

КОДЕКС ПОВЕДЕНИЯ РОБОТА



Искусственный интеллект с каждым годом все глубже внедряется в нашу жизнь. Мы встречаемся с ним практически повсеместно: дома, на транспорте, в медицине, при совершении покупок и при общении с разного рода онлайн-сервисами, в том числе государственными. Мы видим в ИИ помощника и защитника, однако не можем не задумываться о том, какие новые проблемы этического характера принесет активное вторжение машин в мир людей.

ИИ

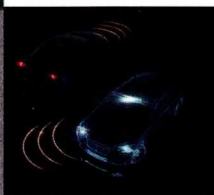
ежду тем эти проблемы уже находятся в центре внимания мирового сообщества. Документы, регламентирующие этические стороны взаимодействия человека и ИИ готовятся в ООН и Евросоюзе. Разрабатывается и российский Национальный кодекс этики в сфере ИИ. Скоро в Москве планируется проведение представительного форума «Этика искусственного интеллекта: начало доверия», организованного органами государственной власти в сотрудничестве с крупными компаниями и некоммерческими исследовательскими организациями. На мероприятии будут обсуждаться ключевые проблемы, связанные с экономикой, образованием, транспортом, безопасностью и IT-сферой в свете новых этических вызовов.

«ПМ» поговорила с рядом экспертов, которые разъяснили нам, в чем заключаются эти ключевые проблемы и что находится в центре внимания государства, разработчиков и эксплуатантов, а также всей заинтересованной общественности.

ЗА РУЛЕМ ИЛИ БЕЗ РУЛЯ?

Возможно, самым ярким и очевидным примером внедрения ИИ в нашу жизнь может служить транспорт – все то, что ездит, летает и плавает без участия человека (или с его участием в качестве контролера). Новости об испытании беспилотных автомобилей или новых дронов мы читаем практически ежедневно. «Даже для того чтобы просто тестировать такие технологии, требуется, как правило, менять регулирование, – говорит Андрей Незнамов, управляющий директор Центра регулирования ИИ ПАО «Сбербанк», кандидат юридических наук, один из разработчиков Национального кодекса этики в сфере ИИ. – Например, в случае с автомобилями необходимо изменить правила дорожного движения, и мировое сообщество сейчас этим занято. Вводятся категории транспортных средств, в которых есть водитель, но он уже не всегда следит за дорогой, и категория полностью автоматизированных. Обсуждается, на каких условиях человек вправе отвлекаться от вождения, а когда может пересесть в пассажирское кресло. И как все это встроить в обычное дорожное движение, в котором участвуют привычные управляемые человеком машины».

Для автомобилей с ИИ необходимо изменить правила дорожного движения, и мировое сообщество сейчас этим занято.

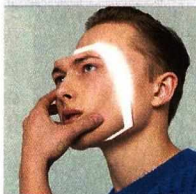


Другой сферой, где ИИ применяется достаточно широко, можно назвать здравоохранение. «Во время пандемии, – продолжает Андрей, – многие увидели огромный потенциал искусственного интеллекта в медицине. Как следствие, в России в прошлом году были приняты документы, разрешающие использование ИИ – не для постановки окончательного диагноза и назначения лечения, но для помощи врачам, например при расшифровке томограмм или рентгеновских снимков. И регулирование в этой области уже далеко продвинулось».

КАК СУДИТ МАШИНА? Отдельным блоком проблем можно считать вопросы, связанные с неприкосновенностью частной жизни, доступом к персональным данным, использованием ИИ в судебной и правоохранительной системах. «При создании систем искусственного интеллекта важно обращать внимание на обеспечение безопасности пользовательских данных, справедливость и недискриминационный характер принятия решений, прозрачность, контролируемость и понятность функционирования ИИ, – рассказывает Алексей Сидорюк, директор направления «Искусственный интеллект» АНО «Цифровая экономика». – В мировой практике при разработке ИИ уже был допущен ряд серьезных ошибок, которые влекли за собой дискриминационные решения. Например, в судах США ИИ в большинстве случаев считал темнокожих людей виновными, поскольку исторически они чаще получали обвинительные приговоры. Другой пример – очернение мусульман в языковом алгоритме GPT-3: этот алгоритм обучался на новостях из интернета, в которых мусульмане нередко упоминаются в связи с угрозой терроризма. В ряде моделей финансового скоринга дискриминировались люди, у которых нет соцсетей: алгоритмы ИИ не могли проверить информацию о человеке и отказывали в выдаче кредитных продуктов». В сфере общественной безопасности вызывают тревогу основанные на ИИ

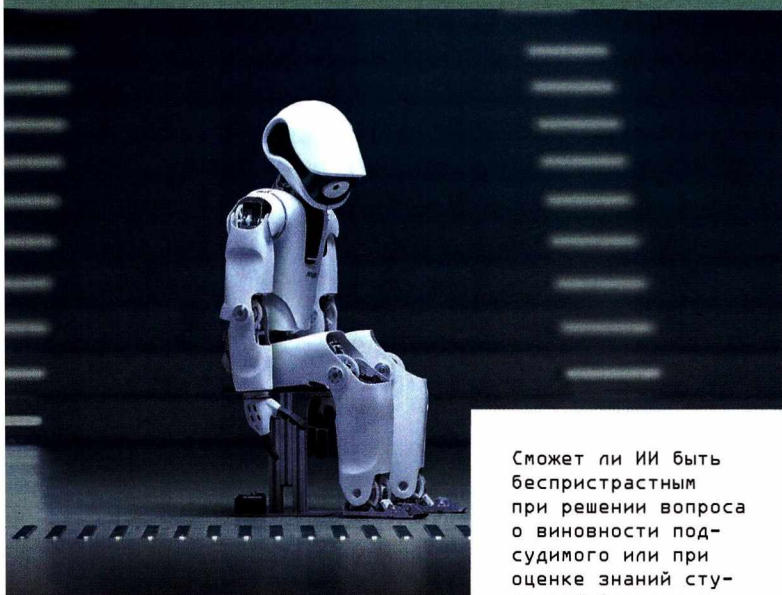
технологии deepfake – создание видео с фальшивыми, но при этом реалистичными изображениями тех или иных людей. Такие технологии могут использоваться в целях социального инжиниринга (подмены личности), киберпреступности, банального шантажа и вымогательства. С другой стороны, обеспечение общественной безопасности за счет алгоритмов компьютерного зрения, когда система распознает конкретных людей на изображениях с установленных в публичных местах камер, создает психологический эффект страха перед «большим братом». В Европе есть тенденция к ограничению работы подобных средств.

ЦИФРОВАЯ ПРОФЕССУРА Казалось бы, использование ИИ в образовании можно оценивать исключительно положительно, однако и здесь есть свои подводные камни. «Искусственный интеллект расширяет спектр возможностей в этой сфере: обеспечивает инклюзивный доступ, позволяет людям учиться на программах и курсах, подходящих именно им, вне зависимости от места жительства, – рассуждает Нина Яныкина из «Университета 2035». – Каждый студент обучается в соответствии со своими целями, в своем темпе и именно тем дисциплинам, которые ему нужны. Чтобы обеспечить такие индивидуальные образовательные траектории в массовом образовании, например в вузе, где тысячи, а то и десятки тысяч студентов, необходимо собирать данные, формировать цифровой след учащихся, анализировать его и принимать решения. Вручную это сделать невозможно – требуется использовать технологии ИИ и анализа больших данных. Здесь они незаменимы. Однако встает вопрос о том, кто в таком случае будет принимать решения и нести ответственность за результат. Если искусственный



Технологии deepfake могут использоваться для подмены личности, шантажа и вымогательства.

ДЛЯ ОРГАНИЗАЦИИ ИНДИВИДУАЛЬНОГО ОБУЧЕНИЯ ТЫСЯЧ СТУДЕНТОВ ОДНОВРЕМЕННО **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НЕЗАМЕНИМ.**



Сможет ли ИИ быть беспристрастным при решении вопроса о виновности подсудимого или при оценке знаний студента? Окончательного ответа на этот вопрос пока нет.

Ключевые опасения – риск мошеннических действий, принятие ошибочных решений, замена людей машинами и последующая безработица. Треть россиян опасаются замены человека искусственным интеллектом в их профессии, однако две трети будут рады сокращению рабочего дня благодаря ИИ при сохранении прежнего уровня дохода. В целом россияне ожидают от ИИ не указаний, а консультаций и советов – с тем, что окончательное решение останется все же за человеком. А от применения новых технологий в сфере управления люди ждут большей честности и справедливости».

ВСТРОЕННАЯ ЭТИКА Мы внедряем искусственный интеллект, потому что во многих случаях он превосходит нас в скорости обработки и анализа информации, действует точнее, исключает огрехи человеческого фактора. В каком-то смысле он обретает «право быть человеком», сотрудничая и общаясь с людьми на равных. Однако этика взаимоотношений, представление о добре и зле, ценности сопереживания и сотрудничества сформировались отчасти в ходе длительной эволюции Homo sapiens как биологического вида, отчасти в ходе развития цивилизации. Мы не можем дать системам ИИ столько же времени: существовать рядом с этически несовершенной машиной будет для нас и некомфортно, и опасно. Именно поэтому сейчас как никогда актуально создание таких документов, как Национальный кодекс этики искусственного интеллекта. «Кодекс – это концентрация норм, естественных для человека, – говорит Сергей Наквасин, заместитель руководителя Аналитического центра, руководитель Центра экспертизы по реализации федерального проекта «Искусственный интеллект». – В документе содержатся те правила, которые мы ценим и которые необходимо заложить в алгоритмы ИИ для того, чтобы люди могли таким системам доверять. Человек – высшая ценность, и все, что делается вокруг человека, должно быть направлено на его благо». **ИИМ**

интеллект проанализировал информацию о студенте и дал определенные рекомендации, обязательно ли студент должен им следовать? Как обеспечить не только сохранение, но и развитие когнитивных функций человека при массовом применении технологий ИИ в обучении? Также необходимо понимать, что при использовании искусственного интеллекта в образовании личные данные учащихся должны быть защищены».

НАДЕЖДА НА СПРАВЕДЛИВОСТЬ Насколько осознает наше общество те этические проблемы, которые активно обсуждают эксперты? С этим вопросом мы обратились к Юлии Грязновой, руководителю дирекции стратегии и аналитики АНО «Национальные приоритеты». «Любое новое явление вызывает беспокойность – если не у всего населения, то у каких-то групп. Это нормально, – говорит Юлия. – Однако в России страхи и надежды, связанные с ИИ, количественно находятся в равновесии.