

Искусственный интеллект: сферы применения, потенциальные угрозы использования

Н.В. Алексеев

студент 3 курса бакалавриата НИЯУ МИФИ, Москва

E-mail: nik_val_alekseev@mail.ru

В.А. Рычков

Старший преподаватель кафедры финансового мониторинга

НИЯУ МИФИ, Москва

E-mail: varychkov@mephi.ru

Аннотация: Возрастающая в последнее время популярность искусственного интеллекта и его развитие могут привести не только к позитивным изменениям в обществе, но и к возникновению новых угроз. Цель данной статьи - дать их описание и предложить возможные методы их минимизации.

Ключевые слова: искусственный интеллект, угрозы, личность, безопасность.

Artificial intelligence: meaning of concept, area and threats of usage

N.V. Alekseev

3rd year bachelor's student at NRNU MPhI, Moscow

E-mail: nik_val_alekseev@mail.ru

V.A. Rychkov

Senior Lecturer of the Department of Financial Monitoring

NRNU MPhI, Moscow

E-mail: VARychkov@mephi.ru

Abstract: Increasing popularity of artificial intelligence and its development in the last time can lead not only to positive changes, but to new threats too. The purpose of this article is to describe them and to offer possible methods of minimization.

Keywords: artificial intelligence, threats, personality, safety.

Введение

В течение последних десяти лет использование искусственного интеллекта от практически нулевого резко выросло: объём рынка достиг

298 млрд. долларов, и распространилось на многие сферы человеческой жизни.

Однако, как и любая новая технология, ИИ несёт в себе не только возможности, но и угрозы безопасности личности и общества, которые желательно нивелировать до тех пор, пока они не получили масштабного распространения. [8, 11].

Цель данной статьи - дать описание понятия “искусственный интеллект” (далее - ИИ), выявить основные сферы, в которых применение ИИ может привести к негативным последствиям, определить потенциальные угрозы, связанные с его использованием.

Искусственный интеллект. Описание понятия. Критические сферы применения. Искусственный интеллект (ИИ) - междисциплинарная технология, направленная на использование больших наборов данных (Big Data), подходящих вычислительных мощностей, а также специальных аналитических процедур и процедур принятия решений для того, чтобы позволить компьютерам выполнять задачи, приближенные к человеческим способностям, а в некоторых отношениях даже превосходящие их. [8, 11]

Если использовать упрощение, ИИ - это сверхмощный вероятностный калькулятор, выдающий решение относительно текущих входных данных на базе предзагруженных данных и предыдущего опыта.

В настоящий момент, продвижение данной технологии в направлении выполнения человеческих задач значительно: ИИ может генерировать текст по запросу: писать код, эссе, аналитические сводки и т.д.; может распознавать и генерировать изображения, звуковые дорожки; может управлять беспилотными системами. В связи с этим, искусственный интеллект стремятся внедрить во многих сферах жизнедеятельности общества с целью повышения эффективности деятельности, снижения затрат. Таким образом, влияние ИИ распространяется взрывными темпами. В связи с этим, началось исследование ИИ учёными с различных точек зрения, но доминирующее количество публикаций было выпущено в трёх сферах:

- Компьютерные науки.
- Когнитивные науки.
- Философия.

Это показывает, что наибольшее внимание притягивает возможное влияние искусственного интеллекта на человека, его способности и мировоззрение. Данная работа также сосредоточится больше на влиянии искусственного интеллекта на человека в сфере психологической и когнитивной, чем на возможных угрозах кибербезопасности и физической безопасности субъектов.

Уязвимости личности и общества перед искусственным интеллектом. Угрозы, исходящие от использования ИИ. Рассматриваются угрозы, разделённые на два отдельных класса: угрозы личности и угрозы обществу. Такое разбиение обусловлено принципиальной разницей в последствиях от реализации выявленных угроз: если для личности такие последствия чаще всего психологические, то для общества они относятся больше к вопросам устойчивости и безопасности.

Для того, чтобы рассмотреть угрозы, возникшие из-за использования ИИ, и предложить меры по минимизации этих угроз, нужно выявить также и уязвимости, позволившие угрозам быть угрозами.

Были выявлены следующие уязвимости личности и общества, которые дают возможность реализоваться угрозам, исходящим от использования ИИ:

Уязвимости личности (следует заметить, что эти уязвимости присущи не всем и не всегда, но определённая общественная тенденция существует):

1. Повышенная тревожность. [5,10]
2. Инфантильность, нежелание принимать на себя ответственность за принимаемые решения. [5,10]
3. Ослабленное критическое мышление. [3, 10]
4. Нечёткость моральных рамок и норм поведения, их относительность.

Уязвимости общества:

5. Недостаточно серьёзное восприятие возможных опасностей, исходящих от использования ИИ, как психологических и угроз безопасности государства и общества, так и угроз физической и кибербезопасности: кибератак, атак с помощью беспилотных средств. [2, 10]

6. Массовая виртуализация жизни, неконтролируемое потребление информации из СМИ, социальных сетей, Интернета; зависимость от такого потребления. [1, 10]

7. Массовое использование ИИ, в основном иностранных. Отсутствие возможности контролировать процесс создания иностранных ИИ.

Из этих уязвимостей вытекают следующие угрозы личности от ИИ и методы их минимизации (следует отметить, что угрозы здесь изложены в намеренно преувеличенном виде, для того, чтобы высветить предельный случай каждой из них; но в той или иной степени они все могут реализоваться для одного человека):

- 1) Мошенничество с помощью DeepFake. ИИ даёт мошенникам возможность имитировать голос начальника, близких людей, что повышает

подверженность человека мошенническим схемам. Возникновению этой угрозы способствуют уязвимости (1) - повышенная тревожность и (3) - ослабленное критическое мышление: ослабленное критическое мышление не даёт возможности оценить ситуацию верно, повышенная тревожность вгоняет в страх, подстёгивающий на совершение необдуманных поступков. Минимизировать эту проблему можно, во-первых, относясь критически к поступающей через системы связи информации, а во-вторых - развивая свою личность, избавляясь от тревожности (стоит отметить, что для этого личности нужно сформировать у себя нравственные идеалы, отличные от идеалов “общества потребления”[4, 10][5, 10], позволяющие действовать несмотря на внешние неблагоприятные обстоятельства; воспитать в себе такие качества, как мужество, самоотверженность - хотя бы в бытовом понимании; если говорить в терминах психологии - обрести субъектность)

2) Персонализированный фишинг. С помощью ИИ появляется возможность создавать фишинговые письма персонализированно, особенно собрав информацию о человеке в соцсетях и прочих открытых источниках; это повышает шансы на то, что человек поддастся на провокацию. Бороться с этим можно критически относясь ко всей поступающей информации и производя проверку фактов. Ключевая уязвимость здесь, при наличии которой угроза получает шансы на реализацию - (3) - сниженное критическое мышление.

3) Управление сознанием с помощью персонализированных рекомендаций, подложной информации. Если кто-то, имеющий доступ к формированию ИИ, злонамеренно скорректирует обучение ИИ - даст неправильную выборку, ложные факты, то, в случае использования ИИ вместо традиционных методов сбора и поиска информации, резко повышается возможность управлять сознанием человека, т.к. в таком случае для каждого человека можно предоставить ложную информацию, не противоречащую его мировоззрения напрямую, но постепенно корректирующую мировоззрение в нужную создателю ИИ сторону. Данная угроза может реализоваться и в другой форме: если тот же приём коррекции ИИ использовать в сфере персонализированного подбора рекомендаций: направить человека на те ресурсы, которые выгодны архитектору ИИ. Для человека последствия такой манипуляции могут выражаться в различных девиациях, социально неприемлемом поведении, психических расстройствах. Уязвимости, которые привели к возможности реализации этой угрозы - (2) инфантильность, (3) ослабленное критическое мышление, (4) нечёткость моральных границ. Возможные методы противодействия - сокращение потребления контента, не необходимого в профессиональной деятельности; самостоятельный анализ и осмысление информации; творческий подход к жизни.

4) Уничтожение творческого потенциала. Инфантильность может привести к перекалыванию ответственности (причём в любой сфере жизни, т.к. ИИ обучен на данных, предоставленных людьми) на ИИ. Стремление не решать творческие задачи, связанные с работой, учёбой, общением и т.д., а просто спросить у ИИ, неизбежно ведёт к деградации когнитивных способностей и творческого потенциала личности, уничтожает субъектность [3, 10]. Минимизировать эту угрозу личность может, осознав, что смысл жизни заключается не в постоянном потреблении, поняв необходимость развития.

Угрозы обществу, исходящие от ИИ:

1) Персонализированный фишинг, мошенничество с помощью DeepFake; создание полностью виртуальной жизни, атомизация общества. Последствия этих угроз на уровне общества могут проявиться как полное недоверие любой информации, публикуемой в Интернете, ещё большей виртуализации сознания населения, атомизации общества, одиночеству, учитывая зависимость от потребления информации: если пользователь не прекратит потребление информации, при этом не считая её истинной, реальной, он будет жить в некотором полностью виртуальном пространстве, в реальность которого он не будет верить, но в котором будет существовать больше, чем в настоящем мире. Стоит отметить, что в той или иной мере это присутствует и сейчас у тех, кто потребляет контент в виде новостных постов, коротких видео и т.п.: их сознание занято той информацией, которая не относится напрямую к их повседневной жизни; но ИИ даёт гораздо более широкие возможности для имитации реальности. [1, 10]

2) Хаотические ложные знания в обществе. Если задать ИИ запрос на выдачу информации или мнения, не сомневаясь в его истинности, ИИ может сгенерировать тот ответ, который удовлетворит пользователя. Пример: ИИ был задан запрос «назови десять советских фильмов, в которых пропагандируются семейные ценности, с их кратким описанием». ИИ выдал список, краткое описание сюжета было правдивым, но выводы были сделаны неверные: из десяти фильмов действительно про семейные ценности был только один. Такие ответы могут привести к усилению атомизации общества: одно дело, если у двух людей просто разные мнения, и другое - если они подтверждены в их сознании данными, сгенерированными с помощью ИИ. Минимизировать эту угрозу можно, сделав в системе образования упор на то, чтобы дать школьникам и студентам не только технические знания, но и культурный и мировоззренческий базис.

3) Систематические ложные знания в обществе. Учитывая, что большинство пользователей ИИ в России сейчас используют зарубежные языковые модели, повышается риск того, что иностранные интересные

отредактируют в моделях информацию, относящуюся к исторической и политической ситуации. Минимизировать эту угрозу можно, создавая свой ИИ, в случае критической ситуации в данной области - сегментировав сеть Интернет, отделив её от общемировой.

4) Разобщённость общества. Инфантильность может привести к тому, что человек может посчитать ИИ более предпочтительным собеседником, чем живой человек, с которым нужно искать компромисс, договариваться, тогда как ИИ может в одной ситуации выступить в роли гуру, на которого можно переложить ответственность за принятие решения в жизненной ситуации, в другой - как личность, полностью зависимую от пользователя, восхищающуюся им. Чтобы минимизировать эту угрозу, нужно распространять в обществе взгляды о том, что общение с людьми происходит не для личного удовольствия, а для того, чтобы принести другому пользу или поддержку; если воспринимать ситуацию более широко, нужно воспитывать в людях ту систему ценностей, которая была бы направлена не на личное удобство человека, а на то, чтобы стать лучше и отдать что-то хорошее в окружающий мир. Следует отметить, что такое мировоззрение воспитать без веры в Бога достаточно тяжело.

5) Уничтожение творческого потенциала общества. Широкое использование ИИ в интеллектуальной работе может привести к тому, что в стране будет всё меньше людей, способных создавать что-то принципиально новое. Недавние исследования [6, 10] показали, что постоянное использование ИИ снижает нейронную связность мозга, ослабляет память, снижает самооценку владения знаниями и когнитивные способности. Также исследования [7, 10] показали, что более 40% студентов используют ИИ в учёбе регулярно. Особенно опасно использование ИИ в сфере образования, т.к. именно там закладывается способность человека к познанию. Минимизировать эту угрозу можно, распространяя в обществе соответствующие исследования о вреде использования ИИ, формируя в людях субъектность.

6) Кибератаки, с которыми крайне тяжело справиться обычными средствами. ИИ может использовать весь опыт человеческих кибератак, при этом действуя на скорости, принципиально недоступной команде хакеров. Как показывает недавний опыт кибератаки на США [10, 11], искусственный интеллект может быть успешно использован на всех этапах кибератаки, от подготовки фишинговых писем до собственно активной атаки на информационную инфраструктуру предприятий. Для минимизации этой угрозы необходимо собрать исследователей из сферы кибербезопасности, организовать анализ возможностей ИИ в сфере кибератак, тренировочные соревнования “синих” и “красных” команд по взлому с использованием ИИ, разработку систем распознавания и противодействия атак с использованием ИИ. [2, 10]

7) Атаки с помощью беспилотных систем. Данная угроза поддаётся только теоретическому рассмотрению ввиду отсутствия известных примеров таких атак, однако учитывая способность ИИ к координации действий гораздо выше человеческой, это может привести к резкому возрастанию смертоносности войн. Минимизировать эту угрозу можно, развивая электромагнитное оружие, выводящее из строя электротехнику.

Заключение

Если обобщить содержимое статьи, то искусственный интеллект - это перспективная разработка, которая переросла в масштабную индустрию. Хотя технология ИИ включает в себе значительные перспективы в области работы с естественным языком, автоматизированного управления техническими системами, анализа больших данных, она усиливает уже существующие угрозы для личности и общества, а также несёт в себе дополнительные.

Неконтролируемое использование ИИ может навредить человеку так же, как вредные привычки: снизить когнитивные способности, усложнить социальную адаптацию, повысить нестабильность психики. В общественном пространстве использование ИИ может повысить разобщённость, привести к массовым манипуляциям сознанием. В деле вооружённого и хакерского противостояния ИИ может дать толчок к развитию новых деструктивных технологий.

Учитывая, что искусственный интеллект стремится внедрить во многих областях: в бизнесе, науке, соцсетях, образовании, нужно в краткие сроки выработать механизмы общественного и личного противодействия угрозам, от него исходящим, а в перспективе - обратить внимание на те слабости, которые привели к возникновению этих угроз и скорректировать вектор своего движения, учитывая их.

Подводя итог, нужно отметить, что сама по себе концепция искусственного интеллекта не представляется как вредоносная; в большинстве приведённых вариантов она может принести вред лишь в сочетании с уязвимостями личности и общества, которые бытуют в современности. Из этого можно сделать вывод, что первично не ограничение самой технологии искусственного интеллекта, а воспитание в человеке здоровой, субъектной личности, которая обладает внутренним стержнем и способна сделать правильный выбор в сложившейся обстановке.

Список использованных источников:

1. Ванновская, О. В. Влияние виртуальной реальности на сознание человека с позиций христианской антропологии и современной психологии / О. В. Ванновская // Вестник Православного Свято-Тихоновского

гуманитарного университета. Серия 4: Педагогика. Психология. – 2023. – № 70. – С. 90-102. – DOI 10.15382/sturIV202370.90-102. – EDN NDLTCI.

2. Artificial intelligence index report 2025, Stanford University

3. Мастерова, А. А. Риски влияния искусственного интеллекта на когнитивные способности людей / А. А. Мастерова, А. В. Тимофеев // Наука XXI века: актуальные направления развития. – 2024. – № 2-2. – С. 230-234. – EDN QMPYXB.

4. Апенешева, Л. Г. Специфика морали в светских культурах в условиях современного общества / Л. Г. Апенешева, Н. В. Кулипанова, Е. И. Тарасова // Социальные процессы в современной западной Сибири : сборник научных трудов XVII Всероссийской научно-практической конференции, Горно-Алтайск, 07–09 июля 2017 года / Горно-Алтайский государственный университет. Том Выпуск 17. – Горно-Алтайск: Горно-Алтайский государственный университет, 2017. – С. 18-21. – EDN ZUPIXN.

5. Smirnova, O. V. Fear of the future: causes and help in overcoming this phobia / O. V. Smirnova // Теория права и межгосударственных отношений. – 2021. – Vol. 1, No. 8(20). – P. 602-605. – EDN PLGULO.

6. Your brain on ChatGPT: accumulation of cognitive debt when using AI assistant for essay writing task. https://www.researchgate.net/publication/392560878_Your_Brain_on_ChatGPT_Accumulation_of_Cognitive_Debt_when_Using_an_AI_Assistant_for_Essay_Writing_Task

7. <https://www.vedomosti.ru/society/articles/2025/03/31/1101211-iskusstvennii-intellekt-dlya-napisaniya-kursovih-rabot>

8. Доктрина ИБ РФ, Раздел 1, п. 2г

9. Беликова, Е. К. Социокультурная сущность искусственного интеллекта / Е. К. Беликова // Bulletin of the International Centre of Art and Education. – 2023. – № 4. – С. 290-297. – EDN ICMVGQ.

10. <https://www.forbes.ru/tekhnologii/549885-kitajskie-hakery-ispolzovali-ii-ot-anthropic-dla-avtomatizacii-kiberatak>