditional artificial intelligence. Machine learning algorithms deny the individual the guarantee of the confidentiality of her mental and psychological state.

Keywords: artificial intelligence (AI); algorithm; machine learning; personality; freedom; AI bot; disinformation; manipulation.

Новые технологии создаются человеком для достижения жизненно важных целей при обеспечении безопасности человека. Одной из жизненно важных целей является свобода и справедливость. За них борются лучшие люди в истории человечества. При этом каждый человек мечтает о справедливости и свободе. Безопасность личности подразумевает неснижающуюся вероятность достижения жизненно важных целей. Новые технологии, изменяющие нашу жизнь во многих отношениях, помогают ли достигнуть свободы?

Человек, использующий компьютер и живущий в интернете, кажется свободным. Или, по крайней мере, более свободным, чем без интернета — расширенного разума. Ранее человек, как показал Э. Фромм, освободился «от»: при рождении «от» уз природы (пуповины матери), затем «от» уз средневековой социальной структуры (привязанности к месту, профессии, общине). С развитием буржуазных отношений он обрел индивидуальность, приобрел «свободу от» [1]. Но в процессе дальнейшего развития общества он оказался в плену больших социальных структур, лишившись «свободы от». Как показали структуралисты, появилось общество структур: человека нет, есть одни структуры. Человек бессилен перед ними. Обесценивается свобода, добытая предшественниками. Вместо свободы появилась новая зависимость «от». Человек — вещь среди вещей. Свобода превращается в иллюзию, в мечту. Экзистенциалисты призывали восстать против бессилия человека, пробудить в нем личность. В сложившемся обществе потребления человек стал жить лучше во многих отношениях. Ортега-и-Гассет показал, что цену за благополучие, которое дал научно-технический прогресс, люди заплатили большую: цена за это — массовый человек [2].

Призрак свободы как основа этики бродит в сердцах людей. Новые технологии, развивая человеческие потребности, снимут ли последнее ограничение «от»: Я — только в массе, Я — массовый человек? Откроется ли дорога к свободе «для», свободе для творчества?

Прежде всего, выделим два вида ИИ: традиционный и нетрадиционный ИИ (ИИ для Big Date). С момента зарождения искусственного интеллекта специалисты традиционного направления пытались создавать системы, способные решать задачи, приближенные к человеческому интеллекту, такие как системы автоматического доказательства теорем, игровые системы, системы перевода с одного языка на другой, системы распознавания зрительных образов, сочинения музыки и подобные им. Для этого не требовалось много данных, но требовалась экспертная оценка.

Развитие области микроэлектроники и ее доступность создали возможность генерации, хранения и обработки больших объемов данных [3]. Это создало предпосылки для решения множества практически важных задач методами машинного обучения. Для алгоритмов машинного обучения важна доступность данных. Это направление составляет подавляющее большинство систем ИИ в настоящее время. Существует большое множество видов моделей машинного обучения.

Традиционный ИИ прекрасно работает и в наше время. Он не порождал проблем со свободой человека. Он огорчал и восхищал человека своими возможностями в сравнении с человеческим разумом: кто умнее. Если традиционный ИИ поражал человека, то ИИ для Big Date должен помогать человеку в его деятельности. Парадигма машинного обучения позволяет получать хорошие результаты. Вместе с тем он порождает определенные сомнения. При практической реализации он сталкивается с рядом трудностей, касающихся взаимоотношения алгоритма с человеком, с условиями деятельности человека — т. е. с проблемой свободы человека.

Принято считать, что человек совершает действие на основе принятия решения. В практике ИИ действие совершается в соответствии с решением

алгоритма. Человек может знать о том, что он реализует решение алгоритма, и может об этом не знать и полагать, что сам самостоятельно принял решение на действие. Примером может служить участие в голосовании под влиянием ИИ-подсказки, подобранной именно для этого человека.

Для проблемы свободы человека важное значение могут иметь новые появившиеся способности ИИ. Алгоритмы искусственного интеллекта стали способны понимать человеческие предпочтения. Это показало исследование, проведенное несколько лет назад по заказу Facebook¹. Оно показало, что фейсбукowski алгоритм разбирается в характерах и настроениях людей лучше, чем их друзья, родители и супруги. В эксперименте участвовало 86 220 волонтеров, которые имели аккаунт в Facebook и ответили на сто вопросов персональной анкеты. Алгоритм предсказывал ответы волонтеров, основываясь на анализе их «лайков» — то есть страниц, изображений и клипов, которые они отметили как понравившиеся. Точность предсказаний зависела от количества лайков [4]. Но несколько сотен лайков было достаточно, чтобы алгоритм ответил на вопросы о человеке более точно, чем его родственники и друзья. В результате этого многие ученые сделали прогноз, согласно которому в будущем люди могут отказаться от собственных субъективных суждений и доверить компьютерам принятие жизненно важных решений типа выбора профессии, любимых занятий и даже романтического партнера.

Возможно, где-то это улучшит качество жизни человека, однако человек, вполне вероятно, будет все больше и больше лишаться свободы выбора. Поэтому проблему свободы, являющуюся одной из важнейших в философии, необходимо будет рассматривать исходя из новых реалий. Например, можно ли считать, что человек, который поступает в соответствии с рекомендациями алгоритма искусственного интеллекта, обладает свободой воли? Или, например, отличается ли действие, совершенное осознанно в соответствии с рекомендациями искусственного интеллекта, и аналогичное действие, но совершенное неосознанно (то есть когда человек не осознает, что он действует под влиянием алгоритма машинного обучения)?

Другой пример возможностей нетрадиционного алгоритма. Алгоритмы Facebook могут быть осведомлены не только о политических предпочтениях миллионов человек, но и о том, кто из них относится к критической группе колеблющихся избирателей и как этих колеблющихся избирателей склонить в ту или другую сторону. Facebook сможет сказать, например, что в определенном штате две партии идут вровень, что какая-то часть избирателей еще не определилась, и даже сформулировать, что именно должен говорить кандидат, чтобы склонить чашу весов в свою сторону. Таким образом, искусственный интеллект можно применять для влияния на электоральные процедуры и при некоторых выраженных политических взглядах крупнейших корпораций сделать невозможным победу оппонента. В связи с этим необходимо переосмысление традиционного понятия либеральной демократии, в которой предполагалось, что человек способен совершать выбор независимо и свободно.

¹ Является продуктом компании Meta, которая в свою очередь является экстремистской организацией, деятельность которой запрещена на территории РФ.

Крупнейшие корпорации, используя алгоритмы искусственного интеллекта, способны влиять на огромное количество людей по своему усмотрению.

Однако влияние на человеческое поведение способны осуществлять не только крупные корпорации или другие большие структуры, но и отдельные люди. В качестве такого примера можно рассмотреть создание с помощью нейронных сетей дипфейков [5] — фотографий или видеороликов с поддельным содержанием. Создать дипфейк сейчас под силу любому человеку, который умеет обращаться с нейронными сетями. Отделение правдивой и ложной информации с каждым годом будет становиться все труднее. Также технологии искусственного интеллекта могут быть использованы и для генерации ложных новостей (с помощью генеративных текстовых моделей).

Итак, хотя алгоритмы искусственного интеллекта обычно создаются для благих целей, эти же алгоритмы могут быть использованы для манипуляции человеком. Так как эти алгоритмы могут достаточно точно понимать психологию человека, то их можно использовать для внедрения многих идей. Тем самым, необходимо переосмысление понятия свободы воли с учетом новых складывающихся реалий, потому что очевидно, что при недобросовестном использовании новые технологии способны сокращать количество свободы.

Нельзя не обратить внимания на тот факт, что современные модели машинного обучения полностью зависят от данных, на которых они обучаются. Поэтому очень важным остается вопрос формирования обширных и разнообразных обучающих выборок. В этом контексте многие, если не подавляющее большинство специалистов, указывают на предвзятость ИИ — на предвзятость машинного обучения. Под предвзятостью машинного обучения понимают тенденции алгоритмов отражать человеческие предубеждения. И они в результате лежат в основе судьбы индивидуума, которой касается алгоритм.

Появление систем ИИ автоматической генерации текста влечет за собой новые этические проблемы. Такие системы могут генерировать текст в больших объемах, который довольно сложно отличить от текста, созданного человеком [6]. В такой ситуации мы наблюдаем размывание ролей между людьми и машинами в производстве текстового контента. Это событие имеет глубокие последствия для общества: что произойдет, если мы в конечном итоге заставим замолчать человеческий голос? Эти алгоритмы ограничивают возможность творчества со стороны человека.

ИИ лишает ли человека возможности создавать новое? Например, создать новое в науке и создать новое в искусстве? Что касается науки, то ИИ помогает ученому увидеть проблему и ее решать. ИИ не вытесняет пока ученого из науки. Интеллектуальные модели позволяют людям делегировать рутинную работу и обратить внимание на творческие задачи.

В художественном творчестве больше неясного. Если машинное обучение использовало для обучения существующие произведения искусств, то это выступает аргументом в пользу утверждения, что, обучаясь на старом, нового создать нельзя. Можно получить только копии. Базы данных — это прошлое, даже не настоящее. И получить будущее из них не представляется возможным. Конечно, гении стоят на головах предшествующих гениев, и они создают новое. Но механизм творчества является до сих пор загадкой, с одной стороны, и сферой свободы личности — с другой.

Для свободы человека могучим врагом выступает манипулирование и дезинформация, столь распространенные в современном мире [7]. Свободу ограничивает наличие недостоверной информации. Полностью разрушается идея этики античности, ориентирующейся на знание. Распространение дезинформации с целью манипулирования общественным мнением встречается все чаще. Опасность неконтролируемого распространения дезинформации была заметна во время акций против вакцинации в период пандемии COVID, президентских выборов в США и Франции в 2016 и 2017 годах и кампании Brexit в Великобритании в 2016 году, когда в социальных сетях широко распространялись фальшивые новости о противоположных сторонах мнений. Проведенные исследования [8] показали, что дезинформация в течении этих кампаний распространялась в основном ИИ-ботами на основе ИИ, которые создавали недостоверную информацию.

Ключевые признаки фальшивых новостей, созданных ботами, могут позволить социальным сетям, таким как Facebook и Twitter, отмечать поддельное содержание публикаций на ранней стадии, чтобы предотвратить его распространение. Однако эти особенности не обнаруживаются, когда содержание фальшивых новостей создается с помощью ИИ последнего поколения, который все чаще заменяет традиционные боты. Ученые отмечают сложность обнаружения фальшивого контента, направленного на пропаганду, из-за стиля написания, который похож на человеческий. Особенность таких моделей, которая делает довольно сложной работу по обнаружению, — это способность учиться на основе реального взаимодействия. Это дает им возможность распространять одно и то же сообщение в различных формах. Такие модели могут автоматически подстраивать стиль материала под разные типы аудитории (например, в зависимости от социальной сети), что делает их содержание еще более сложным для обнаружения.

В итоге с некоторой долей уверенности можно предположить, что личность будет активно использовать возможности алгоритмов при решении своих проблем. Возможно, это позволит улучшить ее жизненный мир. При этом она потеряет долю своей автономии в столь важной для личности области индивидуальной самости и отдаст алгоритмам сугубо человеческое право на принятие решений. Проблема свободы личности приобретает новые очертания.

Список литературы

1. Фромм Э. Бегство от свободы / под ред. П.С. Гуревича. Москва: Прогресс, 1989. 272 с.
2. Ортега-и-Гассет Х. Восстание масс. Москва: АСТ: Ермак, 2005. 509 с.
3. Dan Koboldt [Internet]. The current state of artificial intelligence. [дата обращения 25.12.2022]. Доступ по ссылке: <http://dankoboldt.com/current-artificial-intelligence/>
4. Харари Ю.Н. Homo deus. Краткая история будущего. Москва: Синдбад, 2022. 496 с.
5. Mirsky Y., Lee W. The creation and detection of deepfakes: a survey // ACM Computing Surveys (CSUR). 2021. Vol. 54. No. 1. P. 1–41. DOI: 10.1145/3425780
6. Kreps S., McCain M., Brundage M. All the news that's fit to fabricate: AI-generated text as a tool of media misinformation // Journal of Experimental Political Science. 2022. Vol. 9. No. 1. P. 104–117. DOI: 10.1017/XPS.2020.37
7. Илизова Л.П., Илиев С.Д. Деманипуляция в НЛП: сложносистемный подход // Бионика – 60 лет. Итоги и перспективы. Сборник статей Первой Международной

научно-практической конференции 17–19 декабря 2021 года, г. Москва / под ред. А.П. Карпенко. Москва: Ассоциация технических университетов, 2022. С. 30–40.

8. Seele P, Schultz M.D. From greenwashing to machinewashing: a model and future directions derived from reasoning by analogy // J. Bus. Ethics. 2022. Vol. 178. P. 1063–1089. DOI: 10.1007/s10551-022-05054-9

References

1. Fromm E. Begstvo ot svobody. Ed.by P.S. Gurevicha. Moscow: Progress; 1989. (In Russ.)
2. Ortega-i-Gasset Kh. Vosstanie mass. Moscow: AST: Ermak, 2005. (In Russ.)
3. Dan Koboldt [Internet]. The current state of artificial intelligence. [cited 2022 Dec 25]. Available from: <http://dankoboldt.com/current-artificial-intelligence/>
4. Kharari YuN. Homo deus. Kratkaya istoriya budushchego. Moscow: Sindbad, 2022. (In Russ.)
5. Mirsky Y, Lee W. The creation and detection of deepfakes: a survey. *ACM Computing Surveys (CSUR)*. 2021;54(1):1–41. DOI: 10.1145/3425780
6. Kreps S, McCain M, Brundage M. All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*. 2022;9(1):104–117. DOI: 10.1017/XPS.2020.37
7. Ilieva LP, Iliev SD. Demanipulyatsiya v NLP: slozhnosistemnyi podkhod. In: Bionika — 60 let. Itogi i perspektivy. Sbornik statei Pervoi Mezhdunarodnoi nauchno-prakticheskoi konferentsii 17-19 dekabrya 2021 goda, g. Moskva. Ed. by A.P. Karpenko. Moscow: Assotsiatsiya tekhnicheskikh universitetov; 2022. P. 30–40. (In Russ.)
8. Seele P, Schultz MD. From greenwashing to machinewashing: a model and future directions derived from reasoning by analogy. *J Bus Ethics*. 2022;178:1063–1089. DOI: 10.1007/s10551-022-05054-9

Информация об авторе

Валентина Павловна Казарян — доктор философских наук, профессор, профессор философского факультета МГУ имени М.В. Ломоносова, Москва, Россия; eLibrary SPIN: 8201-4661.
E-mail: kazaryanvp@mail.ru

Information about the author

Valentina P. Kazaryan — Doctor of Philosophy, Professor, Professor of Philosophy, Lomonosov Moscow State University, Moscow, Russia; eLibrary SPIN: 8201-4661. **E-mail:** kazaryanvp@mail.ru