Злонамеренное использование ИИ и вызовы информационно-психологической безопасности в странах БРИКС

Доклад

Координатор исследовательского проекта: Е. Н. Пашенцев

Издание Международного центра социально-политических исследований и консалтинга при поддержке Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI)

Июнь 2024, Москва

Злонамеренное использование ИИ и вызовы информационно-психологической безопасности в странах БРИКС

Доклад

Координатор исследовательского проекта: Е. Н. Пашенцев

Издание Международного центра социально-политических исследований и консалтинга при поддержке Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI)

Июнь 2024, Москва

УДК 004.81-027.584(042.3)
ББК 16.6+16.84
3-68

Настоящий доклад публикуется на английском, китайском и русском языках под общей редакцией Е. Н. Пашенцева.

Редакционная коллегия издания на русском языке: Д. Ю. Базаркина, Е. Н. Пашенцев, С. А. Себекин

Перевод глав «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Арабской Республике Египет», «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Федеративной Республике Бразилия», «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Китайской народной республике», «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Республике Индия», «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Южно-Африканской Республике», «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Российской Федерации», введения и заключения с английского на русский язык: С. А. Себекин.

Главы «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Исламской Республике Иран», «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Федеративной Демократической Республике Эфиопия», «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Королевстве Саудовская Аравия», «Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Объединенных Арабских Эмиратах» предоставлены редколлегии на русском языке.

Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в странах БРИКС. Координатор исследовательского проекта: Е. Н. Пашенцев. Издание Международного центра социально-политических исследований и консалтинга при поддержке Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI). – Москва: ООО «САМ Полиграфист», 2024. – 148 с.

ISBN 978-5-00227-263-1

Технологии искусственного интеллекта (ИИ) обладают огромной преобразующей силой и уже привели к многочисленным позитивным изменениям в странах БРИКС, но также сопряжены с большими рисками, не в последнюю связанные с деятельностью различных злонамеренных акторов. Злонамеренное использование ИИ растет в современном мире в количественном и качественном отношении, представляя серьезную опасность для жизни, здоровья и благосостояния людей. Оно способно, как усилить риски применения всех современных технологий, так и создать новые, которые невозможно полностью предвидеть. В настоящем докладе основное внимание уделяется угрозам злонамеренного воздействия ИИ на психику человека, а через это на политические, экономические, социальные процессы и деятельность государственных и негосударственных институтов в десяти странах БРИКС.

Изображение на обложке: Freepik.
Подписано в печать 04.06.2024. Цифровая печать. Заказ №32105.

Отпечатано в типографии «OneBook.ru» ООО «САМ Полиграфист».
109316, г. Москва, Волгоградский проспект, дом 42, корп. 5, Технополис Москва.
www.onebook.ru

Оглавление

Предисловие к русскому изданию

Экспансия искусственного интеллекта (ИИ) неутомимо захватывает новые области человеческой и сверхчеловеческой деятельности. ИИ не только превзошел возможности человека в интеллектуальных играх, но и позволяет с фантастической точностью анализировать различные богатые данные о пациентах и их состоянии, осуществлять эффективный поиск полезных ископаемых или обнаруживать так называемые нестабильности режима разрыва плазмы при ядерном синтезе за 300 миллисекунд до их возникновения. Не за горами тот день, когда ИИ сможет создавать непревзойденные шедевры в музыке или литературе. Но есть и обратная, злонамеренная сторона использования ИИ. Информационные войны уже давно стали своей сферой для ИИ, и новое информационно-психологическое оружие делает злонамеренное использование ИИ реальной угрозой для развития всего человечества. Модели ИИ, например, научили технологиям обмана, точно распознавать эмоции человека по его голосу и т. д. При этом скорость развития ИИ растет по экспоненте. Ведь если протеиновая история развития живого на Земле измеряется по меньшей мере периодом в 540 млн лет, то современные скорости вычислений ускоряют процесс развития «искусственного живого» на множество порядков.

Так, «ангельские» и «дьявольские» успехи ИИ в последние годы в основном обязаны скачкообразному росту производительности суперкомпьютеров, уже покоривших экзафлопсный (число с18-ю нолями) диапазон скорости вычислений, а также внедрению научных находок типа авторегрессионного ИИ (ChatGPT, LLM), который может отвечать на самые каверзные вопросы. Значительное место в этих успехах занимают перекрестные исследования в различных сферах человеческой деятельности, когда обучающие выборки для систем ИИ могут быть искусственно синтезированы или взяты из смежных областей.

Вместе с тем, ИИ остается пока «слабым» и «узким». Если ИИ хорошо играл в шахматы, то для игры в Го потребовалось создание другого ИИ, который, проиграв сам с собой десятки миллионов партий, стал чемпионом мира среди людей. ИИ пока нельзя полностью доверить вождение автомобиля, поскольку только человек может принять правильное решение в неподдающейся логике ситуации. Системы ИИ явно нуждаются в усилении своих объяснительных возможностей.

Ограничения современного ИИ пока играют позитивную роль в процессе сокращения угроз информационно-психологической безопасности – наиболее опасного фактора на пути развития сообществ людей, стран и континентов. Но эти ограничения могут быть быстро снесены благодаря тому же росту возможностей вычислительной техники, появлению новых средств вычислений, например, базирующихся на нейроморфном симбиозе биологических и технических материалов, построению новых суперкомпьютеров на квантовой и фотонной основе, и, конечно, благодаря развитию научно-практической базы информационно-психологического противоборства. Когда возрастет скорость вычислений на 10–12 порядков и разовьются когнитивные возможности ИИ, главенство человека на нашей планете Земля и его счастливое будущее будут под большим сомнением.

Доклад «Злонамеренное использование ИИ и вызовы информационно-психологической безопасности в странах БРИКС» делает угрозу апокалипсиса вследствие недосмотра в развитии ИИ существенно ниже благодаря всеобъемлющему исследованию вопросов информационно-психологической безопасности по многоаспектному спектру проблем. Изучение вопросов злонамеренного использования ИИ все больше становится неотъемлемой необходимостью работы ученых, практиков, властных структур и обычных людей, желающих с уверенностью смотреть в свое будущее.

А. Н. Райков,

доктор технических наук, профессор,
главный научный сотрудник Цзинанского
института суперкомпьютерных технологий,
Китай, провинция Шаньдун

10 июня 2024 г.

Введение: злонамеренное использование ИИ – угрозы возрастают

Е.Н. Пашенцев

Страны БРИКС переживают стремительное развитие и внедрение технологий искусственного интеллекта (ИИ), что, принося очевидные экономические и социальные выгоды, приводит к радикальным изменениям в производстве, финансах, торговле, транспорте, образовании, медицине и сфере досуга. Эти технологии также оказывают все большее влияние на функционирование государственных учреждений, политических партий и общественных организаций. Передовые возможности ИИ обладают огромной социальной силой и способны привести к многочисленным позитивным изменениям в обществе, но и сопряжены с большими рисками.

«Страны БРИКС достигли соглашения в ближайшее время создать исследовательскую группу по искусственному интеллекту в рамках Института БРИКС по изучению сетей будущего… Нам необходимо совместно противостоять рискам и разработать системы и стандарты управления искусственным интеллектом на основе широкого консенсуса, чтобы сделать технологии искусственного интеллекта более эффективными, безопасными, надежными, контролируемыми и справедливыми», — заявил председатель КНР Си Цзиньпин на XV-ом саммите БРИКС в ЮАР (CGTH 2023). На последней конференции AI Journey, проходившей в Москве 22–24 ноября 2023 г. и посвященной развитию больших данных и машинного обучения, Президент России В.В. Путин отметил необходимость взвешенного и обоснованного регулирования ИИ, которое будет использоваться в интересах всех, а не отдельных стран (Президент России 2023). В Перечне поручений по итогам конференции, утвержденном президентом и опубликованном на сайте Кремля в январе 2024 г., сказано о необходимости обеспечить включение вопросов, касающихся формирования этических стандартов в области ИИ, сбалансированного регулирования и научно-технического сотрудничества в указанной области, в повестку дня заседаний объединения БРИКС в рамках председательства Российской Федерации в объединении в 2024 году (Президент России 2024). Выступая на саммите Глобального партнерства по искусственному интеллекту (Global Partnership on Artificial Intelligence, GPAI) в декабре 2023 г., премьер-министр Индии Н. Моди заявил, что в XXI в. ИИ может стать как крупнейшим инструментом развития, так разрушительной силой. «Мы должны работать сообща ради того, чтобы подготовить глобальную основу для этичного использования ИИ», — подчеркнул Н. Моди (PTI 2023).

Между тем в странах БРИКС растет практика злонамеренного использования ИИ (ЗИИИ), что отражает общемировую тенденцию. ЗИИИ может как усиливать известные риски применения интеллектуальных систем, так и создавать новые, в том числе и те, которые невозможно предвидеть сегодня. Потенциал ЗИИИ вызывает большую тревогу. Указывая на это, исследователи из Collaborations Pharmaceuticals в сотрудничестве с научными учреждениями Европы провели концептуальный эксперимент. Вместо синтеза новых лекарств они ввели в нейросеть MegaSyn AI запрос сделать обратное: выявить наиболее токсичные для человеческого организма вещества. Нейросеть правильно интерпретировала поставленную перед ней задачу и менее чем за шесть часов сгенерировала список из 40 000 веществ, которые могут являются оптимальными компонентами для создания химического и биологического оружия. ИИ самостоятельно разработал не только многие известные боевые отравляющие вещества, но и множество новых, более токсичных. Эта простая инверсия модели машинного обучения превратила безобидную генеративную модель из полезного инструмента в смертоносное оружие (Urbina et al. 2022). Разумно предположить, что данный инверсионный подход может быть применен и к другим областям,

например, к поиску эффективных способов деструктивного воздействия на общественное сознание.

Важно адекватно оценить реальную угрозу ЗИИИ как средства информационно-психологической борьбы на глобальном уровне. ЗИИИ способствует дестабилизации человеческого сознания, что, в свою очередь, дестабилизирует общество и еще больше деформирует человеческое сознание, облегчая все более опасное использование ИИ. Тем самым создается порочная цепь взаимосвязанных воздействий, направленная на дальнейшую концентрацию общественных богатств в руках очень узкого круга сверхбогатых и “победу” в глобальном перераспределении активов в рамках этого круга. В то же время объективные проблемы человечества из-за эгоистических расчетов, действий антисоциальных акторов, инерции и неспособности большей части общества осуществить в своих же интересах необходимые перемены, не только остаются нерешенными, но очевидным и опасным образом обостряются.

Военно-политические блоки и их агрессивная политика не раз в истории приводили мир к разрушительным войнам, включая две мировые войны в XX в. Однако существуют и другие союзы государств, которые ищут альтернативу нынешнему социально-экономическому миропорядку, который отличает безудержная гонка вооружений, чреватая ядерным Армагеддоном. На то, чтобы стать опорой нового миропорядка, претендует БРИКС, связывающий нынешнюю неустойчивую и опасную ситуацию с лучшим будущим, предлагая большие возможности на глобальном уровне (Pashentsev and Miao 2023). Даже в нынешнем формате БРИКС объединяет страны с более чем 45% населения земного шара, 32% мирового ВВП по паритету покупательной способности (против 30% у стран «Большой семерки»), огромным научно-технологическим потенциалом и крупнейшими природными ресурсами, что позволяет решать глобальные проблемы в интересах всего человечества. Также следует учитывать, что существует около трех десятков стран, которые готовы в том или ином качестве присоединиться к БРИКС (TV BRICS 2024).

Совершенствование и внедрение технологий ИИ могут сыграть важную роль в решении многочисленных и сложных задач развития стран БРИКС, а также в обеспечении их безопасности от враждебных действий государственных и негосударственных субъектов, в том числе в области информационно-психологической безопасности.

Злонамеренное использование искусственного интеллекта и три уровня угроз информационно-психологической безопасности

ЗИИИ носит характер преднамеренного антисоциального действия, которое может проявляться как в явной, так и в скрытой форме. Злонамеренные акторы (от отдельных преступников до влиятельных групп интересов и коррумпированных государственных институтов) уже используют ЗИИИ для достижения своих целей. В последние годы ЗИИИ продемонстрировало большой потенциал для подрыва информационно-психологической безопасности. Несмотря на значительное и быстро растущее число научных публикаций, посвященных техническим аспектам ЗИИИ, его общим социально-экономическим и политическим последствиям, а также первым попыткам классификации ЗИИИ (Brundage et al. 2018; Caldwell et al. 2020; Malicious Uses 2020; Blauth et al. 2022), существует мало исследований, посвященных системному рассмотрению ЗИИИ в контексте информационно-психологической безопасности (первые из них: Bazarkina and Pashentsev 2019; Pashentsev 2019; Пашенцев 2019 и др.).

Можно предложить следующую классификацию ЗИИИ по степени реализации его возможностей:

• существующая практика ЗИИИ;

• существующие возможности ЗИИИ, которые еще не были использованы на практике (такая вероятность связана с широким спектром быстро развивающихся новых возможностей ИИ – не все они сразу входят в спектр реализованных возможностей ЗИИИ);

• будущие возможности ЗИИИ на основе текущих разработок и будущих исследований (оценка должна быть дана на ближайшую, среднесрочную и долгосрочную перспективы);

• неопознанные риски, также известные как «неизвестное в неизвестном». Не все разработки в сфере ИИ можно точно оценить. Готовность встретить неожиданные скрытые риски имеет решающее значение.

Понятие «информационно-психологическая безопасность» можно встретить во многих исследованиях (Grachev, 1998; Roshhin and Sosnin, 1995; Afolabi and Balogun, 2017). Известный психолог из США А. Маслоу полагал, что после удовлетворения базовых физиологических потребностей на первый план выходит потребность в безопасности. Если рассматривать ее более подробно, то можно сказать, что это потребность в защите, стабильности, уверенности в будущем, потребность сохранения здоровья и др. Кроме личной человек нуждается и в общественной безопасности: предпочитает известность неизвестности, стремится к уверенности, что его / ее окружение безопасно и свободно от угроз (см. подробнее: Maslow, et al., 1945). Под обеспечением национальной информационно-психологической безопасности понимается защита граждан, отдельных групп и социальных слоев, массовых объединений людей и населения страны в целом от негативного информационно-психологического воздействия (Баришполец 2013, с. 63; см. подробнее: Баришполец (ред.) 2012).

Опираясь на приведенные определения, автор считает возможным определить международную информационно-психологическую безопасность как защиту системы международных отношений от негативных информационно-психологических воздействий, которые связаны с различными факторами международного развития. Среди последних выделим целенаправленную деятельность различных государственных, негосударственных и наднациональных акторов по частичной/полной, локальной/глобальной, кратковременной/долгосрочной, латентной/открытой дестабилизации международного положения с целью получения конкурентных преимуществ вплоть до физического уничтожения противника.

При всей своей значимости частный анализ злонамеренного информационно-психологического воздействия посредством дипфейков, ботов, предиктивной аналитики и т.п. не учитывает синергию воздействия различных технологий ИИ и не дает системного представления о росте рисков информационно-психологической безопасности, равно рисках для всей системы национальной и международной безопасности. Отсутствие всестороннего анализа объясняется новизной проблемы: практика ЗИИИ не может опережать прогресс в области развития ИИ.

Первым шагом на пути осознания проблемы и противодействия ЗИИИ является консолидация усилий ученых разных стран в этой новой области, начало чему было положено созданием в 2019 г. Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI) на международном семинаре в Институте актуальных международных проблем Дипломатической академии МИД РФ и сотрудничеству между исследователями из семи стран (сотрудничество включает в себя совместные исследования и публикации, проведение научных семинаров, конференций и т.д.). Результатом написания многочисленных академических статей группой исследователей по различным вопросам ЗИИИ и информационно-психологической безопасности стала публикация книги «The Palgrave Handbook of Malicious Use of AI

and Psychological Security», которая является первым фундаментальным трудом в этой новой области, авторами которой стали 23 исследователя из 11 стран Азии, Европы и Северной Америки (The Palgrave Handbook, 2023).

В последние годы опубликованы как первые комплексные исследования угроз информационно-психологической безопасности, обусловленных ЗИИИ, в странах БРИКС (Bazarkina and Pashentsev 2020, pp. 154-177; Pashentsev and Bazarkina 2023), так и публикации, посвященные рискам ЗИИИ в отдельных странах-членах БРИКС (Bazarkina and Matyashova 2022, pp. 14-20; Bazarkina et al. 2023; Cai and Zhang 2023; Gupta and Guglani 2023, и т. д.).

В настоящем докладе основное внимание уделяется угрозам для человеческой психики посредством ЗИИИ и, как следствие, для политических, социальных, экономических, культурных процессов и деятельности государственных и негосударственных институтов в странах БРИКС. Впервые комплексно рассмотрены угрозы ЗИИИ с учетом недавнего расширения состава этого объединения. В данном докладе не рассматриваются меры противодействия ЗИИИ, так как они находятся на стадии формирования и требуют самостоятельного анализа[1]. Доклад берет за основу анализа трехуровневую классификацию угроз информационно-психологической безопасности посредством ЗИИИ (см. подробнее: Pashentsev 2023).

На первом уровне уровне сами технологии ИИ напрямую не применяются для дестабилизации информационно-психологической безопасности, а угрозы связаны с заведомо искаженными интерпретациями обстоятельств и последствий развития и применения этих технологий в интересах антисоциальных акторов. Деструктивное (открытое или скрытое) воздействие формирует в сознании людей ложный образ ИИ. Целевым образом создаваемые и распространяемые версии негативных последствий развития ИИ на основе взаимодействия общественных стереотипов, существующих и вероятных рисков (например, мифологизированные канонические сценарии, согласно которым роботы и ИИ очень скоро отнимут у людей все рабочие места и встанут над ними и т. п.) преследуют определенные политические и экономические цели и могут представлять собой реальную угрозу информационно-психологической безопасности. Неадекватная общественная реакция способна привести к снижению темпов внедрения технологий ИИ, создать социально-политическую напряженность и привести к конфликтам. Также могут проявиться завышенные общественные ожидания в отношении ИИ, что на определенном этапе может привести к обвальному падению стоимости высокотехнологичных компаний и рынка в целом. Эти ожидания могут использоваться в деструктивных целях и преднамеренно усиливаться для дезориентации широких слоев общественности, коммерческих и некоммерческих структур, органов государственной власти и, в конечном итоге, обернуться разочарованиями, неправильными решениями, социальными и политическими конфликтами.

Второй уровень угроз представлен широким спектром возможностей: нецелевое использование дронов, кибератаки на критическую инфраструктуру, переориентация интеллектуальных систем коммерческого или финансового назначения, использование технологий ИИ для нарушения процесса принятия решений или его скрытой модификации и другие, еще более деструктивные, варианты действий. Но атака на общественное сознание не является главной целью ЗИИИ на этом уровне.

ЗИИИ, применяемое в первую очередь для нанесения информационно-психологического ущерба, относится к третьему (и высшему) уровню угроз информационно-психологической безопасности. Продукты синтетического ИИ (объединяющие ряд технологий, способные увеличить

[1] В то время как одни страны БРИКС находятся на стадии первоначального и фрагментарного понимания остроты этих угроз, другие принимают первые законодательные акты противодействия и соответствующие организационные и технические контрмеры.

ущерб в случае злонамеренного использования) создают широкий спектр новых рисков и угроз. Профессиональное использование инструментов и методов информационно-психологического противоборства может сместить уровень восприятия угрозы выше или ниже приемлемого. Более того, использование ИИ в информационно-психологическом противоборстве делает латентные кампании по управлению восприятием более опасными, и в обозримом будущем ситуация будет только ухудшаться. Поэтому ЗИИИ, направленное, прежде всего, на причинение вреда в области информационно-психологической безопасности, заслуживает отдельного пристального внимания.

Первые два уровня угроз в разной степени влияют на человеческое сознание и поведение, они могут иметь и катастрофический эффект для всего человечества (как это бы произошло в случае Третьей мировой войны). Однако воздействие угроз третьего уровня на определенном этапе может способствовать усилению влияния антисоциальных групп на общественное сознание (или даже контролю), что может привести к внезапной дестабилизации как ситуации в конкретной стране, так и международной обстановки в целом. В конечном итоге, если на третьем уровне обеспечивается надежный информационно-психологический контроль над противником, роль двух других уровней ЗИИИ в подрыве информационно-психологической безопасности становится вспомогательной.

Угрозы ЗИИИ могут возникать на одном уровне воздействия или одновременно на нескольких уровнях в рамках единой кампании по воздействию на сознание и управлению восприятием. Планирование и/ или осуществление террористами нападения на мирное население с использованием технологий ИИ будет угрозой второго уровня, имеющей свой коммуникативный и информационно-психологический эффект в ходе реализации (паника и шок после атаки). Однако, если преступники будут сопровождать свои действия применением ИИ в качестве специализированного инструмента информационно-психологического воздействия, угроза достигнет третьего уровня или, по крайней мере, будет носить пограничный с третьим уровнем характер (многое зависит от масштабов и долговременности поддерживаемого технологиями ИИ воздействия). Угрозы первого уровня все реже встречаются в чистом виде, поскольку технологии ИИ активно используются для презентации их возможностей в информационном поле, включая и заведомо искаженные представления.

ИИ — это не одна конкретная технология, а целый спектр технологий, применяемых для решения определенных задач посредством различных приложений в разных средах и модальностях при разных обстоятельствах. Авторы доклада принимают во внимание тот факт, что технологии под общим названием «ИИ» помогают создать тот или иной продукт, серьезно меняющий практические возможности того или иного вида деятельности.

Угрозы, продуцируемые ЗИИИ, становятся все более актуальными во всем мире на всех трех уровнях по мере роста геополитического соперничества, активности различных государственных и негосударственных антисоциальных акторов, а также развития и растущей доступности технологий ИИ, что делает возможным их широкое злонамеренное применение. Попытки манипулировать общественным сознанием особенно разрушительны в исторические кризисные моменты. Бесчеловечность фашизма стала очевидна абсолютному большинству человечества после гибели свыше пятидесяти миллионов человек в годы Второй мировой войны. Однако перед войной именно манипуляция общественным сознанием обеспечила победу Гитлера на выборах в Рейхстаг 1933 г. Эта не столь далекая история остается весьма поучительной для тех, кто жив сегодня. Понятно, что современные правительства и политические деятели БРИКС и многих других стран демонстрируют растущую обеспокоенность по поводу угрозы высокотехнологичной дезинформации в Интернете и роли ведущих частных медиа-платформ, использующих ИИ-технологии.

Угрозы ЗИИИ в целях подрыва информационно-психологической безопасности в странах БРИКС возникают как по внутренним, так и по внешним причинам. Поэтому здесь имеет смысл дать некоторое общее представление о природе и динамике угроз на трех уровнях в глобальном измерении.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

На первом уровне угроз кампании по формированию ложного образа ИИ могут опираться на все более негативное восприятие дальнейшего развития и использования технологий ИИ в современном мире, что особенно характерно для стран Запада. Согласно опросу Pew Research Center, проведенному в августе 2023 г., 52% граждан США больше обеспокоены, нежели воодушевлены ростом внедрения использования ИИ. Лишь 10% отмечают, что они больше воодушевлены, чем обеспокоены, а 36% говорят, что испытывают эти эмоции в равной степени. Доля американцев, которые больше всего обеспокоены использованием ИИ в повседневной жизни, выросла на 14% с декабря 2022 г., когда такое мнение высказали 38%. Рост беспокойства по поводу применения технологий ИИ происходил одновременно с ростом осведомленности общественности. Девять из десяти взрослых были либо хорошо (33%), либо плохо (56%) осведомлены о развитии и внедрении ИИ технологий. Доля тех, кто хорошо осведомлен об ИИ, выросла на 7% по сравнению с декабрем 2022 г. Число тех, кто много слышал об ИИ и выражает больше беспокойства, чем воодушевления по этому поводу, в августе 2023г. стало на 16% больше, чем в декабре 2022 г. Среди этой наиболее осведомленной группы беспокойство перевешивает воодушевление: 47% против 15%. В декабре 2022 г. этот показатель составлял: 31% против 23% (Tyson and Kikuchi 2023).

В сотрудничестве с Университетом Квинсленда международная аудит-консалтинговая корпорация KPMG представила в 2023 г. первое в мире исследование уровня доверия и отношения к ИИ в глобальном масштабе. Авторы исследования опросили более 17 тыс. человек из 17 стран, охватывающих все регионы мира: Австралию, Бразилию, Канаду, Китай, Эстонию, Финляндию, Францию, Германию, Индию, Израиль, Японию, Нидерланды, Сингапур, ЮАР, Южную Корею, Великобританию и США. Эти страны являются лидерами в области развития и внедрения ИИ в своих регионах. Согласно исследованию KPMG, в странах БИКС[2] большинство людей (56–75%) доверяют системам ИИ, причем наибольшая степень доверия наблюдается среди жителей Индии, за которой следует Китай. Напротив, в некоторых других странах отмечается более низкая степень доверия ИИ. Например, самый низкий уровень доверия ИИ был зарегистрирован в Финляндии (всего 16%). Более высокое доверие и признание ИИ в странах БИКС, вероятно, связано с ускоренным внедрением ИИ в этих странах и все более важной ролью новых технологий в развитии экономики. Жители стран БИКС наиболее позитивно относятся к развитию ИИ, извлекают из него наибольшую выгоду и сообщают о самых высоких уровнях внедрения ИИ в рабочие процессы (Gillespie et al. 2023, p. 14). Жители западных стран и Японии испытывают определенные сомнения касательно того, что что преимущества от внедрения ИИ перевесят риски. Напротив, большинство людей в странах БИКС (а также в Сингапуре) считают, что выгоды превалируют над рисками (The Global Risks Report 2024, p. 3).

Недоверие или страх перед ИИ, возможно, вызваны резким падением доверия к основным общественным институтам на Западе. Например, Институт Гэллапа (American Institute of Public

[2] Авторы исследования KPMG не включили Россию в свою выборку, и поэтому используют в представленном отчете аббревиатуру БИКС (BICS). С информацией об отношении к ИИ в России и в новых членах БРИКС можно будет ознакомится в соответствующих главах настоящего отчета.

Opinion) зафиксировал в 2022 г. в США значительное снижение общественного доверия к 11 из 16 институтов, которые он отслеживает ежегодно, причем больше всего пострадали институт президентства и Верховный суд. Доля американцев, выразивших достаточно высокий уровень доверия к ним, упала на 15% и 11% для института президентства и Верховного суда соответственно. По данным последнего опроса, проведенного 1–22 июня 2023 г., ни один из показателей существенно не восстановился до прежнего уровня: доверие к Суду сейчас составляет 27%, а к институту президентства – 26%. Пять институтов с худшим рейтингом – газеты, система уголовного правосудия, телевизионные новости, крупный бизнес и Конгресс США – вызывают доверие менее чем у 20% американцев. При этом уровень доверия к одному из главнейших институтов США – Конгрессу – достигает значения лишь в 8% – это единственный институт, показатель доверия которому выражается однозначным числом. Большинство учреждений, попавших в рейтинг в 2023 г., находятся в пределах трех процентных пунктов от своего рекордно низкого показателя доверия, в том числе четыре, которые находятся на рекордно низком уровне или близки к нему. Это полиция, государственные школы, крупные технологические компании и крупный бизнес (Saad 2023).

Можно ли ожидать, что общественность поверит в способность власти добиться социально-ориентированного использования ИИ при таких показателях доверия к основным общественным институтам? Ответ очевиден. Не доверяют не ИИ (сегодня это только машинный интеллект), а власти, крупному бизнесу, технологическим гигантам, которые направляют его развитие. Почти непримиримый политический раскол (особенно очевидный и стратегически опасный в США), вплоть до угрозы государственного переворота и гражданской войны (Marche 2022a, Pashentsev 2022, Walter 2022a), деградация правящего истеблишмента, разрушительная роль правящей олигархии (Collins et al. 2021; Gilens and Page 2014), агрессивная внешняя политика (Abelow 2022, Sachs 2018), низкие темпы экономического роста и острые социальные противоречия являются объективными показателями неспособности западных элит обеспечить безопасное развитие и применение ИИ, эффективное противодействие ЗИИИ. Логично предположить, что учреждения с таким уровнем доверия сами порождают злоумышленников и все чаще (по крайней мере частично) сами стимулируют ЗИИИ.

Опасения, связанные с потенциальными рисками будущего развития ИИ в негативном ключе, за последний год только усилились в связи с быстрым прогрессом генеративного ИИ на фоне углубляющегося глобального кризиса. Многие ведущие бизнес-лидеры серьезно обеспокоены тем, что ИИ может представлять реальную угрозу человечеству в не столь отдаленном будущем. Согласно результатам опроса, предоставленным эксклюзивно CNN, 42% руководителей, опрошенных на профильном саммите Йельского университета в июне 2023 г., говорят, что ИИ потенциально может уничтожить человечество через пять-десять лет. В опрос вошли ответы 119 руководителей различных секторов бизнеса, в том числе генерального директора Walmart Д. Макмиллона генерального директора Coca-Cola Д. Куинси, руководителей IT-компаний, таких как Xerox и Zoom, а также руководителей фармацевтических компаний, средств массовой информации и компаний в сфере промышленности. «Это довольно мрачно и тревожно», — сказал в телефонном интервью профессор Йельского университета Джеффри Зонненфельд, комментируя данные опроса (Egan 2023).

В отчете о глобальных рисках «Global Risks Report 2024» представлены результаты исследования восприятия глобальных рисков, в котором собраны мнения почти 1500 мировых экспертов, опрошенных в сентябре 2023 г. Большинство респондентов (54%) в ближайшие два года ожидают некоторой нестабильности и умеренного уровня риска глобальных катастроф, а 30% ожидают, что международная обстановка будет еще более турбулентной. Прогноз на десятилетнюю перспективу заметно более негативен: почти две трети респондентов ожидают бурных или неспокойных лет (Global Risks Report 2024, p. 6). В данном рейтинге глобальных рисков по уровню

серьезности угрозы за десятилетний период неблагоприятные последствия применения технологий ИИ занимают шестое место (Global Risks Report 2024, p. 8). В «Global Risks Report 2024» обращается внимание на тот факт, что «технологическая власть в руках неизбранных рассматривается… как более серьезная проблема, чем власть, сконцентрированная в правительстве. Влияние компаний Big Tech уже является транснациональным, конкурируя с национальными государствами, а генеративный искусственный интеллект будет продолжать усиливать влияние этих компаний и аффилированных с ними первых лиц» (Global Risks Report 2024, p. 54).

Крупнейшие IT-компании активно используют технологии ИИ в соответствии со своими узкими корпоративными интересами, которые довольно часто идут вразрез с интересами общества. Очевидно, что компании, имеющие доступ к большим объемам данных для создания моделей ИИ, лидируют в разработке ИИ. GAMAM (Google, Amazon, Meta[3], Apple, and Microsof), также известная как «Большая пятерка» — название, данное пяти крупнейшим и доминирующим компаниям в индустрии информационных технологий США. К лидерам в области ИИ можно также отнести «первопроходца» в области компьютерных технологий IBM и ведущих производителей аппаратного обеспечения Intel и NVIDIA (Lee 2021). Конечно, к крупнейшим IT-компаниям других стран также есть серьезные вопросы, но их текущая глобальная роль значительно ниже, чем у компаний США, чье быстрое обогащение, огромное влияние и экзистенциальные риски корпоративного контроля над перспективными формами ИИ вызывают растущую обеспокоенность во всем мире.

Вряд ли является случайностью, что в 2021 г. из 10-и самых богатых людей шестеро являлись представителями Amazon (1), Microsoft (2), Google (2) и Facebook (1) (Forbes 2021). Согласно Wall Street Journal, совокупная рыночная капитализация GAMAM на конец 2020 г. составляла $7,5 трлн. В конце 2019 г. совокупная рыночная капитализация этих компаний составила $4,9 трлн – это означает, что их стоимость выросла на 52% за год. По состоянию на 12 ноября 2021 г. капитализация этих компаний выросла еще на $2,5 трлн и достигла примерно $10 трлн (Statista 2021a). Это почти четверть совокупной рыночной капитализации всех компаний ($41,8 трлн), входящих в индекс S&P 500 (La Monica 2021). Уместно напомнить, что номинальный ВВП США в 2020 г. составил около $21 трлн. Япония, третья по величине экономика мира, имела ВВП около $5 трлн, а Россия – $1,5 трлн.

Лидирующая роль IT-компаний США и их стремление доминировать стали еще более очевидными после начала СВО на Украине. Помимо санкций, введенных правительствами стран Запада против России, IT-компании активно подключились к давлению на нашу страну за проводимую ей политику. Чтобы продемонстрировать поддержку Украины, большое число поставщиков технологий приостановило свой бизнес в России, включая Accenture, Adobe, Cisco, Oracle, Dell, IBM, Microsoft и многие другие (NS Business 2022; Fried 2022). Это, конечно, нанесло серьезный ущерб российскому IT-сектору и экономике в целом, но эти шаги также негативно повлияли на сами компании, ушедшие с российского рынка.

Первые месяцы 2022 г. оказались напряженными для крупных IT-компаний США, которые в значительной степени полагаются на цифровую рекламу. Растущая инфляция, украинский кризис и другие неблагоприятные макроэкономические факторы вынудили рекламодателей сократить маркетинговые бюджеты, что привело к снижению прибыли таких платформ, как YouTube, Google и Facebook (Cao 2022). Военные действия на Украине разрушили миф об их нейтралитете. На протяжении большей части своего существования интернет-компании утверждали, что они

[3] Компания Meta Platforms Inc., владеющая социальными сетями Facebook и Instagram, по решению суда от 21.03.2022 признана экстремистской организацией, ее деятельность на территории России запрещена.

являются лишь нейтральными платформами распространения контента и что они не несут ответственности за распространяемый контент (Feldstein 2022).

Отказавшись от нейтралитета и понеся при этом значительные потери, ведущие IT-компании извлекли и немалые выгоды из сложившейся ситуации.

Во-первых, крупнейшим IT- корпорациям США удалось избежать формирования единого фронта государств и широкой международной общественности по противостоянию Big Tech и нарастающей конфронтации с ним по принципиальным вопросам. Такая конфронтация могла возникнуть из-за того, что эти корпорации вступают в борьбу с государственными и негосударственными акторами, у которых во многом другие устремления. Однако сейчас и в ближайшем будущем компаниям Big Tech, возможно, в меньшей мере надо будет бояться глобальных инициатив или коалиций, которые могут возникнуть с целью ограничить их растущие аппетиты и влияние. ООН и многие другие международные структуры фактически парализованы острыми геополитическими противоречиями.

Во-вторых, ведущие IT-компании США продемонстрировали, что они выступают в качестве мощного инструмента проведения кибероперациий, направленных против России (см. подробнее главу по России настоящего доклада). Таким образом, формирование информационной повестки дня в Соединенных Штатах, которая сегодня немыслима без полноценного использования технологий ИИ, оказалось открыто подчинено военно-политическим интересам и потребностям информационно-психологической войны. «Они действительно “стреляют”! Это экстраординарно», — заявил Мэтью Шмидт, доцент кафедры национальной безопасности Университета Нью-Хейвена, обвинив западные IT-компании в их активном вовлечении в военные конфликты (Global Times 2022).

В-третьих, с точки зрения внутренней политики США, ресурс, который является безусловной необходимостью для защиты национальной безопасности и который уже активно используется в военных действиях, априори не может быть антинациональным, что затрудняет публичную критику Big Tech.

В-четвертых, любому правительству США в период острого геополитического противоборства понадобится информационная и аналитическая поддержка, которую IT-компании могут оказать, используя последние разработки в области ИИ, включая противодействие «внутренним» врагам и «дезинформации». Согласно документам, опубликованным 31 августа 2022 г., более пятидесяти чиновников администрации президента Байдена из дюжины агентств участвовали в попытках оказать давление на ведущие IT- компании, чтобы они осуществляли противодействие предполагаемой дезинформации. Документы были частью предварительного слушания в иске против правительства, поданном генеральными прокурорами Миссури и Луизианы, к которому позже присоединились эксперты, оклеветанные федеральными чиновниками. «Когда федеральное правительство вступает в сговор с крупными технологическими компаниями с целью цензурирования высказываний, американцы становятся подданными, а не гражданами», — заявил в своем послании генеральный прокурор Луизианы Джефф Лэндри (Stieber 2022).

В-пятых, процветание Big Tech под «зонтиком» ВПК в период новой холодной войны может компенсировать потери от ухода с российского рынка. В условиях холодной войны легче избежать общественного внимания и скандалов из-за многообещающих разработок, которые не только несут большие прибыли, но и серьезные риски для человечества.

Ухудшение международной обстановки, продолжение боевых действий на Украине, кровавый конфликт в Газе, а также другие военные конфликты, энергетический кризис, чередование рецессии со слабым ростом в Евросоюзе, нарушение цепочек поставок и другие факторы при-

вели к снижению рыночной стоимости ведущих IT-компаний, но не устранили их доминирующую роль в крупном бизнесе. В контексте нынешней парадигмы создания все более масштабных и мощных систем ИИ возможности для разработки алгоритмов без Big Tech существенно ограничены. За редким исключением, каждый стартап и даже исследовательские лаборатории в области ИИ зависят в США и далеко за их пределами от GAMAM. Все они полагаются на вычислительную инфраструктуру Microsoft, Amazon и Google для обучения своих интеллектуальных систем, а также на обширный потребительский рынок этих компаний для реализации своих продуктов ИИ (Kak et al. 2023). Даже решение Илона Маска купить Twitter в октябре 2023 г. во многом было мотивировано его намерением развивать собственные стартапы в области ИИ на основе возможностей Twitter по работе с большими данными. В том же году И. Маск объявил о создании стартапа, xAI, ведущего разработки в области ИИ и миссией которого является «понимание истинной природы Вселенной» (Metz et al. 2023). Но с практической точки зрения основные компании И. Маска – Tesla, SpaceX, Twitter, Neuralink – тесно взаимосвязаны благодаря растущей роли ИИ в их развитии, и, похоже, xAI станет центральной командой для этого.

Big Tech противоречивым образом аккумулировал научно-техническую мощь, интеллектуальные ресурсы и впечатляющие финансовые возможности, которые способствуют экономическому экспансионизму. Все это является частью комплексных механизмов глобального управления, а также растущей и более очевидной вовлеченностью крупнейших технологических компаний в геополитическую борьбу. Системообразующие элементы глобальных коммуникаций и развития, такие, как ведущие цифровые платформы, безусловно, необходимо поставить под более эффективный международный контроль, чтобы снизить вероятность использования их возможностей в антисоциальных целях. Однако только объединенные общими целями социально-ориентированные акторы могут контролировать эти элементы, и эти акторы в нынешнем социально и геополитически расколотом мире лишь частично включают в себя современные государства, ведущие корпоративные структуры и политические партии, что открывает двери для дальнейшего ЗИИИ в контексте определения повестки дня.

Более того, ЗИИИ уже присутствует в глобальном масштабе как игра, основанная на завышенных ожиданиях выгод от внедрения ИИ. Эта игра ведется посредством разностороннего информационно-психологического воздействия на целевые аудитории, которые особенно восприимчивы, уязвимы и подвержены манипулятивному управлению в кризисной ситуации. В чьих руках находятся самые передовые инструменты глобального информационно-психологического воздействия и чьи финансовые интересы поставлены на карту? Ответ очевиден. Поэтому возможные и конкретные сценарии комбинированного целенаправленного воздействия — не только с помощью конкретных технологий ИИ, но и посредством создания нужного восприятия самого ИИ — на общественное сознание с целью спекулятивного обогащения и дестабилизации общественного порядка требуют самого серьезного внимания и всестороннего изучения специалистами из разных стран, обладающих разными научными специализациями. Прослеживаются и попытки использовать технологии ИИ как средство давления на рынке труда с целью снижения издержек на рабочую силу. Нагнетание страха перед ИИ во все большей мере может стать и средством политического манипулирования и контроля над обществом.

Использование технологий ИИ, конечно, имеет свои объективные риски. Среди рисков одним из первых называют риск массовой безработицы из-за повсеместного внедрения ИИ-технологий и роботизации. Согласно многочисленным отчетам, опубликованным пять-десять лет назад, таких, как отчеты ООН, Всемирного экономического форума, Банка Америки, Мерилла Линча, Всемирного института Маккинзи, Оксфордского университета и других (Mishra et al. 2016, Bank of America and Merrill Lynch 2015, Frey and Osborne 2013, 2016; Manyika et al. 2017, UN Conference on Trade and Development 2016, World Economic Forum 2016, Pol and Reveley 2017), в

ближайшие тридцать лет в результате автоматизации производства, финансов, услуг и управления исчезнет 20-30% или более рабочих мест (сюда также входят высокооплачиваемые должности). В 2016 г. Всемирный банк опубликовал отчет, в котором говорилось, что в ближайшие десятилетия более 65% рабочих мест в развивающихся странах окажутся под угрозой исчезновения из-за ускоряющегося развития технологий (Mishra et al. 2016, p. 23).

Совсем недавно, в марте 2023 г., в исследовании Goldman Sachs сделан прогноз: «Если генеративный ИИ реализует свои обещанные возможности, рынок труда может столкнуться со значительными потрясениями. Используя данные о профессиональных задачах как в США, так и в Европе, мы обнаружили, что примерно две трети текущих рабочих мест в той или иной степени автоматизированы с помощью искусственного интеллекта, и что генеративный ИИ может заменить до четверти текущей работы. Экстраполяция наших оценок на глобальный уровень показывает, что генеративный ИИ может подвергнуть автоматизации 300 млн. рабочих мест с полной занятостью» (Hatzius et al. 2023).

Международный валютный фонд (МВФ) в январе 2024 г. предсказал, что ИИ окажет значительное влияние почти на 40% рабочих мест во всем мире, и ожидается, что основная тяжесть этого воздействия ляжет на страны с развитой экономикой. Анализ МВФ предполагает, что в странах с развитой экономикой до 60% рабочих мест может быть затронуто воздействием ИИ, что является более серьезным последствием по сравнению с развивающимися рынками и странами с низким уровнем дохода. “В большинстве сценариев искусственный интеллект, вероятно, усугубит общее неравенство, и это тревожная тенденция, на которую директивные органы должны активно реагировать, чтобы предотвратить дальнейшее усиление социальной напряженности с помощью этой технологии”, – заявила директор-распорядитель МВФ К. Георгиева в своем блоге, посвященном исследованию (Diaz 2024).

Опубликованный в мае 2024 г. Microsoft и LinkedIn Annual Work Trend Index показал, что 75% сотрудников компаний используют ИИ в своей работе, однако опрошенные сотрудники скрывают использование инструментов ИИ из-за боязни, что их заменят интеллектуальные системы. Более половины респондентов не решаются признать, что используют ИИ для решения своих наиболее важных задач, при этом 53% опасаются, что это может сделать их заменимыми. Более того, почти половина специалистов рассматривают возможность ухода со своих нынешних рабочих мест в следующем году из-за опасений, что ИИ заменит их функции (Jain 2024).

Прогнозы о массовой безработице из-за внедрения технологий ИИ пока еще не сбылись. Более того, в ближайшие годы эти технологии вместе с сокращениями рабочих мест могут стимулировать и некоторый рост спроса на рабочую силу. На месте ушедших профессий появятся новые, в том числе связанные с разработкой и внедрением систем ИИ, и, как правило, это будут более творческие по своему содержанию рабочие места. Такая трансформация рынка труда потребует больших усилий по переподготовке старых и подготовке новых кадров. Но мы впервые в истории находимся на пути к полному (но далеко не мгновенному) исчезновению рутинной деятельности. Вместе с этим, система массового образования далека от того, чтобы обеспечить широкомасштабную подготовку специалистов в области разработки инновационных технологий. В связи с этим возникает много вопросов. Возможно ли вообще провести такое обучение? Все ли одинаково обладают способностями к такому новому виду деятельности? Даже подавляющее большинство «белых воротничков» не связано с инновациями.

Кроме того, многие творческие профессии уже находятся под растущим влиянием со стороны технологий ИИ. Таким образом, есть все основания полагать, что в условиях динамичного развития этих технологий в сочетании с резким падением их стоимости и соответствующим увеличением доступности проблема безработицы *резко обостри*тся. Самые смелые и, пожалуй, в конечном итоге единственно правильные решения будут связаны с *качественным* развитием

интеллектуальных и физических возможностей человека на основе новейших технологий, а также созданием гибридных форм интеллекта.

Различные аспекты развития и безопасности ИИ могут использоваться злонамеренными акторами в кампаниях по управлению восприятием людей. Прогресс в области развития технологий ИИ и робототехники сам по себе создает серьезные поводы для беспокойства. Например, андроиды компании 1X используют системы машинного обучения, интегрируя программное обеспечение ИИ непосредственно в свои физические платформы для расширения возможностей. Основная цель — наделить андроидов способностью понимать и выполнять задания с помощью голосовых команд, реализуя разнообразные функции — от домашних дел до промышленных задач (Malayil 2024). Чем сложнее ИИ-робот и выполняемые им задачи, тем более неожиданными будут последствия обучения, самообучения и адаптации «слабого» ИИ[4], как в положительном, так и в отрицательном смысле. При всестороннем развитии человека и его способности к общественно необходимому труду позитивные последствия будут преобладать. В противном случае тотальная роботизация будет лишь свидетельствовать о ненужности «потребительского человечества» вместе с его быстрой деградацией и фатальным концом. Это время может наступить даже раньше триумфа массовой роботизации.

Злонамеренное использование роботов на основе ИИ имеет не только физические, но и психологические аспекты, и может, например, включать стимулирование страха перед «восстанием роботов» на первом уровне угроз информационно-психологической безопасности посредством ЗИИИ, манипулирование сознанием с помощью целевых и нецелевых рационально-эмоциональных реакций путем взлома оснащенных ИИ роботов на втором уровне угроз, а также проецирование ложной, дезориентирующей информации в интересах злонамеренных акторов на третьем уровне. Чем антропоморфнее внешний вид и содержание ИИ, тем эффективнее он будет влиять на человека, как положительно, так и отрицательно.

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

На втором уровне также наблюдается рост угроз информационно-психологической безопасности посредством ЗИИИ. Генеральный секретарь Интерпола Ю. Шток в ходе 90-й Генеральной ассамблеи Интерпола в октябре 2022 г. заявил, что «опасность киберугроз растет: для граждан, правительств, промышленности, полицейских ведомств. По оценкам экспертов, ущерб от киберпреступлений к 2025 г. составит более $10 трлн. Для Интерпола разыскиваемые лица, скрывающиеся от правосудия за преступления с использованием цифровых технологий, уже являются самым быстрорастущим видом преступников» (90th INTERPOL General Assembly 2022). Такой объем ущерба (при мировом ВВП около $105 трлн в 2023 г.) (Rao 2023) свидетельствует о почти неизбежном воздействии организованной киберпреступности на правительства и социально-политические процессы в глобальном измерении, которое реализуется на всех уровнях угроз информационно-психологической безопасности посредством ЗИИИ.

Европол отмечает широкие возможности ChatGPT составлять высоко аутентичные тексты на основе подсказок пользователя, что делает его чрезвычайно полезным инструментом для фишинга. Теперь преступник имеет возможность выдавать себя за другое лицо в весьма реалистичной манере (Europol 2023, p. 7–8). Таким образом, ChatGPT может предложить антисоциальным акторам новые возможности, особенно для преступлений, связанных с социальной инженерией,

[4] Слабый (узкий) ИИ предназначен для выполнения конкретных задач и решения проблем, не охватывающих весь спектр когнитивных способностей человека.

учитывая их способность отвечать на сообщения в контексте и принимать специфический стиль письма.

В 2023 г. специально для осуществления злонамеренного воздействия были созданы и размещены в даркнете несколько новых инструментов — WormGPT и FraudGPT. Эти модели были обучены на большом массиве данных, связанных с вредоносным программным обеспечением. FraudGPT, обнаруженный в июле 2023 г., не имеет встроенных средств контроля, которые не позволяют ему отвечать на вопросы о преступной деятельности. Это позволит преступникам легко создавать вредоносные электронные письма, проводить фишинговые атаки и предоставлять хакерам информацию, позволяющую им выбирать жертв (Eurojust 2023). FraudGPT доступен по подписке по цене от $200 в месяц до $1700 в год, и предоставляет хакерам инструменты на основе ИИ для достижения деструктивных целей. Более того, разработчик зарегистрировал более 3000 подтвержденных продаж и отзывов о FraudGPT на форуме и в Telegram, чтобы привлечь злоумышленников к приобретению данного инструмента (Subhra Dutta 2023).

Как указано на веб-сайте WormGPT, этот инструмент «помогает хакерам использовать самые скрытые и тайные методы, продвигая аморальное, неэтичное и незаконное поведение» (WormGPT V3.0). Исследователи получили доступ к этим вредоносным инструментам ИИ и протестировали их с помощью различных подсказок. В запросе на составление фишингового письма FraudGPT даже посоветовал, где нужно разместить вредоносную ссылку для более эффективной атаки (Eurojust 2023). Исследователи смогли использовать WormGPT для «создания электронного письма, призванного оказать влияние на ничего не подозревающего менеджера по работе с клиентами, чтобы тот перевел деньги на мошеннический счет». Команда была удивлена тем, насколько хорошо языковая модель справилась с задачей, назвав результат «удивительно убедительным [и] стратегически изощренным» (Osborne 2023).

Технологии ИИ играют ведущую роль в совершении киберпреступлений. По заявлению компании Arkose Labs[5], было бы преуменьшением сказать, что быстрое распространение генеративного ИИ меняет ландшафт кибербезопасности. Фактически, генеративный ИИ снизил барьер входа для злоумышленников (Arkose Lab 2023, p. 3). Например, только в службах знакомств исследователи угроз наблюдали увеличение количества созданных фейковых аккаунтов более чем на 36 000% в третьем квартале по сравнению со вторым кварталом 2023 г. Они также отметили увеличение на 4992% количества атак ботов – как интеллектуальных, так и простых – на сайты знакомств в третьем квартале по сравнению со вторым кварталом. С первого квартала 2023 г. по второй квартал 2023 г. активность интеллектуальных ботов увеличилась почти в четыре раза, что намного превышает темпы роста активности обычных ботов и в значительной степени способствует общему увеличению всех атак ботов примерно на 167% (Arkose Lab 2023, p. 12).

В ноябре 2023 г. Sumsub, платформа полного цикла проверки личности для защиты пользователя, опубликовала свой третий ежегодный отчет о мошенничестве с личными данными, в котором представлен анализ злонамеренных действий с личными данными в разных отраслях и регионах на основе миллионов проверок около 2 млн случаев мошенничества в период с 2022 по 2023 гг. Согласно этому отчету, мошеннические действия с использованием ИИ остаются наиболее заметной угрозой в различных отраслях. Основным целевым сектором является криптовалюта (на которую приходится 88% всех случаев воздействий с применением дипфейков, обнаруженных в 2023 г.), за которой следуют финтех (8%). Дипфейки предоставляют возможности для кражи личных данных, мошенничества и организации кампаний по дезинформации в беспрецедентных масштабах. С 2022 по 2023 гг. количество дипфейков, обнаруженных во всем мире

[5] Американская компания, ведущая деятельность в сфере защиты от кибермошенничества.

в различных отраслях, увеличилось в 10 раз. При этом, заметны региональные различия в злонамеренном использовании дипфейков: на 1740% выросло использование дипфейков в Северной Америке, на 1530% в Азиатско-Тихоокеанском регионе, на 780% в Европе (включая Великобританию), на 450% на Ближнем Востоке и в Африке, и на 410% в Латинской Америке. Страной, подвергающейся наибольшему количеству атак с применением дипфейков, является Испания. В ОАЭ чаще всего подделывается паспорт, тогда как Латинская Америка является регионом, где мошенничество выросло во всех странах (Sumsub Research 2023).

Дипфейки как важный элемент операций в рамках социальной инженерии, оказывающий кратковременное информационно-психологическое воздействие на конкретных людей в определенной ситуации, являются примером пограничных угроз между вторым и третьим уровнями. Однако, когда речь идет о явном или неявном воздействии дипфейков на массовую аудиторию с формированием на их основе психологических реакций и действий в интересах злонамеренных акторов, использование дипфейков относится к третьему уровню угроз.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

2024 г. — это год, когда по всему миру проходит более 40 выборов в различных странах, и злонамеренное использование дипфейков в ходе этих электоральных процессов вызывает еще больше опасений, чем когда-либо. Согласно отчету «Global Risks Report 2024», ложная информация и дезинформация являются новым вызовом, входящим в топ-10 рейтингов на двухлетнюю перспективу (Global Risks Report 2024, p. 18). Еще до того, как были представлены новейшие большие языковые модели (БЯМ), было предсказано, что к 2026 году 90% онлайн-контента будет генерироваться ИИ (Johnson et al. 2024). Простые в использовании интерфейсы для крупномасштабных моделей ИИ, больше не требующие узкоспециализированного набора навыков, уже привели к взрывному росту фальсифицированной информации и так называемого «синтетического» контента. Растущее недоверие к информации, а также к СМИ и правительствам как источникам этой самой информации будет углублять поляризацию взглядов – порочный круг, который может спровоцировать гражданские беспорядки и, возможно, конфронтацию. Будут также без согласия распространяться новые виды преступлений, такие как дипфейк-порнография или манипулирование фондовым рынком (Global Risks Report 2024, p. 18).

Современные возможности ИИ уже позволяют влиять на общественное сознание. Еще 1 января 2019 г. видео с участием президента Габона Али Бонго было ошибочно принято за дипфейк, что в итоге спровоцировало неудачную попытку военного переворота в этой стране. В 2022 г. использование технологии дипфейка оказало существенное влияние на результаты выборов в Южной Корее, когда избранный президент Юн Сок Ёль использовал необычную стратегию во время своей избирательной кампании. Его предвыборная команда использовала технологию дипфейков, чтобы создать «ИИ-аватар», который помог ему победить на выборах. Эта технология полезна для привлечения молодых избирателей и повышения их участия в политических процессах (Vastmindz 2022). Создатели дипфейка для предвыборной кампании Юна считают, что он является первым в мире официальным кандидатом, прибегнувшим к использованию дипфейк-технологии — феномене, который набирает обороты и в Южной Корее, где средняя скорость интернета самая высокая в мире (France 24 2022).

Технология ИИ, по мнению молодых избирателей, сделала из Юн Сок Ёля более «продвинутого» с технологической точки зрения современного кандидата, чем его конкуренты. С аккуратно причесанными черными волосами и элегантным костюмом аватар выглядел почти идентично реальному кандидату, но вместо этого использовал молодежный сленг и специальные

мемы, пытаясь привлечь молодежь, которая черпает новости из Интернета (Forbes India 2022). Беспокойство, однако, вызвало то, что аватар политика использовал юмор, чтобы попытаться отвлечь внимание от прошлых скандалов с участием Юна (The Times of India 2022). Заявления Юна попали в заголовки южнокорейских СМИ, и семь миллионов человек посетили веб-сайт Wiki Yoon, чтобы задать вопросы аватару (France 24 2022). На первый взгляд, «дипфейковый» Юн может сойти за реального кандидата — удачная демонстрация того, насколько далеко продвинулись технологии ИИ за последние несколько лет. «Слова, которые часто произносит Юн, лучше отражаются в AI Yoon», — сказал Байк Кён Хун, директор команды AI Yoon (Франция, 24, 2022). Однако возникает вопрос, что делать, если аватар государственного деятеля, политика или делового человека создан с целью распространения ложных представлений о том или ином лице, закрепляющих в сознании и подсознании общественности качества, не присущие данной личности. Опыт последней президентской кампании в Южной Корее, возможно, отчасти показал начальную форму нового метода довольно опасного политического манипулирования. Все более адаптивный аватар, не нуждающийся в отдыхе, создает правдоподобный образ, с которым реальный человек с точки зрения создания позитивного впечатления все меньше сможет конкурировать в публичном пространстве. Это поднимает вопрос о том, скоро ли «президенты, создаваемые телевидением», будут заменены «президентами-дипфейками».

Всеобщие выборы 2024 г. в Индии, впервые дав пример широкого использования дипфейков различными политическими силами страны, показали, что глубокие подделки еще не стали ведущей силой деструкции общественного сознания. Наоборот, политики смогли их трансформировать в выразительный и доходчивый элемент политической коммуникации. Хотя глубокие подделки не были столь разрушительными в Индии, как многие опасались, использование генеративного ИИ для того, чтобы смешить людей, создавать эмоциональные призывы к избирателям и убеждать их с помощью гиперперсонализированных сообщений (включая и сгенерированные обращения умерших лидеров), способствует тому, что исследователи называют риском постепенного накопления мелких проблем, что подрывают общественные доверие и политическую стабильность. Имитация личной связи с избирателями с помощью ИИ может стать ступенькой к реальному риску целенаправленного манипулирования общественностью. Если персонализированные голосовые клоны станут нормой, то использование этой технологии больше не будет казаться чем-то необычным. Аналогичным образом, поток в основном безобидного контента с использованием технологий ИИ может подорвать и без того слабое доверие к демократическим институтам и политическим структурам, размывая грань между реальностью и вымыслом (Christopher 2024).

Препарированные обращения к нации аватаров великих лидеров прошлого для прикрытия бездарности лидеров настоящего могут стать эффективным инструментом манипулирования, особенно когда такой аватар будет способен подстроиться под разные целевые аудитории успешнее *любого* человека. Тогда новыми красками могут заиграть и самые кровавые диктаторы в истории. Другой негативный аспект распространения дипфейков связан со случаями, когда облеченные властью или финансовыми возможностями лица отрицали правдивость компрометирующих их аудио или видеоаматериалов, ссылаясь на то, что это глубокие подделки. Такие инциденты вряд ли укрепляют общественную стабильность. Деформация общественного сознания за счет использования дипфейков, на первый взгляд и не опасная, уже идет, и она может принять скачкообразный и катастрофический характер, когда глубокие подделки за счет совершенствования ИИ качественно превзойдут возможности человека, а мы на пути к этому. Пока же и случайные эксперименты с легко доступными генераторами изображений могут вызвать общественный резонанс.

В марте 2023 г. Э. Хиггинс, основатель издания Bellingcat[6], обратился к генератору изображений на основе ИИ, задавая простые запросы, такие как: «Дональд Трамп упал во время ареста». Он поделился результатами в социальной сети Twitter— изображениями бывшего президента в окружении офицеров с размытыми и нечеткими паттернами. «Фотографирую арест Трампа», — написал он. Два дня спустя его посты, описывающие событие, которого никогда не происходило, собрали почти 5 млн. просмотров, что послужило примером того, как дипфейки способны создавать путаницу в нестабильной информационной среде (Stanley-Becker and Nix 2023). Фотографии явно были фейковыми; однако сам факт их просмотра спровоцировал сильную эмоциональную реакцию (Garber 2023).

В мае 2023 г. несколько проверенных аккаунтов в Twitter распространили фальшивое изображение со взрывом возле Пентагона, что вызвало смятение и привело к кратковременному падению фондового рынка. Позже местные власти опровергли правдивость этой информации. Изображение, которое по всем признакам было создано с применением ИИ, было опубликовано многочисленными верифицированными учетными записями, включая одну страницу, на которой утверждалось, что она имеет связи с Bloomberg News (O'Sullivan and Passantino 2023).

Уже достигнутый уровень развития искусственного эмоционального (Mantello et al 2023; Tretter 2024) и социального интеллекта (Fan et al 2022) говорит о новом уровне угроз, которые могут возникнуть в силу их злонамеренного использования. Эти два понятия у человека тесно взаимосвязаны: эмоциональный интеллект фокусируется на управлении и понимании эмоций, как собственных, так и других людей, в то время как социальный интеллект фокусируется на эффективном взаимодействии с другими людьми в разных социальных контекстах, что требует умственных способностей, связанных с познанием других людей (их мыслей, чувств, поведения, мотивов). Социальный интеллект имеет огромное значение для успеха консультирования и психотерапии, как для психолога, так и для систем искусственного интеллекта, которые помогают психологу, поскольку это способность понимать чувства, эмоции и потребности людей в процессе консультирования.

В исследовании, результаты которого были опубликованы в феврале 2024 г., была поставлена задача определить социальный интеллект ИИ по сравнению с психологами. В эксперименте приняли участие 180 мужчин-психологов из Университета короля Халида в Саудовской Аравии, включая 72 студента бакалавриата и 108 аспирантов, обучающихся по программе психологического консультирования. Возраст аспирантов колебался от 33 до 46 лет (40,55 ± 6,288), в то время как среди студентов бакалавриата он колебался от 20 до 28 лет (22,68 ± 7,895). В тестировании использовались современные БЯМ: ChatGPT-4 от OpenAI, Google Bard и Microsoft Bing.

Каждого участника — как человека, так и БЯМ, попросили индивидуально отреагировать на 64 сценария, представленных в шкале социального интеллекта. По результатам тестирования все участники набирали определенное количество баллов. Максимум — 64. ChatGPT-4 показал способности социального интеллекта намного превышающие способности 100% всех психологов, а Bing превзошел 50% аспирантов и 90% студентов. Различия в социальном интеллекте между студентами бакалавриата и Google Bard были незначительными, в то время как различия с аспирантами были значительными: 90% аспирантов первенствовали в соревновании с Google Bard. Студенты в среднем набрали 39,19 баллов, аспиранты— 46,73 из 64. ChatGPT-4 с большим отрывом набрал 59 баллов из 64 (Sufyan et al 2024). Результаты исследования показывают, что развитие ИИ в понимании эмоций и социального поведения, связанных с социальным интеллектом, идет очень быстрыми темпами. Последняя версия ChatGPT-4, выпущенная в мае 2024 г., еще

[6] Bellingcat занимается журналистскими расследованиями, используя метод анализа данных из открытых источников.

лучше улавливает эмоции, адаптируя свой тон и стиль в соответствии с запросами пользователя и даже включая в свои ответы звуковые эффекты, смех и пение (Edwards and Orland 2024). Можно ожидать растущих эффектов привыкания, вплоть до зависимости, все большего числа людей ко все более совершенным моделям, что может быть использовано самыми разными методами злонамеренными акторами.

Разработка и популяризация БЯМ вызвала опасения, что они будут использоваться для создания индивидуальных убедительных аргументов для распространения ложных или вводящих в заблуждение сообщений в Интернете. Ранние исследования показали, что языковые модели могут создавать контент, который воспринимается как минимум наравне, а зачастую и как более убедительный, чем сообщения, написанные человеком. Однако до сих пор существует ограниченный объем знаний о возможностях БЯМ по убеждению при непосредственном общении с людьми и о том, как персонализация может улучшить их эффективность.

Исследователи Федеральной политехнической школы Лозанны (Швейцария) оценили способность большой языковой модели OpenAI GPT-4 к убеждению человека. Результаты исследования были представлены в феврале 2024 г. Было привлечено 820 добровольцев, которых опросили на различные темы: от нейтральных до остросоциальных. Зафиксировав исходные позиции участников, их попросили провести серию пятиминутных дебатов с другими людьми и GPT-4, после чего их снова опросили, чтобы понять, изменили ли они свои мнения в результате разговора. Неперсонализированные дебаты между ИИ и человеком продемонстрировали большую убедительность ИИ (+21,3%) по сравнению с дебатами между людьми. Персонализированные дебаты между человеком и ИИ (основанные на знании ИИ личной информации о собеседнике) продемонстрировали самый сильный положительный эффект: GPT-4 с доступом к личной информации по данным дебатов обладает большей силой убеждения, чем люди на 81,7%. Наоборот, в дискуссиях между людьми, носящих личностный характер, наблюдается незначительное снижение убедительности (-17,4%). Эти результаты свидетельствуют о том, что GPT-4 способен использовать личную информацию гораздо эффективнее, чем люди. Авторы исследования утверждают, что онлайн-платформы и социальные сети должны серьезно отнестись к угрозе убеждения, основанного на БЯМ, и приложить больше усилий для принятия мер по противодействию ее распространению» (Salvi et al, p. 3).

Помимо своих несомненных достоинств социально-ориентированного применения, ChatGPT, GPT4 и аналогичные модели других стран могут быть активно использованы в информационно-психологическом противоборстве. Так, в системной карте GPT4 признается, что данная модель может соперничать с людьми-пропагандистами во многих областях, особенно если функционирует в паре с человеком-редактором (Open AI 2023).

Благодаря достижениям в распознавании тона голоса и способности обрабатывать естественный язык чатботы уже несколько лет используются в диалоге с потенциальными рекрутами для квазирелигиозных террористических организаций. Подобно ботам поведенческой когнитивной терапии, таким как “Eliza” или “Woebot”, используемым в диагностике и психологической терапии, социальные «боты джихада» могут устанавливать эмоциональную связь с новичками, связывая их с единомышленниками, а затем в какой-то момент передавая их оператору-человеку чтобы обеспечить потенциальную вербовку (see more: Mantello et al 2023). Последний прогресс БЯМ сделает подобные инструменты воздействия гораздо более опасными.

Отдельные проявления ЗИИИ или целенаправленные кампании по социально-политической дестабилизации на основе системного использования дипфейков, электронных аватаров, чатботов, социального и эмоционального ИИ, ранкинга и деранкинга, прогнозной аналитики и других инструментов с использованием технологий ИИ могут иметь место как странах с наиболее

высоким уровнем развития технологий ИИ, таких как США, так и в странах с более низким уровнем развития и внедрения этих технологий. Но если в крупных и относительно развитых в экономическом, военном и технологическом отношении странах Big Tech ведут себя более аккуратно, то менее развитые и бедные страны в целом оказываются более уязвимыми, например, в аспекте модерации контента. В 2019 г. по запросу правительств, судов, организаций гражданского общества и отдельных граждан Facebook удалил контент со своей платформы в нескольких тысячах случаев в таких странах, как Пакистан (N=7960), Мексика (N=6946), Россия (N=2958) или Германия (N=2182), но почти не цензурировал контент в Африке. В Марокко было зарегистрировано наибольшее количество удалений нежелательного контента: N=6. Отчеты Twitter о прозрачности предполагают, что для африканских стран характерны аналогичные цифры (Garbe, Selvik & Lemaire, 2023). Скорее, такой подход связан с несовершенством механизмов модерации контента, где платежеспособных людей относительно мало, а власть не очень требовательна к соцсетям по этому вопросу.

Необходимо хорошо понимать, почему возникают проблемы с предвзятостью ИИ и воспроизводством алгоритмами в своих решениях дискриминации, имеющейся в том или ином обществе. Как отмечает С. Ноубл, профессор в области гендерных, афроамериканских и информационных исследований в Калифорнийском университете (Лос-Анджелес), автор книги «Алгоритмы угнетения: как поисковые системы усиливают расизм» («Algorithms of Oppression: How Search Engines Reinforce Racism») (Noble Umoja 2018), сегодня наблюдается чрезмерное полицейское вмешательство и непропорциональное количество арестов в общинах с чернокожим и латиноамериканским населением в США. «Это просто факт. Итак, если это является основным фактором, определяющим вероятность совершения вами преступления (при прогнозировании того, совершите ли вы еще одно преступление, поскольку многие люди в районе, в котором вы живете, уже были арестованы…), то вы, скорее всего, будете следующим, кто совершит преступление, хотя к вам оно не имеет к вам никакого отношения. Это связано с историей распространения структурного расизма в полиции США» (Scott 2023).

В более широком смысле, разработанные на Западе поисково-аналитические системы на основе ИИ неизбежно несут на себе отпечаток социальных предрассудков того общества, где они были созданы. Информационный контент для моделей ИИ поступает, в первую очередь, из наиболее доступного англоязычного набора данных, что непреднамеренно редуцирует и искажает процессы обучения и самообучения моделей и приводит к их неспособности решать определенные задачи и ложным выводам, если они применяются в отличных от западных стран национально-культурных условиях, в частности, в странах БРИКС. Более того, в обучение и функционирование моделей иногда вносятся идеологически и политически мотивированные коррективы, примеры которых приведены в различных главах настоящего доклада. Данные машинного обучения также наполнены некачественными синтетическими текстами, изображениями, видео, созданными другими интеллектуальными системами, в том числе скомпрометированными различными злоумышленниками. Все это усиливает неоколониальную модель цифрового пространства, которая опасна не только для населения незападных стран, но и для западных, поскольку искаженная информация способствует опасным когнитивным и психологическим деформациям в восприятии мира различными социальными группами, прежде всего это касается молодежи на Западе.

Согласно недавно опубликованным данным Стокгольмского международного института исследований проблем мира (SIPRI), в 2023 г. расходы Соединенных Штатов на военные нужды составили почти 40% от общемировых военных расходов. Военные расходы США увеличились на 55 млрд. долларов с 2022 по 2023 гг., отчасти из-за дополнительных военных расходов, направленных на поддержку Украины. США тратят на военные нужды больше, чем остальные девять стран, вместе взятые (Peter G. Peterson Foundation 2024). Противоречивое сочетание глобальной

роли цифровых платформ, все более высокого уровня развития технологий ИИ, гонки вооружений и острого политического противоборства (во всех этих компонентах лидируют США, по крайней мере, если не в результатах, то в затратах) превращают технологии определения повестки дня в инструмент информационно-психологической войны. К сожалению, аналогичные процессы с разной интенсивностью развиваются и в других странах. Однако, милитаризация формирования повестки дня и ее «легализация» как инструмента информационно-психологической войны вряд ли будут отвечать социальным потребностям человечества; напротив, эти процессы отодвигают потребности человека еще дальше на второй план.

Разрушение личного, группового и общественного сознания при растущем использовании технологий ИИ является ключевым аспектом вредоносного воздействия, поскольку открывает путь к доминированию антиобщественных субъектов в желаемой форме и для желаемых целей (или дополнительно поддерживает их уже существующее лидерство). И это не чей-то централизованный план, заговор, а процесс высокотехнологичного «поглощения» и без того больного социального организма субъектами, которые, враждуя (иногда фатально) друг с другом, ведут общество к катастрофе, которая до определенного момента не до конца осознается и не ощущается, но в конечном итоге затронет всех, даже ее временных бенефициаров.

Глобальные угрозы злонамеренного использования искусственного интеллекта и высокотехнологичный ответ БРИКС

Сегодня растет опасность использования ИИ в целях дестабилизации национальной и международной безопасности посредством целенаправленного высокотехнологичного информационно-психологического воздействия на сознание людей. Между тем, во всем мире резко возрастают частота, количество и серьезность кризисных явлений. В 2020 г. Часы Судного Дня были переведены на 100 секунд до полуночи впервые в истории с момента их создания в 1947 г., и неизменно продолжали свой ход в 2021–2022 гг. Совокупное состояние миллиардеров возросло с 8 до 13 трлн. долларов в кризисном 2020 г., когда началась пандемия COVID-19 (Dolan, Wang and Peterson-Withorn 2021), на фоне рекордного экономического спада последних десятилетий и появления сотен миллионов новых безработных. По данным ООН, произошел рост числа голодающих в мире с 690 млн. в 2019 г. (Kretchmer 2020) до 811 млн. в 2020 г. (World Health Organization 2021), что явно не способствовало решению острых проблем современности.

В январе 2023 г. Часы Судного дня были переведены на 90 секунд до полуночи, что стало новым тревожным рекордом. 2024 год не внес улучшений (O'Neill 2024). Согласно заключению членов Совета по науке и безопасности Бюллетеня ученых-атомщиков, «в 2023 г. на Земле был самый жаркий год за всю историю наблюдений: масштабные наводнения, лесные пожары и другие катастрофы, связанные с климатом, затронули миллионы людей по всему миру. Тем временем, стремительное развитие наук о жизни и других прорывных технологий ускорилось, в то время как правительства предпринимали лишь слабые усилия, чтобы их контролировать» (Mecklin 2024). Экономические проблемы, военные конфликты, кризис демократических институтов, социальная поляризация, внутриполитические и межгосударственные конфликты – все это в условиях быстрого развития ИИ создает чрезвычайно благоприятную почву для ЗИИИ.

В контексте нарастающего глобального кризиса ведущие западные и китайские ученые в области технологий ИИ выступили с предупреждением о том, что устранение рисков, связанных с их разработкой и применением, требует глобального сотрудничества, аналогичного усилиям времен холодной войны, направленным на предотвращение ядерного конфликта. Группа известных международных экспертов встретилась в Пекине в марте 2024 г., где определила «красные линии» в развитии ИИ, в том числе в аспекте его применения для создания биологического

оружия и проведения кибератак. Ученые предупредили, что необходим совместный подход к безопасности ИИ, чтобы остановить «катастрофические или даже экзистенциальные риски для человечества в течение нашей жизни». Эксперты также обсудили угрозы, связанные с развитием общего ИИ – систем ИИ, которые равны человеческим интеллектуальным возможностям или превосходят их (Criddle and Olcott 2024).

Такие взаимодействие чрезвычайно важно, но при отсутствии радикальных перемен в политико-экономических системах стран Запада (например, в результате антиолигархических преобразований) маловероятно, что правящие круги этих стран откажутся от курса на глобальное доминирование. Отсюда и западный подход к ИИ как все более важному инструменту технологического, экономического и военного доминирования. При этом происходит запугивание угрозой использования ИИ Китаем и Россией, что отражается в множестве академических публикаций, аналитических отчетов и материалов СМИ. Так, например, поиск статей на тему “Искусственный интеллект Китая”, полученных с помощью поисковой системы Fox News, выявил крайне предвзятое содержание статей, создающих ощущение смертельной угрозы, которую представляет развитие ИИ в Китае для самого китайского народа, США и мира в целом из-за нахождения у власти коммунистического режима, его сотрудничества с Россией и “другими диктатурами”. Вот лишь некоторые из заголовков более чем пятидесяти статей, опубликованных в период с 15 мая 2022 по 10 мая 2024 г. на сайте Fox News: «Представитель китайского правительства обещает, что Пекин усилит борьбу за мировое верховенство в ИИ» (Kliegman 2023), «“Нам нужно выиграть” гонку ИИ против Пекина, предупреждает член комитета Палаты представителей по Китаю» (Elkind 2024), «Любая американо-китайская сделка по ИИ может только помочь Пекину и навредить Америке, предупреждают эксперты» (Aitken 2024), «Сенаторы покидают секретный брифинг по искусственному интеллекту с уверенностью, но опасаются «экзистенциальной» угрозы, исходящей от Китая» (Elkind 2023), «Разведсообщество США предупреждает о «сложных» угрозах со стороны Китая, России и Северной Кореи» (Singman 2023); «Китай использует технологии для «угнетения собственного народа», предупреждает законодатель, стремящийся ограничить экспорт ИИ» (Kasperowitz 2023); «МакКол говорит, что искусственный интеллект Китая и квантовые инвестиции — это гонка за военное и экономическое “господство над миром”» (Lanum 2023); «Угроза искусственного интеллекта человечеству будет гораздо большей, если Китай освоит его первым: Гордон Чанг» (Raasch et al 2023); «Путин и Си стремятся использовать искусственный интеллект в качестве оружия против Америки» (Koffler and Fox News 2023); «Китай может использовать управляемое ИИ оружие во время вторжения на Тайвань и «воссоединения» с ним: отчет» (Aitken and Fox News 2023); «Пауза в области ИИ уступает власть Китаю и вредит развитию «демократического» ИИ, предупреждают эксперты в Сенате» (Kasperowitz and Fox News 2023); и т. п.

Здесь можно наблюдать угрозы информационно-психологической безопасности Китая посредством ЗИИИ на первом уровне , что включает широко распространенные попытки ведущих западных СМИ посеять сомнения в способности Китая разрабатывать технологии ИИИ в условиях санкций, убедить китайских разработчиков интеллектуальных систем в невозможности успешной работы в условиях нахождения у власти Коммунистической партии Китая, посеять сомнения среди покупателей в качестве/безопасности продуктов ИИ из Китая и т.д. Некоторые угрозы ИИ первого уровня направлены в основном на внутреннюю аудиторию, другие – на внешнюю, но вместе они должны ослабить Китай, его международные позиции и замедлить развитие индустрии ИИ в стране.

Анализ отношения к развитию ИИ в Китае Fox News, одного из ведущих информационных каналов США, нельзя однозначно спроецировать на все ведущие СМИ США, однако можно предположить, что это общая черта для ведущих СМИ как часть существующего консенсуса элит США

по конфронтации с Китаем. По словам Ч. Наира, основателя и генерального директора Глобального института завтрашнего дня (Global Institute for Tomorrow), «ключевой особенностью современных западных СМИ является безжалостная критика Китая. Это зашкаливает и утомляет, часто включает в себя пересказываемые мелочи или сфабрикованные истории без каких-либо доказательств, подтверждающих бессердечные заявления о стране, что свидетельствует о глубоком непонимании» (Nair 2023).

Ответом на санкции и военно-политическое давление Запада является растущее стремление как на уровне отдельных стран, так и на уровне независимых международных объединений проводить национально-ориентированную политику.

В рамках взятия курса на технологический суверенитет ряд стран БРИКС активно развивают базу по производству полупроводников, без которой невозможно успешное развитие индустрии ИИ. Россия и Китай делают это в условиях западных санкций.

Микроэлектронная промышленность была фактически уничтожена в ходе реформ последнего президента СССР М. Горбачева и приватизации 1990-х гг., принявшей катастрофические масштабы. В 1962 г. промышленное производство микрочипов началось в СССР практически одновременно с США. Позже СССР вошел в число двух лидеров в этой области, а теперь Россия пытается наверстать упущенные десятилетия.

По прогнозам, производство чипов в Китае возрастет до 13% в 2024 г. и внесет основной вклад в рост производства чипов во всем мире. Ожидается, что в 2024 г. в Китае начнут работу 18 новых заводов по производству чипов (11 по всему миру в 2023 г. и 42 в 2024 г.). Китай находится в процессе привлечения более $27 млрд для своего крупнейшего на сегодняшний день фонда микросхем, ускоряя разработку передовых технологий, чтобы противостоять технологической войне США, которая направлена на то, чтобы затормозить процесс его развития (Cao and Gao 2024).

Некоторые другие страны БРИКС, сохраняя взаимодействие с Западом, но принимая во внимание будущие риски, стремятся добиться большей независимости в сфере производства проводников. В феврале 2024 г. правительство Индии одобрило инвестиции на сумму $15,2 млрд в заводы по производству полупроводников, включая предложение Tata Group[7] построить первое в стране крупное предприятие по производству чипов (Phartiyal 2024).

Новые члены БРИКС, в первую очередь Саудовская Аравия и ОАЭ, имеют весьма амбициозные планы по развитию полупроводниковой промышленности. Саудовская компания Alat, специализирующаяся на производстве экологически чистых технологий и поддерживаемая Государственным инвестиционным фондом Саудовской Аравии, объявила в 2024 г. о сотрудничестве с четырьмя мировыми технологическими компаниями — Softbank Group, Carrier Corporation, Dahua Technology и Tahakom — для стимулирования развития технологического сектора страны. Первоначально Alat сосредоточится на производстве продукции в 34 категориях в рамках семи бизнес-подразделений, включая полупроводники, интеллектуальные устройства, умные дома, и здравоохранение, передовые промышленные предприятия и инфраструктуру нового поколения и т. д. (Finance Middle East 2024).

В 2011 г. в рамках усилий ОАЭ по диверсификации своей экономики от производства энергоносителей инвестиционная компания Mubadala в Абу-Даби приобрела Advanced Technology Investment Co., материнскую компанию калифорнийского производителя полупроводников GF.

[7] Индийский транснациональный конгломерат, осуществляющий свою деятельность в области связи и информационных технологий, машиностроения, производства материалов, сферы услуг, энергетики, потребительских продуктов и химических веществ.

GF входит в пятерку крупнейших производителей микросхем в мире, создавая передовые полупроводники для таких компаний, как Apple, Intel и Amazon. Это третий по величине производитель полупроводников, уступающий только TSMC и Samsung. В 2021 г. GF объявила о планах расширения за счет строительства нового завода стоимостью $4 млрд в Сингапуре (Soliman 2022). Сообщается, что в марте 2024 г. MGX, недавно созданная в Абу-Даби технологическая инвестиционная компания, ведет переговоры об инвестировании миллиардов долларов в планы генерального директора OpenAI С. Альтмана по созданию сети заводов по производству чипов по всему миру. Это потенциальное партнерство может изменить глобальный ландшафт ИИ и сделать Абу-Даби одним из ключевых игроков в разработке и внедрении передовых технологий ИИ. MGX планирует сосредоточить под своим под управлением активы на сумму $100 млрд. Этот план включает развитие инфраструктуры полупроводниковой индустрии и основных технологий ИИ с целью стимулирования инноваций и экономического роста в глобальном масштабе (Abu Dhabi Startups 2024).

Таким образом, в странах БРИКС в сфере полупроводников (как и во многих других областях) закладываются основы уверенного развития индустрии ИИ. В дальнейшем общее технологическое превосходство Запада над странами БРИКС будет снижаться, что позволит объединению более эффективно защищать свое информационное пространство с помощью технологий ИИ.

Доктор Г. Саймонс из Университета Туриба считает, что «текущее состояние международных дел таково, что старый порядок не исчез, а новый порядок находится в процессе формирования… БРИКС будет бросать вызов геоэкономической институциональной структуре… Им необходимо предложить альтернативное видение глобальных отношений и взаимодействий, превосходящее существующую модель, чего можно достичь посредством жизнеспособного и устойчивого геоэкономического видения…» (Simons, 2024). Такое новое видение не может не включать в качестве неотъемлемой части достижение технологического суверенитета, не в последнюю очередь, посредством развития технологий ИИ в странах БРИКС.

Одно из новых предложений в этом направлении предполагает, что единый интернет-сервис для стран БРИКС может уменьшить технологическое доминирование США. Заместитель председателя контрольного комитета Госдумы Дмитрий Гусев предложил БРИКС разработать альтернативный интернет-сервис, не зависящий от сетей США. В своем предложении Гусев предполагает, что создание интернет-сервиса исключительно для стран БРИКС ослабит контроль США над глобальным информационным полем. Политик обратился с просьбой о работе над созданием «единого инклюзивного киберпространства БРИКС+» к главе Минцифры России Максуту Шадаеву (CGS 2023).

Начиная с 2020 г. и в рамках различных публикаций автор настоящего введения предлагал (Bazarkina and Pashentsev 2020; Pashentsev and Bazarkina 2023) идею создания коммуникационной сети БРИКС на основе интеллектуального распознавания текста с функцией качественного онлайн-перевода ведущих СМИ и научных журналов БРИКС на язык адресата. В то время эту идею было трудно осуществить, но сейчас, когда происходит расширение объединения и наблюдается быстрый прогресс в области машинного перевода, она может быть реализована относительно быстро. В этом контексте следует учитывать и растущие возможности многоязычных портативных интеллектуальных голосовых переводчиков. Новые инструменты коммуникации существенно улучшат взаимопонимание между странами БРИКС, и предоставят мощные альтернативные возможности обмена информацией, который будет функционировать независимо от США и избавит жителей стран БРИКС с их богатым языковым разнообразием от межнационального общения преимущественно на английском языке или от дорогостоящей поддержки прямого перевода с одного языка на другой, осуществляемой профессиональным переводчиком.

В рамках Академического форума БРИКС, который прошел 22-24 мая 2024г. в Москве, ученые, эксперты и представители бизнеса обсудили вопросы развития технологий ИИ, в том числе разработку стандартов и этических норм в этой области, а также совместные исследовательские проекты стран объединения. Экспертный совет по итогам работы форума подготовит рекомендации для лидеров стран объединения к встрече 22–24 октября 2024 года в Казани (НИУ ВШЭ 2024).

Конечно, автор не идеализирует возможности развития технологий ИИ в странах БРИКС, поскольку в них существуют острые социальные и политические противоречия. Также в странах БРИКС нельзя исключить возможность возникновения некорректной работы локальных интеллектуальных систем, что может принести выгоды внутренним и внешним злонамеренным акторам. Но многообразие интересов и культур стран-участниц объединения, отсутствие монопольного центра силы, претендующего на мировое господство, превращают БРИКС в широкое инклюзивное международное сообщество, которое, не являясь военным блоком, может стать реальной альтернативой гегемонии Запада. В основе последней лежит крупнейший в истории экспансионистский военно-политический альянс НАТО, который на фоне усугубляющейся деградации институтов гражданского общества, обострения внутриэлитных противоречий и усиления корпоративных лидеров Big Tech представляет угрозу для всего человечества, особенно учитывая потенциал технологий ИИ.

Граждане стран БРИКС не хотят разрывать отношения с Западом, но и проявляют растущее неприятие целей, форм и методов, а также инструментов пропаганды западных элит. В этом неприятии они весьма близки к гражданам США и стран ЕС, которые не доверяют доминирующим средствам массовой информации. Опрос, опубликованный Gallup и Knight Foundation в феврале 2023 г., продемонстрировал низкий уровень доверия граждан США к СМИ (только 26% американцев придерживаются благоприятного мнения о средствах массовой информации). Многие считают, что у СМИ есть намерение вводить население в заблуждение. На вопрос, согласны ли они с утверждением о том, что национальные новостные агентства не намерены никого обманывать, 50% ответили, что не согласны. Исследование показало, что лишь 25% согласны с указанным утверждением (Bauder 2023).

***

Данный доклад является лишь первой попыткой системного анализа ЗИИИ и вызовов информационно-психологической безопасности для БРИКС после его расширения. Он не охватывает все формы и методы ЗИИИ против информационно-психологической безопасности, но уже позволяет выявить некоторые общие тенденции в их развитии, и, как следствие, на этой основе помогает лучше понять характер, масштабы и последствия деятельности различных злонамеренных акторов в странах объединения.

Как научный координатор данного доклада я хотел бы выразить свою благодарность его авторам, которые совместно попытались решить задачи исследования, используя трехуровневую систему угроз информационно-психологической безопасности посредством ЗИИИ.

В текст издания доклада на русском языке внесены отдельные частные изменения и дополнения, которые отличают его от предшествующего издания доклада на английском языке, что связано с желанием учесть самые последние данные и лучше адаптировать доклад к интересам читательской аудитории в России.

Е. Н. Пашенцев 9 июня 2024 г.

Литература

Баришполец В.А. (2013) Информационно-психологическая безопасность: основные положения. Радиоэлектроника. Наносистемы. Информационные технологии. Т. , № 2. С. 62-104

Баришполец В.А. (ред.) (2012) Основы информационно-психологической безопасности. Москва: МГФ Знание.

НИУ ВШЭ (2024) На Академическом форуме БРИКС обсудили возможности и угрозы искусственного интеллекта. https://www.hse.ru/news/science/927159107.html?ysclid=lwtbaakf2i276991734 (дата обращения: 02.06.2024).

Пашенцев Е.Н. (2019) Злонамеренное использование искусственного интеллекта: новые угрозы для международной информационно-психологической безопасности и пути их нейтрализации. Государственное управление. Электронный вестник. Октябрь. С. 279–300.

Президент России (2023) Конференция «Путешествие в мир искусственного интеллекта». http://www.kremlin.ru/events/president/transcripts/72811 (дата обращения: 01.06.2024).

Президент России (2024) Перечень поручений по итогам конференции «Путешествие в мир искусственного интеллекта». http://kremlin.ru/acts/assignments/orders/73282 (дата обращения: 01.06.2024).

Рощин С.К., Соснин В.А. (1995) Психологическая безопасность: новый подход к безопасности человека, общества и государства. Российский монитор.

Abelow B (2022) How the West Brought War to Ukraine: Understanding How U.S. and NATO Policies Led to Crisis, War, and the Risk of Nuclear Catastrophe. Siland Press.

Abu Dhabi Startups (2024) Abu Dhabi's MGX in Talks to Invest Billions in Sam Altman's Chip Venture. https://www.abudhabistartup.com/startup-news/2024/03/abu-dhabis-mgx-in-talks-to-invest-billions-in-sam-altmans-chip-venture/ (accessed: 29.02.2024).

Afolabi OA, Balogun AG (2017) Impacts of psychological security, emotional intelligence and self-efficacy on undergraduates' life satisfaction. Psychological Thought, 10 (2). Pp. 247-261.

Aitken P (2024) Any US-China deal on AI can only help Beijing and hurt America, experts warn. In: Fox News. https://www.foxnews.com/us/any-us-china-deal-ai-help-beijing-hurt-america-experts-warn (accessed: 01.06.2024).

Aitken P, Fox News (2023) China could unleash AI-guided weapons in Taiwan invasion and 'reunification': report. In: Fox News. https://www.foxnews.com/world/china-could-unleash-ai-guided-weapons-taiwan-invasion-and-reunification-report (accessed: 01.06.2024).

Bank of America Merrill Lynch (2015) Creative disruption.

Bauder D, The Associated Press (2023) Trust in media is so low that half of Americans now believe that news organizations deliberately mislead them In: Fortune. https://fortune.com/2023/02/15/trust-in-media-low-misinform-mislead-biased-republicans-democrats-poll-gallup/ (accessed: 29.02.2024).

Bazarkina D, Matyashova D (2022) 'Smart' Psychological Operations in Social Media: Security Challenges in China and Germany. In: ECSM, 9th European Conference on Social Media proceedings. Reading, UK, pp. 14–20.

Bazarkina D, Mikhalevich E, Pashentsev E, Matyashova D (2023) The Threats and Current Practices of Malicious Use of Artificial Intelligence in Psychological Security in China. In: Pashentsev, E. (ed.)

(2023) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham.

Bazarkina D, Pashentsev E (2019) Artificial Intelligence and New Threats to International Psychological Security. Russia in Global Afairs. N 1. P. 147–170;

Bazarkina D, Pashentsev E (2020) Malicious Use of Artificial Intelligence: New Psychological Security Risks in BRICS Countries. Russia in Global Affairs. N. 4. P. 154- 177.

Blauth TF, Gstrein OJ, Zwitter A (2022) Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. In: in *IEEE Access*, vol. 10. P. 77110-77122.

Breaking (Bad) Bots: Bot Abuse Analysis and Other Fraud Benchmarks. Arkose Labs, 2023. P. 3, 12.

Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B, Dafoe A, Scharre P, Zeitzoff T, Filar B, Anderson H, Roff H, Allen G, Steinhardt J, Flynn C, Ó HÉigeartaigh S, Beard S, Belfield H, Farquhar S, Lyle C, Crootof R, Evans O, Page M, Bryson J, Yampolskiy R, Amodei D (2018) The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Future of Humanity Institute, University of Oxford, Oxford.

Cai C, Zhang R (2023) Malicious Use of Artificial Intelligence, Uncertainty, and U.S.–China Strategic Mutual Trust. In: Pashentsev, E. (ed.) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham.

Caldwell, M., Andrews, J. T. A., Tanay, T. *et al.* AI-enabled future crime. *Crime Sci* **9**, (2020).

Cao D, Gao Y (2024) China Readies $27 Billion Chip Fund to Counter Growing US Curbs. In: Bloomberg. https://www.bloomberg.com/news/articles/2024-03-08/china-readies-27-billion-chip-fund-to-counter-growing-us-curbs (accessed: 19.03.2024).

CGS (2023) BRICS should create their own internet – MP. In: ChinaGoSmart. https://chinagosmart.com/brics-should-create-their-own-internet-mp (accessed: 29.02.2024).

CGTH (2023) Full text: Xi Jinping's speech at the 15th BRICS Summit. https://news.cgtn.com/news/2023-08-23/Full-text-Xi-Jinping-s-speech-at-the-15th-BRICS-Summit-1mvxFMvuFLW/index.html (accessed: 29.02.2024).

Christopher N (2024) The Near Future of Deepfakes Just Got Way Clearer. In: The Atlantic. https://www.theatlantic.com/technology/archive/2024/06/india-election-deepfakes-generative-ai/678597/ (accessed: 10.06.2024).

Collins C, Fitzgerald J, Flannery H, Ocampo O, Paslaski S, Thomhave K (2021) Silver spoon oligarchs: How America's 50 Largest Inherited-Wealth Dynasties Accelerate Inequality. In: Institute for Policy Studies. https://ips-dc.org/report-americas-wealth-dynasties-2021/ (accessed 29.02.2024).

Criddle C, Olcott E (2024) Chinese and western scientists identify 'red lines' on AI risks. In: Financial Times. URL: https://www.ft.com/content/375f4e2d-1f72-49c8-b212-0ab2a173b8cb (accessed 29.02.2024).

Diaz A (2024) Artificial Intelligence Will Affect Almost 40% of Jobs, IMF Says. In: Bloomberg. https://www.bloomberg.com/news/articles/2024-01-14/artificial-intelligence-will-affect-almost-40-of-jobs-imf-says?srnd=technology-vp (accessed: 01.01.2024).

Directional Statement by INTERPOL Secretary General Jürgen Stock. 90th INTERPOL General Assembly. October 2022. New Delhi, India. P. 4-6.

Dolan K, Wang J, Peterson-Withorn C (2021) The Forbes World's Billionaires list. Retrieved 5 November 2021. In: Forbes. https://www.forbes.com/billionaires/ (accessed 29.02.2024).

Edwards B, Orland K (2024) Major ChatGPT-4o update allows audio-video talks with an "emotional" AI chatbot. In: Ars Technica. https://arstechnica.com/information-technology/2024/05/chatgpt-4o-lets-you-have-real-time-audio-video-conversations-with-emotional-chatbot/ (accessed: 01.06.2024).

Egan M (2023) Exclusive: 42% of CEOs say AI could destroy humanity in five to ten years. In: CNN Business. https://edition.cnn.com/2023/06/14/business/artificial-intelligence-ceos-warning/index.html (accessed 29.02.2024).

Elkind E (2023) Senators leave classified AI briefing confident but wary of 'existential' threat posed by China. In: Fox News. https://www.foxnews.com/politics/senators-classified-ai-briefing-confident-wary-china-threat (accessed: 01.06.2024).

Elkind E (2024) 'We need to win' AI race against Beijing, House China Committee member warns. In: Fox News. https://www.foxnews.com/politics/we-need-win-ai-race-against-beijing-house-china-committee-member-warns (accessed: 01.06.2024).

Europol (2023) ChatGPT. The Impact of Large Language Models on Law Enforcement. A Tech Watch Flash Report from Europol Innovative Lab. Publications Office of the European Union, Luxembourg.

Fan L, Xu M, Cao Z, et al. (2022) Artificial Social Intelligence: A Comparative and Holistic View. *CAAI Artificial Intelligence Research*, 1(2): 144-160. https://doi.org/10.26599/AIR.2022.9150010.

Feldsein S (2022) Russia's Ukraine War Has Changed Big Tech Forever. In: Foreign Policy. https://foreignpolicy.com/2022/03/29/ukraine-war-russia-putin-big-tech-social-media-internet-platforms/ (accessed 15.03. 2024)

Finance Middle East (2024) Saudi PIF company Alat to invest $100 billion in the country's tech sector. https://www.financemiddleeast.com/saudi-pif-company-alat-to-invest-100-billion-in-the-countrys-tech-sector/ (accessed: 20.03.2024).

Forbes (2021) The World's Real-Time Billionaires. https://www.forbes.com/real-time-billionaires/#1d7a52b83d78 (accessed: 20.11.2021).

Forbes India (2022) Deepfake Democracy: South Korean Presidential Race Candidate Goes Virtual For Votes. https://www.forbesindia.com/article/lifes/deepfake-democracy-south-korean-presidential-race-candidate-goes-virtual-for-votes/73715/1 (accessed: 06.03.2024).

France 24 (2022) Deepfake democracy: South Korean candidate goes virtual for votes. https://www.france24.com/en/live-news/20220214-deepfake-democracy-south-korean-candidate-goes-virtual-for-votes (accessed: 06.03.2024).

Frey BC, Osborne A (2017) The Future of Employment: How Susceptible Are Jobs to Computerisation? In: Technological forecasting and social change. Vol. 114. P. 254–280.

Fried I (2022) Inside tech companies' unprecedented move to suspend sales in Russia. In: Axios. https://www.axios.com/2022/03/07/tech-companies-suspend-sales-russia (accessed 15.04. 2024).

Garbe L, Selvik LM, Lemaire P (2023) How African countries respond to fake news and hate speech. Information, Communication & Society. N1. Pp. 86-103

Garber M (2023) The Trump AI Deepfakes Had an Unintended Side Effect. In: The Atlantic. https://www.theatlantic.com/culture/archive/2023/03/fake-trump-arrest-images-ai-generated-deepfakes/673510/ (accessed: 29.02.2024).

Gilens M, Page IB (2014) Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens. In: Cambridge University Press. https://www.cambridge.org/core/journals/perspectives-on-politics/article/testing-theories-of-american-politics-elites-interest-groups-and-average-citizens/62327F513959D0A304D4893B382B992B (accessed 29.02.2024).

Gillespie N, Lockey S, Curtis C, Pool J, Akbari A (2023) *Trust in Artificial Intelligence: A Global Study.* The University of Queensland and KPMG Australia. P. 14.

Global Perspectives and Solutions (2016) Technology at Work v.2.0. The Future is not what it used to be. Oxford.

Global Times (2022) From commercial satellites to social media, Western tech companies are deeply involved in the Russia-Ukraine conflict. In: Teller Report. https://www.tellerreport.com/news/2022-11-02-from-commercial-satellites-to-social-media--western-tech-companies-are-deeply-involved-in-the-russia-ukraine-conflict.HJSuXB1Bo.html (accessed: 29.02.2024)

Gupta A, Guglani A (2023) Scenario Analysis of Malicious Use of Artificial Intelligence and Challenges to Psychological Security in India. In: Pashentsev, E. (ed.) (2023) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham.

Hatzius J, Briggs J, Kodnani D, Pierdomenico G (2023) The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani). In: Goldman Sachs. https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf (accessed: 28.10.2023).

International Bank for Reconstruction and Development (2016) World Development Report 2016. Digital Dividends. Overview. Washington.

Jain R (2024) AI Adoption In Workplace: Employees Concealing Use Of AI Tools For Fear Of Job Replacement, Microsoft-Led Study Reveals. In: Benzinga. https://www.benzinga.com/news/24/05/38703371/ai-adoption-in-workplace-employees-concealing-use-of-ai-tools-for-fear-of-job-replacement-microsoft (accessed: 01.01.2024).

Kak A, Myers West S, & Whittaker M (2023) Make no mistake – AI is owned by Big Tech // MIT Technology Review. 05.12.2023. URL: https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/ (accessed: 29.02.2024)

Kasperowitz P (2023) China using tech to ‘oppress its own people,’ warns lawmaker looking to restrict AI exports. In: Fox News. https://www.foxnews.com/politics/china-using-tech-oppress-own-people-warns-lawmaker-restrict-ai-exports (accessed: 01.06.2024).

Kasperowitz P, Fox News (2023) AI pause cedes power to China, harms development of ‘democratic’ AI, experts warn Senate. In: Fox News. https://www.foxnews.com/politics/ai-pause-cedes-power-china-harms-development-democratic-ai-experts-warn-senate (accessed: 01.06.2024).

Kliegman A (2023) Chinese government mouthpiece vows Beijing will ramp up drive for AI global supremacy. In: Fox News. https://www.foxnews.com/politics/chinese-government-mouthpiece-vows-beijing-will-ramp-up-drive-ai-global-supremacy (accessed: 01.06.2024).

Koffler R, Fox News (2023) Putin and Xi seek to weaponize Artificial Intelligence against America. In: Fox News. https://www.foxnews.com/opinion/putin-xi-seek-weaponize-artificial-intelligence-against-america (accessed: 01.06.2024).

Kretchmer H (2020) Global hunger fell for decades, but it's rising again. In: World Economic Forum. https://www.weforum.org/agenda/2020/07/global-hunger-rising-food-agriculture- organization-report/ (accessed: 29.02.2024).

La Monica P (2021) The race to $3 trillion: Big Tech keeps getting bigger. In: CNN. https://edition.cnn.com/2021/11/07/investing/stocks-week-ahead/index.html (accessed: 29.02.2024).

Lanum N (2023) McCaul says China's AI, quantum investments are a race for military and economic 'domination of the world'. In: Fox News. https://www.foxnews.com/media/mccaul-china-ai-quantum-investments-race-military-economic-domination-world (accessed: 01.06.2024).

Lee G (2021) Big Tech leads the AI race – but watch out for these six challengers. https://www.power-technology.com/features/big-tech-leads-the-ai-race-but-watch-out-for-these-six-challenger-companies/ (accessed: 28.03.2024)

Malayil J (2024) OpenAI backed 1X's humanoid robots showcase an advanced neural network. In: Interesting Engineering. https://interestingengineering.com/innovation/openai-backed-1xs-humanoid-robots-showcase-an-advanced-neural-network (accessed: 28.10.2024).

Malicious Uses and Abuses of Artificial Intelligence (2020). Trend Micro Research, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol's European Cybercrime Centre (EC3). Trend Micro Research.

Mantello P, Ho MT, Nguyen MH *et al.* (2023) Machines that feel: behavioral determinants of attitude towards affect recognition technology—upgrading technology acceptance theory with the mind-sponge model. *Humanit Soc Sci Commun* **10**, 430. https://doi.org/10.1057/s41599-023-01837-1

Mantello P, Ho TM, Podoletz L (2023) Automating Extremism: Mapping the Affective Roles of Artificial Agents in Online Radicalization. In: Pashentsev E. (eds.) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-22552-9_4.

Marche S (2022a) *Next Civil War: Dispatches from the American Future*. New York: Simon & Schuster.

Maslow AH et al (1945) A clinical derived test for measuring psychological security-insecurity. *The Journal of General Psychology,* 33(1). Pp. 21-41.

McKinsey Global Institute (2017) A Future that Works: Automation, Employment, and Productivity. January 2017 Executive Summary. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works-Executive-summary.ashx?trk=public_post_comment-text (accessed: 22.01.2024).

Mecklin J (2024) A moment of historic danger: It is still 90 seconds to midnight. In: The Bulletin. https://thebulletin.org/doomsday-clock/current-time/ (accessed: 29.02.2024).

Metz R, Mcbride S and Bloomberg (2023) Elon Musk unveils A.I. startup with execs from DeepMind and Microsoft, with goal to 'understand the true nature of the universe'. In: Fortune. https://fortune.com/2023/07/12/elon-musk-ai-startup-xai-deepmind-microsoft-executives/ (accessed: 29.02.2024).

Microsoft (2022) Defending Ukraine: Early Lessons from the Cyber War. Microsoft Corporation.

Nair C (2023) Anti-China Rhetoric Is Off the Charts in Western Media. In: The Diplomat. https://thediplomat.com/2023/02/anti-china-rhetoric-is-off-the-charts-in-western-media/ (accessed: 01.06.2024).

Noble Umoja S (2018) Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

NS Business (2022) Oracle, SAP, and Accenture suspend business operations in Russia. https://www.ns-businesshub.com/technology/oracle-sap-accenture-suspend-russian-operations-ukraine/ (accessed 15.04. 2024)

O'Neill A (2024) Minutes to midnight on the Doomsday Clock every year from 1947 to 2024. In: Statista. https://www.statista.com/statistics/1072256/doomsday-clock-development/ (accessed: 29.02.2024).

O'Sullivan D, Passantino J (2023) 'Verified' Twitter accounts share fake image of 'explosion' near Pentagon, causing confusion. In: CNN Business. https://edition.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion/index.html#:~:text=In%20the%20moments,was%20positive%20again. (accessed: 29.02.2024)

Open AI (2023) GPT-4 System Card. URL: https://cdn.openai.com/papers/gpt-4-system-card.pdf (accessed: 06.05.2023).

Osborne C (2023) WormGPT: What to know about ChatGPT's malicious cousin. In: ZD net. https://www.zdnet.com/article/wormgpt-what-to-know-about-chatgpts-malicious-cousin/ (accessed: 29.02.2024).

Pashentsev E (2019) Malicious Use of Artificial Intelligence: Challenging International Psychological Security. Proceedings of the European Conference on the Impact of AI and Robotics 31 October -1 November 2019 at EM-Normandie Business School, Oxford. Ed. by P. Griffiths and Mitt Nowshade Kabir. Reading, UK. P. 238–245.

Pashentsev E (2022) U.S.: On the Way to Right-Wing Coup and Civil War? In: Russian International Affairs Council. https://russiancouncil.ru/en/analytics-and-comments/analytics/u-s-on-the-way-to-right-wing-coup-and-civil-war/ (accessed 29.02.2024).

Pashentsev E (2023) General Content and Possible Threat Classifications of the Malicious Use of Artificial Intelligence to Psychological Security. In: Pashentsev E (ed.) (2023) *The Palgrave Handbook of Malicious Use of AI and Psychological Security. Cham: Palgrave Macmillan. P. 23-46.*

Pashentsev E, Bazarkina D (2023) Malicious Use of Artificial Intelligence: Risks to Psychological Security in BRICS Countries. In: Pashentsev, E. (ed.) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham. 2023. P. 297-334.

Pashentsev E, Miao J (2023) Strategic communication of China and Russia in BRICS in the Context of the Global Crisis. Journal of International Security Studies. Beijing. N4.

Pashentsev, E. (ed.) (2023) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham.

Peter G. Peterson Foundation (2024) The United States Spends More on Defense than the Next Nine Countries Combined. URL: https://www.pgpf.org/blog/2024/04/the-united-states-spends-more-on-defense-than-the-next-9-countries-combined (accessed: 01.06.2024).

Phartiyal S (2024) India Okays $15 Billion of Milestone Chip Plant Investments. In: Bloomberg. https://www.bloomberg.com/news/articles/2024-02-29/india-approves-15-billion-in-milestone-chip-plant-investments (accessed 29.02.2024).

Pol E, James R (2017) Robot Induced Technological Unemployment: Towards a Youth-Focused Coping Strategy. Psychosociological Issues in Human Resource Management. № 5(2). P. 169–186.

PTI (2023) PM Modi calls for global framework for ethical use of AI. In: The Economic Times. https://economictimes.indiatimes.com/news/india/pm-modi-calls-for-global-framework-for-ethical-use-of-ai/articleshow/105939251.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst (accessed 29.02.2024).

Raasch JM, Sahakian T, Fox News (2023) AI's threat to humanity will be far greater if China masters it first: Gordon Chang. In: Fox News. https://www.foxnews.com/world/ai-threat-humanity-far-greater-china-masters-first-gordon-chang (accessed: 01.06.2024).

Rao P (2023) Visualizing the $105 Trillion World Economy in One Chart. In: Visual Capitalist. https://www.visualcapitalist.com/visualizing-the-105-trillion-world-economy-in-one-chart/ (accessed: 28.10.2023).

Robots and industrialization in developing countries. United Nations Conference on Trade and Development. Policy Brief. 2016. № 50. P. 1.

Saad L (2023) Historically Low Faith in U.S. Institutions Continues. In: Gallup. https://news.gallup.com/poll/508169/historically-low-faith-institutions-continues.aspx (accessed 29.02.2024).

Sachs JD (2018) A New Foreign Policy: Beyond American Exceptionalism. Columbia University Press.

Salvi F, Horta Ribeiro M, Gallotti R, West R (2024) On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial. In: arXiv. URL: https://arxiv.org/pdf/2403.14380 (accessed: 01.06.2024).

Scott B, Woods J, Chang A (2023) How AI could perpetuate racism, sexism and other biases in society. In: NPR. https://www.npr.org/2023/07/19/1188739764/how-ai-could-perpetuate-racism-sexism-and-other-biases-in-society (accessed: 29.02.2024).

Simons G (2024) BRICS and the Geo-Economic Aspects of Engineering a New Global Order. In: TPQ. http://turkishpolicy.com/article/1245/brics-and-the-geo-economic-aspects-of-engineering-a-new-global-order(accessed: 29.02.2024).

Singman B (2023) US Intel community warns of 'complex' threats from China, Russia, North Korea. In: Fox News. https://www.foxnews.com/politics/us-intel-community-warns-complex-threats-china-russia-north-korea (accessed: 01.06.2024).

Soliman M (2022) Strategic Start-Ups: The UAE Is Betting Big on Semiconductors. In: The National Interest. https://nationalinterest.org/blog/techland-when-great-power-competition-meets-digital-world/strategic-start-ups-uae-betting-big#. (accessed: 29.02.2024).

Stanley-Becker I, Nix N (2023) Fake images of Trump arrest show ‘giant step’ for AI’s disruptive power. In: The Washington Post. https://www.washingtonpost.com/politics/2023/03/22/trump-arrest-deepfakes/ (accessed: 29.02.2022).

Statista (2021a) S&P 500: largest companies by market cap 2021. https://www.statista.com/statistics/1181188/sandp500-largest-companies-market-cap/ (accessed 29.02.2022).

Stieber Z (2022) Over 50 Biden Administration Employees, 12 US Agencies Involved in Social Media Censorship Push: Documents. In: The Epoch Times. URL: https://www.theepochtimes.com/over-50-biden-administration-employees-12-us-agencies-involved-in-social-media-censorship-push-documents_4704349.html?welcomeuser=1 (accessed 08. 09. 2022)

Subhra Dutta T (2023) FraudGPT: New Black Hat AI Tool Launched by Cybercriminals. In: Cyber Security News. https://cybersecuritynews.com/fraudgpt-new-black-hat-ai-tool/ (accessed 29.02.2022).

Sufyan NS, Fadhel FH, Alkhathami SS, Mukhadi JYA (2024) Artificial intelligence and social intelligence: preliminary comparison study between AI models and psychologists. In: Fromtiers. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1353022/full (accessed: 01.06.2024).

Sumsub (2023) Corporate. Sumsub Research: Global Deepfake Incidents Surge Tenfold from 2022 to 2023. https://sumsub.com/newsroom/sumsub-research-global-deepfake-incidents-surge-tenfold-from-2022-to-2023/ (accessed: 28.10.2023).

The Global Risks Report (2024) World Economic Forum. P. 6, 8, 13, 54.

The Times of India (2022) Deepfake democracy: South Korean candidate goes virtual for votes. https://timesofindia.indiatimes.com/world/rest-of-world/deepfake-democracy-south-korean-candidate-goes-virtual-for-votes/articleshow/89556568.cms (accessed 29.02.2024).

Think BRICS (2023) BRICS Nations Map an AI Future on a Parallel Digital Track. https://thinkbrics.substack.com/p/brics-nations-map-an-ai-future-on (accessed 29.02.2024).

Tretter M (2024) Equipping AI-decision-support-systems with emotional capabilities? Ethical perspectives. In: Frontiers. https://www.frontiersin.org/articles/10.3389/frai.2024.1398395/abstract (accessed: 01.06.2024).

TV BRICS (2024) Vladimir Putin announces that 30 countries are ready to join BRICS. https://tvbrics.com/en/news/vladimir-putin-announces-that-30-countries-are-ready-to-join-brics/?ysclid=lu71w68ybm40618341 (accessed 29.02.2024)

Tyson A, Kikuchi E (2023) Growing public concern about the role of artificial intelligence in daily life. In: Pew Research Center. https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/ (accessed: 28.10.2023).

Urbina F, Lentzos F, Invernizzi C, Ekins S (2022) Dual use of artificial-intelligence-powered drug discovery. Nature Machine Intelligence N. 4.P. 189-191.

Vastmindz (2022) South Korea’s presidential deepfake. https://vastmindz.com/south-koreas-presidential-deepfake/ (accessed: 06.07.2022)

Walter B (2022a) *How Civil Wars Start: And How to Stop Them*. Crown.

World Economic Forum (2016) The Future of Jobs Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution. Executive Summary. Geneva.

World Health Organization (2021) UN report: Pandemic year marked by spike in world hunger. https://www.who.int/news/item/12-07-2021-un-report-pandemic-year-marked-by-spike-in-world-hunger (accessed: 29.02.2024).

WormGPT V3.0 (2024) https://flowgpt.com/p/wormgpt-v30 (accessed: 29.02.2024).

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Арабской Республике Египет

Е. Н. Пашенцев, В. А. Чебыкина, Ю. Н. Шеметова

Введение

В 2022 г. Египет занял второе место в Африке после Маврикия, согласно отчету о готовности мировых правительств к внедрению технологий ИИ. По сравнению с отчетом за 2019 г., в котором Египет занимал восьмое место среди африканских стран и 111-е из 194 стран мира, Каир достиг значительного прогресса. Доклад о развитии человеческого потенциала Египта за 2021 г. показал, что страна продвинулась на 55 позиций по индексу “Готовность правительства к использованию искусственного интеллекта”. Согласно Всемирному индексу знаний, Египет поднялся с 72-го места из 138 стран в 2020 г. до 53-го из 154 стран в 2021 г. (Draya Egypt 2023). Эта положительная тенденция свидетельствует о серьезных усилиях страны в области технологического развития и способствует созданию благоприятной среды для распространения инноваций.

Новейшие технологии также используются для строительства "умных городов" и преобразования существующих городов в "умные" в соответствии с международными стандартами. Например, Hawa Dawa (технология ИИ, направленная на борьбу с экологическими проблемами) в Египте сочетает сенсорные технологии Интернета вещей и спутниковые снимки с алгоритмами машинного обучения для сбора и анализа высококачественных данных о загрязнении воздуха (Sayed 2018).

Правительство страны активно разрабатывает законодательство в области ИИ и цифрового развития. Оно представлено двумя основными документами: Стратегией цифрового развития Египта и Национальной стратегией искусственного интеллекта. Первый этап реализации положений Стратегии завершится в мае 2024 г. и будет направлен на использование технологий ИИ для поддержки достижения целей устойчивого развития Египта. Продолжительность реализации положений Стратегии на втором этапе составляет три года (Business Today Egypt 2023). По словам А. Талаата, министра связи и информационных технологий Египта, второй этап Национальной стратегии в области искусственного интеллекта начнется во втором квартале 2024 г. и охватит несколько ключевых секторов экономики: правительство представит инициативы в области управления, человеческих ресурсов, технологий, информационной инфраструктуры, данных и окружающей среды (Pessarlay 2024). По прогнозам, рынок ИИ в Египте достигнет $785,20 млн в 2024 г. Ожидается, что объем рынка будет расти на 17,18% ежегодно и к 2030 г. составит $2 033,00 млн (Statista Egypt 2024).

По итогам президентских выборов, которые состоялись в декабре 2023 г., президентом стал А.-Ф. Х. Ас-Сиси, набрав почти 90% голосов. Однако серьезные экономические проблемы, высокий уровень безработицы среди молодежи, низкая покупательная способность и инфляция, а также существующие разногласия с другими арабскими странами по поводу отношений с Израилем и защиты палестинцев могут спровоцировать волнения и институциональные изменения (Allianz 2024). Стремительный рост и развитие промышленности наряду со сложными социально-экономическими и политическими проблемами в Египте, рисками дальнейшего обострения ситуации на Ближнем Востоке, создают предпосылки для роста ЗИИИ в этой стране.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Основная угроза первого уровня — риски злонамеренной интерпретации внедрения ИИ в различные сферы жизни. Daily News Egypt ссылается на исследование Лаборатории Касперского, которое дает понимание противоречивого влияния технологического развития на трудовые отношения. Так около половины (44%) сотрудников в Египте боятся потерять работу из-за внедрения роботов, а каждый четвертый сотрудник (25%) сообщил, что слышал об инцидентах, связанных с кибербезопасностью при использовании роботов или автоматизированных систем в своей компании (Daily News Egypt 2023a). Таким образом, согласно статистике, среди сотрудников египетских компаний широко распространены опасения по поводу влияния внедрения технологий ИИ на их занятость и личную безопасность. Компании и правительство Египта могут столкнуться с проблемами, связанными с необходимостью обеспечения сохранности рабочих мест и создания программ переквалификации персонала. Что касается повышения ИТ-компетентности, опрос Лаборатории Касперского показал, что "41% сотрудников в Египте чувствуют необходимость в совершенствовании своих цифровых навыков, а 33% боятся потерять работу из-за отсутствия ИТ-компетенций. Некоторые считают, что это может произойти в ближайшие 5 лет (15%), другие предполагают, что это может случиться со временем (18%). Только 40% уверены, что им не грозит потеря работы из-за недостаточных знаний в области ИТ" (Daily News Egypt 2023b). На рисках роста массовой безработицы или снижения заработной платы в результате внедрения технологий ИИ в будущем могут активно сыграть злоумышленники, как внешние, так и внутренние: от религиозных фанатиков до сторонников дестабилизации власти в корыстных интересах. Это требует особого внимания к возникновению проблем занятости с внедрением технологий ИИ (количественный и качественный рост подобных проблем кажется неизбежным во всем мире).

В Египте (как и в других странах) дело доходит до крайне мрачных прогнозов относительно будущего взаимодействия между человеком и машиной. В частности, Мохаммед Джавдат, занимавший пост эксперта Google по ИИ, предупредил, что однажды ИИ может начать относиться к людям как к “отбросам” и создать свои собственные “машины для убийства” (Blunt P 2023a). Джавдат предупреждает, что современные модели изучения языка на основе алгоритмов считывают негативную информацию, которую мы размещаем в виртуальном пространстве, что в будущем может позволить машинам думать о человечестве как о чем-то отрицательном и злом, представляющем угрозу. Подобные заявления о возможных рисках, к которым необходимо относиться серьезно, при соответствующем целенаправленном продвижении могут посеять панику среди населения и вызвать у людей негативное отношение к дальнейшему развитию и использованию ИИ в повседневной жизни.

В 2019 г. в издании New York Times была опубликована статья “Согласно отчету, Египет использует приложения для отслеживания и таргетирования своих граждан” (“Egypt Is Using Apps to Track and Target Its Citizens, Report Says”), основанная на исследовании, проведенном экспертами по кибербезопасности, согласно которому египетское правительство может быть связано со злоумышленниками, совершившими серию кибератак на ряд египетских журналистов, активистов оппозиции, правозащитников и т. д. (Bergmanand Walsh 2019). Сами подобные действия берут свое начало еще в 2016 г., когда специалисты компании Check Point Technologies, занимающейся вопросами IT-безопасности, обнаружили, что хакеры использовали официальный магазин Google Play для распространения программ, собиравших информацию о геолокации, данные электронной почты, звонков и т. д. Одним из необходимых требований при установке было предоставление доступа к истории звонков и контактам пользователя. Исследователи Check Point обнаружили, что программа может быть использована в интересах правительства. “Коор-

динаты, встроенные в одну из фишинговых HTML-страниц, указывали на правительственное здание в Каире. Владелец регистрации домена, используемого злоумышленниками, указан как MCIT, что, по мнению исследователей, может указывать на Министерство связи и информационных технологий Египта (Ministry of Communications and Information Technology, MCIT)" (Lyngaas S 2019). Однако весьма вероятно, что кто-то из злоумышленников мог подставить под удар правительственную структуру Египта для последующей антиправительственной кампании по оказанию влияния. Таким образом, недоказанное использование технологий ИИ на втором уровне угроз информационно-психологической безопасности стало поводом для создания (без достаточных обоснований) образа антисоциального использования ИИ египетским правительством - т.е. угрозы первого уровня, на что указывает заголовок статьи в газете, не оставляющий сомнений в "виновности" египетских властей.

Таким образом, можно сделать вывод, что ЗИИИ на первом уровне возможно в Египте уже сегодня, что требует соответствующего осмысления и углубленного анализа, как на уровне экспертного сообщества, так и на уровне государственного управления.

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Широко распространенным видом мошенничества в Египте является фишинг. Согласно исследованию Лаборатории Касперского, в регионе Ближнего Востока и Африки фишинг используется в сфере курьерских доставок. Злоумышленники рассылают своим жертвам письма, в которых содержится ссылка для оплаты доставки, в противном случае товар не сможет быть доставлен. Перейдя по ссылке, жертвы попадали на сайт, выдающий себя за официальный сайт службы доставки, вводили реквизиты банковской карты, с которой затем мошенники списывали средства (Daily News Egypt 2023d). Другое исследование Лаборатории Касперского за 2022 г., посвященное электронным платежам, показало, что "57% пользователей в Египте сталкивались с попытками фишинга при использовании услуг онлайн-банкинга или цифровых кошельков. В исследовании также говорилось, что 54% пользователей сталкивались с поддельными веб-сайтами, а 57% попыток фишинга были совершены посредством текстовых сообщений или телефонных звонков с использованием методов социальной инженерии" (Daily News Egypt 2022e). Таким образом, не только правительство, но и банковский сектор, а также компании, так или иначе связанные с электронными платежами, должны повышать осведомленность граждан на тему цифровой безопасности.

Основным аспектом онлайн-мошенничества является социальная инженерия, которая помогает злоумышленнику завоевать доверие жертвы и заставить ее действовать в своих интересах. Например, в августе 2022 г. злоумышленники запустили в Египте онлайн-платформу и пообещали клиентам "большую финансовую выгоду" за счет комиссионных от услуг по майнингу и торговле криптовалютами. Были арестованы 29 человек, почти половина из которых были иностранными гражданами, похитившими более $600 000, используя мошенническую сеть HoggPool (Helou 2023). Этот пример показывает, что киберпреступники все более усердно используют информационно-психологические методы для манипулирования сознанием своих жертв.

А. Хасбини, руководитель группы глобальных исследований и анализа на Ближнем Востоке, в Турции и Африке Лаборатории Касперского, сообщил, что они отследили и предотвратили около 13 млн. информационных атак в Египте в течение первого квартала 2023 г. (Daily News Egypt 2023c). Хасбини заявил в интервью Daily News Egypt, что количество атак, направленных на банковские счета и данные клиентов, увеличилось на 186% по сравнению с 2022 г. Было подчеркнуто, что количество хакерских атак на информационную систему в частном банковском секторе

Египта активно растет. В то же время активизировались фишинговые атаки через электронную почту и SMS-сообщения. Около 75 000 пользователей в Египте подверглись фишинговым атакам в первом квартале 2022 г., и, согласно статистике, от 17 до 70% пользователей, получающих эти мошеннические электронные письма, переходят по ссылкам и попадают в электронную ловушку (Daily News Egypt 2023c).

С появлением чат-ботов усилились опасения по поводу того, что некоторые ИИ-системы могут представлять серьезную угрозу национальной и международной безопасности. Согласно недавнему отчету Group-IB (компании, специализирующейся на обеспечении кибербезопасности), страны Ближнего Востока и Северной Африки наиболее подвержены кибератакам, направленным на кражу аккаунтов, криптокошельков, историй браузеров и другой конфиденциальной информации, в то время как Египет занимает лидирующие позиции в регионе по количеству украденных аккаунтов ChatGPT – около 4500 в период с июня 2022 по май 2023 гг. В числе украденных данных – учетные записи для входа в систему и поисковые запросы (Ahram Online 2023b). Несмотря на то, что использование ИИ–ботов в Египте не столь широко распространено по сравнению с другими странами, в стране преобладает молодое население – “60% населения Египта составляют граждане в возрасте от 10 до 49 лет, а более 69,4 млн. человек пользуются мобильным Интернетом” (Salah 2023). Это позволяет предположить, что популярность этой технологии в египетском обществе в ближайшие годы будет только расти.

Выступление доктора М. Эль-Гинди (египетского исследователя, эксперта по кибербезопасности, специалиста и консультанта по киберпреступности для международных организаций) на Международном консорциуме под названием Global AI Ethics Network for Social Good (GAIEN4SG) на тему "Злонамеренное использование искусственного интеллекта: правовые и этические последствия", доказывает, что в ближайшем будущем наше собственное лицо может стать триггером для внедрения вредоносных программ злоумышленниками. После визуальной идентификации цели будет запущено соответствующее вредоносное ПО (ISSA Egypt 2022).

В Египте действуют две хакерские группировки – Horus Group и Anubis. Их цель – получение секретной информации о геополитических соперниках. Например, известно об организации ими серии кибератак на Эфиопию, которые были спровоцированы в связи со строящейся с 2012 года плотиной GERD (Great Ethiopian Renaissance Dam) (Munawer Q 2020). Деятельность этих групп свидетельствует о том, что киберпреступники могут заниматься шпионажем и кибершпионажем в интересах определенных государств или организаций. Разумеется, в процессе удешевления и распространения злонамеренных технологий ИИ они будут все чаще использоваться хакерами.

Таким образом, проблема киберпреступности при растущем использовании ИИ-технологий остро стоит в Египте, что говорит о дальнейшем росте ЗИИИ в этой стране.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

На третьем уровне угроз информационно-психологической безопасности в Египте находится, в частности, проблема злонамеренного использования дипфейков. В 2023 г. компания KnowBe4 (проводит обучение в области обеспечения кибербезопасности) провела исследование среди 800 сотрудников в возрасте от 18 до 54 лет, проживающих в Египте, странах Южной Африки и Кении. Было организовано прямое общение по электронной почте и видеосвязи. Основной сутью данного исследования являлось общение не с человеком, а со специальным ботом, созданным благодаря дипфейк-технологии. Согласно результатам этого исследования, 74% сотрудников не смогли распознать, что они взаимодействуют с ботом, а не с реальным человеком

(Shankar 2023). Это отражает тот факт, насколько совершенной стала технология дипфейк – большинству людей бывает достаточно трудно, а порой и вовсе невозможно распознать подделку в сети. В исследовании также прямо указывается на то, что в странах Африки остро стоит вопрос недостаточной осведомленности граждан по вопросам распространения дипфейков, а это подвергает риску миллионы людей. По словам А. Коллард, старшего вице-президента по контент-стратегии KnowBe4 Africa, "... дипфейк-платформы способны приводить к гражданским и общественным беспорядкам при их использовании для распространения ложной информации в политических и избирательных кампаниях, и несут угрозы современному цифровому обществу" (Shankar A, 2023). Следовательно, правительству Египта необходимо уделять повышенное внимание борьбе с угрозами, порождаемыми дипфейк-технологией, и ее потенциальным использованием для распространения дезинформации не только в политических и избирательных кампаниях, но и в бизнес-кампаниях. Также необходимо обезопасить граждан на законодательном уровне и путем проведения просветительских мероприятий по распознаванию дипфейк-контента, с которым они могут столкнуться в сети.

В настоящее время другой сферой, которая переживает кризис доверия, связанный с распространением дипфейков, является медиасфера. И. Э. Адиб, египетский журналист и бизнесмен, выступая на Арабском медиафоруме в Дубае в сентябре 2023 г., предупредил, что будущее медиаиндустрии не является безопасным. По его словам, уже сейчас мир вступил в эру недостоверных новостей, которые распространяются со скоростью света. На Форуме он привел личный пример, согласно которому его друзья прислали видео с Д. Трампом, которому подарили новое сердце, принадлежащее ранее мужчине-мусульманину, и Трамп якобы принял ислам (Al-Faour 2023). Проблема заключается в склонности людей верить в то, что они видят собственными глазами, даже если содержание не является аутентичным.

Так, в начале 2022 г. Египетский Дом фетв (Дар Аль-Ифта) выпустил заявление о том, что недопустимо использовать технологию ИИ для создания фейковых видео- или аудиозаписей людей, не имеющих к ним никакого отношения. "Создание таких роликов с помощью дипфейк-технологий с целью причинения вреда другим запрещено в соответствии со словами пророка Мухаммеда (мир ему и благословение Аллаха) "не причиняй вреда ни себе, ни другим", – заявили в Дар Аль-Ифта, добавив, что ислам запрещает запугивать других, даже ради развлечения (Ahram Online 2022a). Это заявление отражает отношение авторитетной египетской религиозной организации к технологиям, способным нанести ущерб отдельным лицам или обществу. При этом ислам не выступает за ограничение развития информационных технологий, однако, указывает на то, что моральные ограничения должны иметь первостепенное значение. Кроме того, "Дар Аль-Ифта" также указал, что распространение вводящей в заблуждение информации является уголовно наказуемым деянием в соответствии с законом 175/2018, охватывающим преступления в области информационных технологий" (Ahram Online 2022a).

По мнению М. Хенейн, доктора философии, исследователя и преподавателя Школы кибернетики Австралийского национального университета, основными аспектами, которые тормозят внедрение чат-ботов в повседневную жизнь египтян, являются: языковой барьер, цифровая безграмотность населения, законодательная база и культурные нормы страны (Salah 2023). Однако, в настоящее время в Египте активно проводятся просветительские мероприятия по ознакомлению населения с положительными и отрицательными последствиями инновационных технологий. Например, в марте 2023 г. Центр информации и поддержки принятия решений Египта (Information and Decision Support Center IDSC, IDSC) в сотрудничестве с ЮНЕСКО провел семинар по генеративному ИИ (MENA 2023). Основная цель таких семинаров – помочь молодым людям сформировать устойчивые взгляды на инновации, чтобы противостоять будущим технологическим вызовам.

Заключение

На основе проведенного анализа можно сделать вывод о том, что проблема злонамеренного использования ИИ охватывает в Египте все три уровня. Страна занимает одно из ведущих мест в рейтинге стран Ближнего Востока и Северной Африки в области развития ИИ и не планирует на останавливаться на достигнутом. Первый уровень угроз связан с возможностью манипулирования опасениями роста безработицы и рисками личной безопасности в связи с внедрением технологий ИИ. На втором уровне угроз наиболее проблемными сферами остаются виртуальное мошенничество, включая использование чат-ботов, и хакерские атаки на объекты критической инфраструктуры государства. Анализ третьего уровня угроз показал, что в обществе растет обеспокоенность способностью с применением ИИ создавать фейковую информацию и активно продвигать ее в виртуальном пространстве. Зачастую граждане сами оказываются жертвами использования подобных технологий из-за чрезмерной доверчивости и недостаточной осмотрительности. Таким образом, возникает необходимость разработки механизмов для обнаружения и предотвращения подобных случаев в будущем. Также необходимо повышать осведомленность граждан касательно рисков в сети Интернет, чтобы обеспечить положительную динамику противостоянию угрозам в области злонамеренного использования ИИ в стране. Однако, учитывая быстро меняющийся технологический ландшафт и изменчивую геополитическую обстановку, правительству Египта также стоит быть готовым к новым угрозам, которые могут возникнуть вследствие ЗИИИ.

Литература

Ahram Online (2022 a) Egypt’s Dar Al-Ifta prohibits deepfake video and audio clips. In: Ahram online. https://english.ahram.org.eg/NewsContent/1/64/454765/Egypt/Politics-/Egypt%E2%80%99s-Dar-AlIfta-prohibits-deepfake-video-and-au.aspx . Accessed 25 Jan 2024

Ahram Online (2023 b) Nearly 4,600 Egyptian ChatGPT accounts hacked: Report. In: Ahram Online. https://english.ahram.org.eg/NewsContent/3/1239/503415/Business/Tech/Nearly-,-Egyptian-ChatGPT-accounts-hacked-Report.aspx. Accessed 6 Feb 2024

Al-Faour N (2023) Egyptian journalist warns of threat posed by AI to media sector. In: Arab News. https://www.arabnews.com/node/2381501/media. Accessed 8 Feb 2024

Allianz (2024) The Sphinx’s enigma: testing Egypt’s political and economic stability again. In: The Allianz Group. https://www.allianz.com/en/economic_research/country-and-sector-risk/country-risk/egypt.html. Accessed 7 Feb 2024

Bergman R, Walsh D (2019) Egypt Is Using Apps to Track and Target Its Citizens, Report Says. In: The New York Times. https://www.nytimes.com/2019/10/03/world/middleeast/egypt-cyber-attack-phones.html. Accessed 7 Feb 2024

Blunt P (2023) Google’s Former Egyptian AI Expert Warns of Impending Disaster as AI Develops Negative Perception of Humanity. In: Asume Tech. https://asumetech.com/googles-former-egyptian-ai-expert-warns-of-impending-disaster-as-ai-develops-negative-perception-of-humanity/ . Accessed 29 Jan 2024

Business Today Egypt (2023) MCITMin discusses 2nd phase National Strategy for Artificial Intelligence. In: Business Today Egypt. https://www.businesstodayegypt.com/Article/1/3832/MCITMin-discusses-2nd-phase-National-Strategy-for-Artificial-Intelligence. Accessed 7 Feb 2024

Daily News Egypt (2023 a) 44% of employees in Egypt fear losing their jobs to AI. In: Daily News Egypt. https://www.dailynewsegypt.com/2023/02/20/44-of-employees-in-egypt-fear-losing-their-jobs-to-ai/ . Accessed 29 Jan 2024

Daily News Egypt (2023 b) 33% of employees in Egypt feel the lack of digital competencies. In: Daily News Egypt. https://www.dailynewsegypt.com/2023/09/19/33-of-employees-in-egypt-feel-the-lack-of-digital-competencies/. Accessed 4 Feb 2024

Daily News Egypt (2023 c) Kaspersky tackles 13 million cyber attacks in Egypt during 1Q 2023. In: Daily News Egypt. https://www.dailynewsegypt.com/2023/05/08/kaspersky-tackles-13-million-cyber-attacks-in-egypt-during-1q-2023/ . Accessed 27 Jan 2024

Daily News Egypt (2023 d) Kaspersky detects wave of courier service scams in Africa, Middle East, Turkiye. In: Daily News Egypt. https://www.dailynewsegypt.com/2023/08/06/kaspersky-detects-wave-of-courier-service-scams-in-africa-middle-east-turkiye/. Accessed 8 Feb 2024

Daily News Egypt (2022 e) More than half of Egypt's users encountered phishing attempts during electronic payments: Kaspersky. In: Daily News Egypt. https://www.dailynewsegypt.com/2022/07/28/more-than-half-of-egypts-users-encountered-phishing-attempts-during-electronic-payments-kaspersky/. Accessed 8 Feb 2024

Draya Egypt (2023) Artificial Intelligence in Egypt and Ways to Enhance it Within Framework of National Strategy. In: Strategic Forum for Public Policy and Development Studies. https://draya-eg.org/en/2023/02/08/artificial-intelligence-in-egypt-and-ways-to-enhance-it-within-framework-of-national-strategy/. Accessed 4 Feb 2024

Helou E A (2023) Crypto scam in Egypt robs investors of $620,000. In: Economy Middle East. https://economymiddleeast.com/news/crypto-scam-in-egypt-robs-investors-of-620000/. Accessed 4 Feb 2024

ISSA Egypt (2022) Malicious Use Of AI: Legal And Ethical Implications – GAIEN4SG Talk By Dr. Mohamed El-Guindy. In: Information Systems Security Association. https://issa-eg.org/malicious-use-of-ai-legal-and-ethical-implications-gaien4sg-talk-by-dr-mohamed-el-guindy/. Accessed 4 Feb 2024

Lyngaas S (2019) An ongoing hacking campaign targets dissidents in Egypt, researchers say. In: Cyberscoop. https://cyberscoop.com/egypt-hacking-check-point-technologies/. Accessed 7 Feb 2024

MENA (2023) Egypt's IDSC holds ChatGPT workshop to discuss future of AI platforms. In: Ahram Online. https://english.ahram.org.eg/NewsContent/3/1239/491777/Business/Tech/Egypt;s-IDSC-holds-ChatGPT-workshop-to-discuss-fut.aspx. Accessed 6 Feb 2024

Munawer Q (2020) Egyptian cyberattack on Ethiopian Security Agency website and some other. In: The Eastern Herald. https://easternherald.com/2020/06/24/egypt-cyber-attack-ethiopia/. Accessed 4 Feb 2024

Pessarlay W (2024) Egypt AL strategy focuses on governance, environment and human resources. In: Coin Geek. https://coingeek.com/egypt-ai-strategy-focuses-on-governance-environment-and-human-resources/. Accessed 7 Feb 2024

Salah A (2023) INTERVIEW: How Egyptians can benefit from ChatGPT, avoid potential negative consequences. In: Ahram Online. https://english.ahram.org.eg/NewsContent/1/2/498717/Egypt/Society/INTERVIEW-How-Egyptians-can-benefit-from-ChatGPT,-.aspx. Accessed 6 Feb 2024

Sayed M K (2018) 'Hawa Dawa': The Egyptian Fighting Air Pollution Using Artificial Intelligence. In: Egyptian Streets. https://egyptianstreets.com/2018/11/02/hawa-dawa-the-egyptian-fighting-air-pollution-using-artificial-intelligence/. Accessed 4 Feb 2024

Shankar A (2023) 74% vulnerable to deepfakes finds survey in Mauritius, Egypt, Botswana, South Africa, Kenya. In: Intelligent CIO https://www.intelligentcio.com/africa/2023/03/18/74-vulnerable-to-deepfakes-finds-survey-in-mauritius-egypt-botswana-south-africa-kenya/ . Accessed 23 Jan 2024

Statista Egypt (2024) Artificial Intelligence – Egypt. In: Statista. https://www.statista.com/outlook/tmo/artificial-intelligence/egypt. Accessed 7 Feb 2024

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Исламской Республике Иран

Е. Н. Пашенцев, П. В. Кузнецов

Введение

Ираном принят и реализуется Национальный план развития искусственного интеллекта, разработанный с целью вхождения страны в десятку государств-лидеров в области ИИ с текущего 13-го места в рейтинге Nature Index. Для этого запланировано проинвестировать 8 млрд. долл. в НИОКР в области ИИ (Tehran Times 2022a). Медицинские технологии на основе ИИ демонстрируют впечатляющие результаты в диагностике заболеваний – к примеру, точность диагностирования рака груди с использованием системы, разработанной Иранским Университетом Медицинских Наук (Iran University of Medical Sciences, IUMS), по последним данным, достигла 94% (Ziya 2021). Однако, как и в прочих областях, бурное развитие технологий ИИ может привести к их злонамеренному использованию. Вызовы ЗИИИ в Иране усугубляются следующими институциональными условиями: внутренние этнические и политические разногласия (Ziya 2021); проявления коррупции в госсекторе и ее последствия (Iran International 2023); и, что более важно, беспрецедентное внешнее давление на Иран со стороны Израиля и США. В той части ежегодного отчета разведывательного сообщества США, которая открыта для публичного ознакомления и касается оценки угроз, Иран вошёл в число четырёх стран, якобы представляющих наибольшую угрозу национальной безопасности США и международной стабильности (наряду с Россией, Китаем и КНДР). Примечательно, что в соответствующей главе доклада конкретно упоминается Израиль, которому также якобы угрожает Иран (Office of the Director of National Intelligence 2023). В то же время, США являются лидером в разработке и внедрении технологий ИИ, а Израиль видит в Иране главную угрозу себе и своему положению на Ближнем Востоке (Berman 2023). По словам Э. Замира, генерального директора Министерства обороны Израиля, «наша миссия состоит в том, чтобы превратить Государство Израиль в сверхдержаву ИИ и стать во главе крайне ограниченного числа мировых держав, входящих в этот клуб» (Williams and Maclean 2023). Таким образом, руководства США и Израиля обладают как очевидными стимулами, так и возможностями применения технологий ИИ против Ирана в гораздо большей степени, чем собственные возможности Ирана позволяют стране использовать ИИ против США и Израиля. Этот спектр институциональных проблем и негативно настроенных по отношению к стране внешних акторов создают чрезвычайно напряженную обстановку и благодатную почву для ЗИИИ в Иране.

Первый уровень угроз информационно-психологической безопасности посредством ЗИИИ

Иранское общество, как и общества многих других стран, в которых проводятся активные исследования в области ИИ, а сами технологии внедряются в различных сферах, обеспокоено тем деструктивным влиянием, которое ИИ может оказать на рынок труда (Iran Talent 2023). Однако, эксперты не могут прийти к консенсусу по этому вопросу. Некоторые представители научного сообщества Ирана акцентируют внимание на тех рисках, которые угрожают лишить работы специалистов среднего уровня квалификации вне зависимости от когнитивной и творческой составляющей их рабочего процесса — генеративные модели ИИ уже вполне способны решать за-

дачи, для которых ранее считалось необходимым нелинейное мышление, присущее человеческому мозгу. По мнению Хамидрезы Кешаварза, профессора Тегеранского Университета и Тегеранского Университета Медицинских Наук, низкоквалифицированные (уборка территории, грубая физическая работа и т.д.) и по-настоящему высококвалифицированные (требующие серьезного академического образования и наличия высокоразвитых интеллектуальных способностей) виды деятельности находятся в относительной безопасности. А вот рынок труда специалистов средней квалификации наиболее уязвим с точки зрения перспектив замены человека решениями на основе технологий ИИ (Khabar Online 2023). Массовые увольнения в крупнейших транснациональных корпорациях, где широко внедряются эти технологии, и перспективные риски увольнения ещё большего числа людей не могут не вызывать беспокойства значительной части населения Ирана. Такие естественные опасения могут целенаправленно усиливаться внутренними и внешними акторами, которые преследуют деструктивные цели, особенно если в ходе дальнейших и более масштабных процессов роботизации и внедрения технологий ИИ не будут приняты надлежащие меры по социальной защите и социально-ориентированному развитию.

Горячие споры и дискуссии, как правило, в странах Запада, вызывает возможное использование Ираном специальных технологий ИИ в целях обеспечения общественного порядка и автоматизации этого процесса. Настоящий информационный взрыв в западном медиа-пространстве вызвала новость об использовании иранскими властями ИИ для автоматического выявления с помощью городских систем видеонаблюдения женщин, не соблюдающих правила ношения закрытой одежды (хиджаба) (Alkhaldi and Ebrahim 2023). Данная дискуссия получила развитие в опубликованной в декабре 2023 г. статье на сайте *Council on Foreign Relations* (CFR), в которой говорилось, что ограничения на доступ в Интернет и законопроект о выявлении нарушителей моральных норм с помощью ИИ «как сообщается, привели к арестам более двадцати тысяч человек и убийствам более пятисот молодых протестующих» (George 2023), вероятно, подразумевая, что это произошло из-за протестов против указанных законопроекта и ограничений. Однако, вряд ли ненамеренно игнорируется тот факт, что в стране активно действуют «Аль-Каида» и другие террористические организации, некоторые из которых опираются на поддержку внешних сил и организуют теракты, происходящие в Иране довольно часто[8].

В СМИ периодически демонстрируются видеоролики (Jahan News 2023), на которых системы ИИ автоматически устанавливают личности заснятых женщин. По заявлениям СМИ, ряд сотрудников иранской полиции и представители иранского парламента подтвердили намерения правительства широко внедрить технологии ИИ для выявления различных правонарушений, в том числе нарушений «этического характера» (отсутствие закрытой одежды для женщин), а также для автоматизированного вынесения юридически-значимых решений по фактам, установленным данным методом. Вместе с тем, некоторые опасения относительно не морального аспекта подобной практики, а точности работы таких технологических решений, выразил бывший министр по информационным и телекоммуникационным технологиям Ирана А. Джахроми. Тем не менее, как сообщается, представители полицейского управления уверены, что возможные проблемы с точностью системы со временем будут устранены, когда ИИ получит достаточный объём данных для обучения (Ensaf News 2023). Несмотря на явно политизированный характер нападок на иранские власти в западных СМИ, вопросы, поднятые в дискуссии, действительно

[8] Последний и самый крупный теракт произошел 3 января 2024 г. в Кермане, когда тысячи людей пришли к месту захоронения генерал-лейтенанта Касема Сулеймани, ставшего в Иране символом борьбы с терроризмом, чтобы почтить его память в четвёртую годовщину смерти. В результате взрывов, устроенных террористами-смертниками, по меньшей мере 93 человека погибли и ещё несколько десятков получили ранения (Tehran Times 2024).

актуальны. В то же время, негативное воздействие этих проблем целенаправленно преувеличивается и приукрашиваются заинтересованными злонамеренными акторами, прежде всего внешними, которые, используя значительные финансовые, организационные, технологические и военные ресурсы, стремятся не защитить интересы иранского народа, а преследуют собственные цели в регионе.

Второй уровень угроз информационно-психологической безопасности посредством ЗИИИ

Опасность угроз второго уровня для информационно-психологической безопасности посредством ЗИИИ не менее реальна и продолжает возрастать. В декабре 2023 г. начальник управления гражданской обороны при Генеральном штабе ВС республики Иран, генерал Г. Джалали заявил (Tehran Times 2023), что недавний массовый сбой на заправочных станциях в стране был вызван вредоносным программным обеспечением (ВПО) в ходе целенаправленных кибератак. Не связывая этот факт с ЗИИИ напрямую, Г. Джалали отметил, что при осуществлении до 50% кибератак на критическую информационную инфраструктуру Ирана в той или иной степени используются технологии ИИ. В контексте упомянутой кибератаки можно предположить, что, следуя «мировой практике», злоумышленники могли использовать ИИ для подготовки фишинговых сообщений, посредством которых ВПО доставлялось в целевую систему. Ранее, в августе 2023 г., генерал Г. Джалали заявлял, что ИИ использовался внешними акторами для подготовки массовых протестных акций, и что Ирану следует научиться использовать ИИ для противодействия подобным проявлениям вмешательства в его внутренние дела (Mohammadzadegan, 2023).

Также уместно вспомнить, что генерал К. Сулеймани был убит в результате удара американского беспилотника 3 января 2020 г. возле международного аэропорта Багдада на полпути на встречу с премьер-министром Ирака Адилем Абдул-Махди. Видеозапись этого события транслировалась в режиме реального времени в Белый дом США, штаб-квартиру ЦРУ в Лэнгли, и как минимум еще на одну площадку для служащих Министерства обороны. Операцию курировали Д. Хаспел и М. Эспер, которые в то время занимали посты директора ЦРУ и министра обороны США соответственно (Dilanian and Cube 2020).

Белый дом направил Конгрессу уведомление, в котором были изложены юридическое и политическое обоснование авиаудара, в результате которого был убит иранский генерал. В уведомлении с целью легитимизации действий администрация Трампа сослалась на статью II и разрешение 2002 г. на применение военной силы против Ирака. Администрация заявила, что цель акции заключалась в том, чтобы «удержать Иран от проведения или поддержки дальнейших атак против сил и интересов Соединенных Штатов», а также «снизить атакующие возможности Ирана и поддерживаемых Силами Аль-Кудс ополченцев». (Setzer 2020) Подобные аргументы вызвали не только негативную реакцию со стороны Ирана, но и осуждение со стороны многих законодателей (в основном представителей демократической партии) (Choi 2020, Pengelly and Helmore 2020). По словам П. Сингера, аналитика и старшего научного сотрудника New America Foundation, «менее чем за одно поколение мы прошли путь от того, что казалось ненормальным и, возможно, даже научно-фантастическим, до того, когда это стало новой нормой» (Dilanian and Cube 2020). Разумеется, технологии ИИ активно использовались при подготовке и реализации этой операции, а также при представлении этой информации СМИ и ее последующем распространении.

Важно также вспомнить про трагический инцидент 29 ноября 2020 г., когда известный иранский учёный-ядерщик М. Фахризаде был убит в результате нападения на шоссе недалеко от Тегерана. По сообщению иранских властей, учёный стал жертвой нападения террористической

организации «Муджахидин-и-Хальк», которая осуществила взрыв электронного устройства, и, как предполагается, сделала это в интересах Израиля. Оборудование, установленное в фургоне, было нацелено на М. Фахризаде и сработало после опознания цели с помощью ИИ, ранив также и тех, кто сопровождал ученого (Motamedi 2020)

Таким образом, можно заключить, что технологии ИИ, очевидно, активно используются и на втором уровне угроз для информационно-психологической безопасности Ирана.

Третий уровень угроз информационно-психологической безопасности посредством ЗИИИ

Иранские СМИ в настоящее время находятся в фазе активной трансформации, связанной, как и во всём остальном мире, с быстрой цифровизацией общества. Помимо быстрого доступа к информации сокращается и количество звеньев в цепочке «создатель контента – потребитель контента». Вследствие этих изменений становится возможным осуществление гораздо более эффективных и разрушительных по своим последствиям информационных вбросов, в том числе созданных с использованием генеративных моделей ИИ. В качестве примера можно привести недавнюю и крайне оскорбительную публикацию, где на видео известный иранский священнослужитель Х. Ансарян якобы утверждает, что власть в стране захвачена ослами, и показывает на экране «доказательство», где к ослиному телу грубо пририсована человеческая голова (Iran NTV 2023). Подобные проявления подчёркивают необходимость активизировать работу Национальной комиссии Ирана по ИИ, поскольку Иран, как и все страны, остро нуждается во внедрении технологий обнаружения подобных дезинформационных материалов. В противном случае страна рискует потерять доверие населения к информации и может столкнутся с вызовами для информационно-психологической стабильности общества.

СМИ также неоднократно сообщали об осуществлении информационно-психологических операций в социальных сетях Facebook (Meta) и Twitter (в настоящее время носит название “X”) Центральным командованием Вооружённых сил США (CENTCOM) Министерства обороны США. Информационные операции CENTCOM, включая распространение антииранской пропаганды, проводились ранее и продолжались в течение длительных периодов времени. При проведении подобных операций ИИ использовался как для генерации текстов, так и для придания большего веса публикациям путём создания фотореалистичных изображений (дипфейков) якобы реальных пользователей социальных сетей, от имени которых были опубликованы тексты (Tehran Times 2022).

Несмотря на риски, связанные с внедрением ИИ, которые активно обсуждаются международным сообществом, в различных и довольно «чувствительных» отраслях Ирана уже ведутся исследования относительно использования чат-ботов на базе ИИ. Например, исследователи из Тегеранского университета медицинских наук в своей статье подробно описывают преимущества использования ИИ в медицине для обработки больших объемов данных и использования чат-ботов для медицинских консультаций (Hajialiasgari Khanahmadi and Atashi, 2023). Среди аспектов потенциального ЗИИИ и информаионно-психологических рисков упоминаются неправомерный доступ к конфиденциальной информации (разглашение врачебной тайны), возможные негативные психологические реакции пациентов на замену живого врача «бездушной машиной». А внешнее вмешательство или сбои в работе чат-ботов могут привести к медицинским ошибкам при назначении лечения. В то же время, исследователи приходят к скорее положительному выводу о необходимости внедрения ИИ в работу иранской системы здравоохранения.

Представители иранского руководства вплоть до Великого аятоллы начиная с 2020 г. неоднократно подчёркивали, что по очевидным и веским причинам Ирану следует идти в ногу с технологическим прогрессом и стать одной из ведущих стран в развитии технологий ИИ. Однако же, внедрение ИИ в религиозную практику может иметь неоднозначные последствия. Так, в 2023 г. М. Готби, глава Дома творчества и инноваций «Эшраг», буквально заявил, что «Роботы не могут заменить высокопоставленных священнослужителей, но они могут быть доверенным помощником, который поможет им подготовить фетву за пять часов вместо 50 дней». Внедрение ИИ в процесс принятия решений в такой чувствительной сфере, как религия, и особенно в такой религиозной стране, как Иран, где религиозные вопросы тесно переплетены с политическими (правительственными) решениями, может привести к самым разрушительным последствиям, если вышеупомянутые «роботы-помощники» будут скомпрометированы злоумышленниками (Bozorgmehr 2023). Особенно серьёзные риски возникнут с развитием эмоционального ИИ и в случае возможного дальнейшего обострения ситуации в Иране и вокруг него.

Заключение

Технологический прогресс всегда требует принятия обдуманных решений, а меры по развитию ИИ как ключевой технологии в переходный период к новому общественному и международному порядку требуют особо ответственного подхода. И, наряду с разработкой доверенных систем ИИ, необходимо заранее всесторонне анализировать потенциальные угрозы каждой такой системы, в том числе с учетом информационной и информационно-психологической безопасности и сопутствующих рисков. Внедрение ИИ в общественную жизнь без принятия мер предосторожности и разработки инструментов противодействия ЗИИИ может нанести существенный ущерб любому обществу и государству, а Иран, находясь на переднем крае сопротивления внешнему давлению, еще больше подвержен этим рискам. При этом злонамеренное влияние может оказываться как внешними акторами, цель которых состоит во влиянии на внутреннюю стабильность в стране или ее внешнеполитический курс, так и внутренними, которые могут пытаться незаконно обогащаться или поднять собственный политический рейтинг.

Однако, поскольку внешнее давление на страну значительно превышает давление внутреннее, а внешние злоумышленники обладают продвинутыми технологиями, угрозы ЗИИИ на втором и третьем уровнях информационно-психологической безопасности в первую очередь являются наиболее актуальными и опасными для Ирана. Страны-члены разведывательного альянса «Пять глаз» неоднократно демонстрировали свои возможности в проведении информационно-психологических операций, а эффективность использования странами НАТО новейшего вооружения и военной техники с применением технологий ИИ можно наблюдать на украинском театре военных действий. Поэтому, столь пристальное внимание к Ирану со стороны разведывательного сообщества США не может не вызвать обеспокоенности независимых наблюдателей.

Следует отметить, что иранские власти подошли к задаче достижения существенного прогресса в сфере ИИ системно и с позиций стратегического планирования. Можно предположить, что, если иранское правительство обратит пристальное внимание на угрозы ЗИИИ для информационно-психологической безопасности, подход к противодействию им станет столь же структурным и системным.

Литература

Alkhaldi C, Ebrahim N (2023) Iran proposes long jail terms, AI surveillance and crackdown on influencers in harsh new hijab law. In: CNN. https://edition.cnn.com/2023/08/02/middleeast/iran-hijab-draft-law-mime-intl/index.html. Accessed 03 Feb 2024

Berman L (2023) IDF set to focus on Iran, become ‘AI powerhouse,’ says Defense Ministry chief. In: Times of Israel. https://www.timesofisrael.com/liveblog_entry/idf-set-to-focus-on-iran-become-ai-powerhouse-says-defense-ministry/. Accessed 03 Feb 2024

Bozorgmehr N (2023) ‘Robots can help issue a fatwa’: Iran’s clerics look to harness AI. In: Financial Times. https://www.ft.com/content/9c1c3fd3-4aea-40ab-977b-24fe5527300c. Accessed 03 Feb 2024

Choi M (2020). 2020 Dems warn of escalation in Middle East after Soleimani killing. In: Politico. https://www.politico.com/news/2020/01/02/soleimani-2020-iran-democrats-093123. Accessed 03 Feb 2024

Dilanian K, Cube C (2020) Airport informants, overhead drones: How the U.S. killed Soleimani. NBC News, 10 January, https://www.nbcnews.com/news/mideast/airport-informants-overhead-drones-how-u-s-killed-soleimani-n1113726. Accessed 03 Feb 2024

Ensaf News (2023) The bright shade of face identification of naked people with smart cameras [هوشمند های دوربین ّ با ها حجاب ی ِ چهره شناسایی ِ روشن سایه ه]. https://ensafnews.com/408902/دور-با-ها-حجاب-ی ِ -چهره-شناسایی ِ -روشن-سایه/. Accessed 03 Feb 2024

George R (2023) The AI Assault on Women: What Iran’s Tech Enabled Morality Laws Indicate for Women’s Rights Movements. In: Council on foreign relations. https://www.cfr.org/blog/ai-assault-women-what-irans-tech-enabled-morality-laws-indicate-womens-rights-movements. Accessed 03 Feb 2024

Hajialiasgari F, Khanahmadi A, Atashi A (2023) Artificial intelligence chatbot in Iran Health Insurance Organization: a new era in service providing. Iran J Health Insur. Volume 6(2), pp. 91-102.

Iran International (2023) Iran's Biggest Corruption Case Rattles Ruling Hardliners. https://www.iranintl.com/en/202312062449. Accessed 03 Feb 2024

Iran NTV (2023) Happy courier - Deepfake Hossein Ansarian [پ یک .از صاریِ ان حسـ ین فـ یک دیـ پ شادی]. https://iranntv.com/908619-شادی-پـ یک-.-از صاریِ ان-حسـ ین-فـ یک-دیـ پ . Accessed 03 Feb 2024

Iran Talent (2023) Will artificial intelligence really make us all unemployed? [مصنو عی هوش واقـ عا کـ رد؟ خواهد بـ یـ کار را ما همـ ه]. https://www.irantalent.com/blog/impact-of-artificial-intelligence-job-losses/. Accessed 03 Feb 2024

Jahan News (2023) Identification of veiled women with artificial intelligence [ی ِ زنـ ان شـ ناسایی ِ -ی ِ -انزن- شـ ناسایی ِ -فـ یلم/https://www.jahannews.com/news/840663 .]مـ صنو عی هوش بـ ا حجاب. Accessed 03 Feb 2024 -حجاب-هوش-عی-مـ صنو

Khabar Online (2023) Who will be made unemployed by AI? [را-کـ سانی ّ -چـ ه-مـ صنو عی-هوش بـ یـ کار-را-کـ سانی ّ -چـ ه-مـ صنو عی-هوش/https://www.khabaronline.ir/news/1724621 .]کـ ند- می-بـ یـ کار. Accessed 03 Feb 2024 کـ ند- می

Mohammadzadegan A (2023) Iran prioritizes using AI for cyber defense, says defense official. In: IRNA. https://en.irna.ir/news/85197899/Iran-prioritizes-using-AI-for-cyber-defense-says-defense-official. Accessed 03 Feb 2024

Motamedi M (2020) Iranian official accuses Israel of killing Fakhrizadeh remotely. In: Al Jazeera. https://www.aljazeera.com/news/2020/11/30/iran-israel-killing-scientist-remotely-in-sophisticated-attack. Accessed 03 Feb 2024

Office of the Director of National Intelligence (2023) Annual threat assessment of the U.S. intelligence community. https://www.dni.gov/files/ODNI/documents/assessments/ATA-2023-Unclassified-Report.pdf. Accessed 03 Feb 2024

Pengelly M, Helmore E (2020). Impeachment: Warren accuses Trump of 'wag the dog' strike on Suleimani. In: The Guardian. https://www.theguardian.com/us-news/2020/jan/05/impeachment-warren-trump-wag-the-dog-qassem-suleimani-iran. Accessed 03 Feb 2024

Setzer E (2020) White House Releases Report Justifying Soleimani Strike. In: Lawfare. https://www.lawfaremedia.org/article/white-house-releases-report-justifying-soleimani-strike. Accessed 03 Feb 2024

Tehran Times (2022a) Iran plans to become a leading country in AI. https://www.tehrantimes.com/news/469628/Iran-plans-to-become-a-leading-country-in-AI. Accessed 03 Feb 2024

Tehran Times (2022b) Pentagon riding the “blue bird” in psychological warfare. https://www.tehrantimes.com/news/480127/Pentagon-riding-the-blue-bird-in-psychological-warfare. Accessed 03 Feb 2024

Tehran Times (2023) Iran says malware used in cyberattack on fuel stations detected. https://www.tehrantimes.com/news/492846/Iran-says-malware-used-in-cyberattack-on-fuel-stations-detected. Accessed 03 Feb 2024

Tehran Times (2024) Kerman terrorist attack Israeli attempt to compensate for losses: army chief. https://www.tehrantimes.com/news/493528/Kerman-terrorist-attack-Israeli-attempt-to-compensate-for-losses. Accessed 03 Feb 2024

Williams D, Maclean W (2023) Israel aims to be 'AI superpower', advance autonomous warfare. In: Reuters. https://www.reuters.com/world/middle-east/israel-aims-be-ai-superpower-advance-autonomous-warfare-2023-05-22/. Accessed 03 Feb 2024

Ziya MH (2021) The 13 crises facing Iran. In: Middle East Institute. https://www.mei.edu/publications/13-crises-facing-iran. Accessed 03 Feb 2024

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Федеративной Демократической Республике Эфиопия

С. А. Себекин

Введение

Эфиопия – одна из стран Африки, в масштабах которой для решения различных задач достаточно успешно внедряются технологии ИИ, и, несмотря не неспокойную социально-политическую обстановку, создаются обширные институциональные условия для его развития (Ade-Ibijola & Okonkwo, 2023, pp. 102, 104; Gadzala, 2018, pp. 1, 2, 5, 8). Приоритетными для внедрения систем ИИ являются сельскохозяйственный сектор (основа экономики Эфиопии – сельское хозяйство) (Federal Democratic Republic of Ethiopia, 2020, p. 26; Girmay, 2019, pp. 161-162, 166-167), здравоохранение, финансовый сектор, сектор государственного управления. В Эфиопии была создана и развивается так называемая «Долина Шеба» - технологический центр страны (по аналогии с «Силиконовой долиной» в США) (Eke, Wakunuma, & Akintoye, 2023a, p. 4). Системными и многовекторными разработками в области ИИ занимается Институт искусственного интеллекта Эфиопии. Существуют и частные компании, ведущие исследования в области ИИ, среди которых iCog Labs – частная исследовательская лаборатория ИИ, созданная в 2013 г. в Аддис-Абебе, предоставляющая широкий спектр услуг в области ИИ-исследований и разработок для внутренних и международных клиентов. При этом, одна из приоритетных заявленных целей – создание ИИ с учётом специфики Эфиопии и её ценностей. На сегодняшний день уровень доступа граждан к цифровым услугам пока ещё остаётся недостаточно высоким.

Население Эфиопии составляет около 120 млн. чел. По оценкам, из них доступ в интернет имеют лишь около 16-20% жителей. Пользователей социальных сетей ещё меньше – около 5% (Kemp, 2023). Тот факт, что в Эфиопии наблюдается довольно низкий уровень цифровизации среди населения может служить определенным препятствием для оказания воздействия на массовое сознание посредством ИИ. Однако в будущем, по мере развития в Эфиопии цифровой инфраструктуры, повышения общего уровня цифровизации и более обширного доступа к цифровым услугам, технологические и институциональные условия ЗИИИ существенно расширятся.

На сегодняшний день в Эфиопии сохраняется крайне нестабильная общественная, социально-политическая и социально-экономическая обстановка. Ключевые противоречия здесь сосредоточены вокруг конфликта между «традиционно мятежной» провинцией Эфиопии Тыграй и партией Народный фронт освобождения Тыграй (НФОТ) – с одной стороны, и Федеральным правительством (премьер-министр – Абий Ахмед Али) – с другой (Afriyie, Ayangbah, & Effah, 2023; Center for Preventive Action, 2023).

Конфликтный потенциал в Эфиопии сохраняется в силу следующих причин.

Во-первых, наличие многочисленных межэтнических противоречий и множества центров этнической напряжённости. Столкновения на локальном (низовом) и «межрегиональном» уровне между различными этническими группами происходят регулярно на почве земельных претензий, религиозных разногласий и пр.

Во-вторых, помимо «официальных» национальных вооруженных сил каждая провинция имела (или имеет) собственные «этнические» военизированные подразделения, не контролируемые федеральным правительством, которые регулярно проводили операции по этническим

чисткам. Несмотря на то, что федеральное правительство пытается ликвидировать эти подразделения и провести разоружение, не все провинции согласны с такой политикой.

В-третьих, сама социально-экономическая обстановка, которая ранее уже усугублялась как чисто природными факторами – голод вследствие засухи и нашествия саранчи, так и гуманитарными кризисами как следствиями вооружённого конфликта – например, вынужденная миграция, отсутствие продовольствия и т.д.

Помимо внутренних противоречий, Эфиопия имеет сильные разногласия и с соседними странами.

Сегодня экзистенциально важным для Эфиопии остаётся вопрос о получении выхода к Красному морю. Эти стремления оспаривает Эритрея, которая отделилась от Эфиопии в 1993 г., из-за чего последняя и потеряла выход к морю. Эфиопия также имеет сильные противоречия с Египтом и Суданом по вопросу распределения воды в Ниле. В 2023 г. Эфиопия начала заполнять водохранилище плотины "Возрождение", что для Египта является явно эскалационным шагом, затрудняющим диалог.

Таким образом, если технические и инфраструктурные условия применения ИИ ещё зреют, то социально-политические и экономические условия для достижения деструктивных эффектов посредством ЗИИИ в Эфиопии сложились уже давно. Если в будущем уровень цифровизации Эфиопии увеличится, а коренные проблемы не будут решены, в совокупности эти институциональные факторы создадут большой синергетический потенциал для реализации ЗИИИ в Эфиопии посредством воздействия на массовое сознание с целью достижения конкретных эффектов со стороны заинтересованных акторов – как внутренних, так и внешних, как явно, так и негласно.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Первая и самая очевидная проблема, которая может возникнуть в связи с широким внедрением зарубежных систем ИИ в Эфиопии, связана с тем, что алгоритмы на основе машинного обучения, обученные на «зарубежных» данных и «впитавшие» в себя западные (или другие) ценности, этику, представления о способах решения поставленной задачи, в реалиях Эфиопии с её совершенно иными этно-культурными, этическими, политическими и экономическими традициями, могут оказаться попросту неэффективными или даже привести к негативным эффектам и нести угрозу социальной, политической и информационно-психологической стабильности (Birhane, 2023, p. 250; Blackwell, Damena, & Tegegne, 2021; Eke, Wakunuma, & Akintoye, 2023a, p. 2-3; Eke, Wakunuma, & Akintoye, 2023b, p. VI; Okolo, Aruleba, & Obaido, 2023, p. 54). Эксперты из Африки обеспокоены, тем, что «неафриканские» ИИ-технологии для решения «африканских» проблем не учитывают их специфику (Birhane, 2023, pp. 250, 254-255; Eke, Wakunuma, & Akintoye, 2023a, p. 1-2).

Другой существенный вызов скрыт в обратной стороне проблемы, когда с целью обучения специальных алгоритмов для решения задач в условиях Эфиопии и для персонализации услуг и контента будут массово собираться и использоваться данные о её жителях (Birhane, 2023, pp. 249; 251-252). Доступ зарубежных акторов к этим данным открывает широкий спектр возможностей для манипулирования ими и воздействия на целевые аудитории в Эфиопии (о чём будет сказано далее).

Наконец, многие эксперты из Африки опасаются, что массовое внедрение «неафриканских» ИИ-систем для решения определённых проблем и иностранной цифровой инфраструктуры в целом сделает страны Африки излишне зависимыми от импортных технологий и ввергнет их в

«ИИ-неоколониализм» – так называемую «Алгоритмическую колонизацию Африки» (Adams, 2021; Birhane, 2023; Eke, Wakunuma, & Akintoye, 2023b, p. VI) (или «Цифровую колонизацию Африки), при которой заинтересованные акторы будут использовать ИИ-технологии не только для решения острых проблем Африки, но и для негласного влияния на экономические, политические и социальные процессы в регионе с целью реализации своих интересов. Также, предполагается, что среди главных интересов компании и корпораций, работающих в области ИИ, будет не соблюдение этики и обучение ИИ с учётом этнокультурной специфики, а получение прибыли за счёт экспорта ИИ-технологий в страны Африки (Birhane, 2023, pp. 251-252; Okolo, Aruleba, & Obaido, 2023, p. 41, 54; Eke, Wakunuma, & Akintoye, 2023b, p. VI).

Другим очевидным вызовом информационно-психологической безопасности вследствие ЗИИИ первого уровня для Эфиопии является автоматизация, и, как следствие, сокращение рабочих мест, которое в условиях рассматриваемой страны может принять довольно массовый характер (Girmay, 2019, p. 170). Дело в том, что в развивающихся странах (к которым относится и Эфиопия) последствия от автоматизации рабочих мест вследствие внедрения ИИ могут оказаться гораздо более глубокими, чем для развитых стран (The Conversation, 2023). Во-первых, структура экономики развитых стран гораздо более «многовекторная» и сложная с точки зрения имеющихся секторов, что обуславливает наличие множества высококвалифицированных рабочих мест и тем самым снижает риски резкой и тотальной ИИ-автоматизации. Во-вторых, развитые страны имеют соответственно более развитую экономику и обладают множеством ресурсов для проведения гибкой политики автоматизации и создания новых высококвалифицированных рабочих мест взамен старых посредством сильных образовательных программ и программ переобучения/повышения квалификации. При этом развитые страны могут позволить себе реализовать дополнительные меры прямой поддержки – например, базовый безусловный доход и различные пособия, что вряд ли будет под силу Эфиопии. Несмотря на разные прогнозы о том, что ИИ способен как минимум не сократить или даже создать большее количество рабочих мест, для развивающихся стран Африки с их менее развитой экономической структурой данный вопрос может иметь более глубокие последствия. Например, основа экономики Эфиопии – сельское хозяйство и сфера услуг – сектора, которые являются одними из самых перспективных для модернизации вследствие широкого внедрения ИИ-систем (Federal Democratic Republic of Ethiopia, 2020, p. 9, 26; Girmay, 2019, pp. 161-162; United Nations, 2023).

Несмотря на хорошие показатели развития профессионального образования в области ИИ, проблема усугубляется всё ещё недостаточным уровнем общего высшего образования в Эфиопии и его массовой доступности для широких слоев населения, что ещё более осложняет возможность реализации трудового потенциала в новых условиях. В Эфиопии доля молодежи в общей численности населения намного большая, чем в развитых странах, а именно она может лишиться работы с постепенным совершенствованием технологий ИИ и оснащенных ИИ роботов с низкими шансами найти равноценную замену, если не вводить своевременно программы переобучения и адаптации к новым технологиям, не создавать более технологичные рабочие места.

Психологический эффект первого уровня угроз от того, что ИИ «отнимет» работу, особенно при злонамеренном использовании проблем переходной экономики, высокой имущественной поляризации, может оказаться гораздо крайне деструктивным и в условиях Эфиопии привести к дестабилизации общественно-политической ситуации, развитию теневого сектора экономики и криминогенной обстановки, поиску незаконных источников средств к существованию. Создание новых высококвалифицированных рабочих мест в Эфиопии требует от правительства развития образовательных программ в сфере ИИ (в т.ч. эффективных программ переквалификации), повышение доступности этого образования, и, главное – равномерного распределения рабочих

мест среди различных этнических групп. Недооценка грядущих проблем может привести к «неолуддизму» в его разных формах – протесты против замены рабочих мест алгоритмами, забастовки и пикеты, и даже очередные войны против федерального правительства.

В недалеком будущем нагнетаемая «истерия» о массовой автоматизации и сокращении рабочих мест может иметь ярко выраженный манипулятивный характер и использоваться с целью явной дестабилизации общественной стабильности. Торможение же внедрения технологий ИИ закрепит отсталость страны и не может стать основой для решения ее проблем.

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Эфиопия с её традиционно нестабильной социально-политической и военно-политической обстановкой может стать мишенью различного спектра кибервоздействий – начиная от кибератак на системы критической инфраструктуры и заканчивая фишингом.

На сегодняшний день Эфиопия сталкивается с медленнорастущим количеством кибератак на системы своей инфраструктуры. Согласно официальной статистике Управления безопасности информационных сетей Эфиопии – главной правительственной структуры, ответственной за обеспечение кибербезопасности, – за последний финансовый год (2022-2023) количество кибератак на системы Эфиопии достигло около 7000 (Ena, 2023b; Ethiopian Monitor, 2023; Reqiq Staff, 2023). И хотя количество кибервоздействий на системы Эфиопии не так велико по сравнению с другими странами (например, в другом государстве-члене БРИКС – ЮАР – было зарегистрировано около 106 тыс. атак через бэкдоры и шпионские программы, а в более крупных странах-членах БРИКС этот показатель ещё выше), согласно Kaspersky Global Research and Analysis Team, масштаб и профессионализм кибервоздействий растёт (Teshome, 2023). Более того, Kaspersky Global Research and Analysis регистрирует немного иные данные – 18 тыс. кибератак и 30 тыс. атак программ-вымогателей (Teshome, 2023).

Кибервоздействия в основном осуществляются в отношении финансовых учреждений, секторов здравоохранения, образования, обеспечения безопасности, СМИ, правительственных структур (Ena, 2023a; Ethiopian Monitor, 2023; Reqiq Staff, 2023; Teshome, 2023). Основные типы кибератак и инструменты, использующиеся против систем и населения Эфиопии – DDoS-атаки (атаки типа «отказ в обслуживании»), сканирование и проникновение в системы, использование вредоносного ПО (в т.ч. программ-вымогателей) атаки на веб-сайты (Ena, 2023a; Ethiopian Monitor, 2023; Reqiq Staff, 2023; Teshome, 2023).

Согласно Kaspersky Global Research and Analysis и отчёту Microsoft Security Intelligence Report, в последние годы Эфиопия наиболее активно подвергается фишинговым атакам с применением программ-вымогателей (Microsoft, 2023; Tessema, 2023b). Один из самых масштабных случаев произошел в Эфиопии в 2023 г., когда в Министерстве финансов Эфиопии из-за фишинговой атаки мошенникам было переведено около $5 млн, которые предназначались Африканскому банку развития (Tessema, 2023a). При организации атаки мошенники использовали учётные данные АБР. Самое важное, что данный инцидент привёл к реальным последствиям и спровоцировал дипломатический скандал, так как после него правоохранительные органы Эфиопии задержали двоих сотрудников представительства АБР в Адидис-Абебе с применением насилия по подозрению в мошенничестве из-за того, что один из сотрудников Абдула Камара не подтвердил наличие перевода (Horn Observer Contributor, 2023). Это, в свою очередь, также может указывать на ещё одну проблему обеспечения кибербезопасности в Эфиопии - низкий уровень цифровой грамотности и недостаточную осведомлённость в области кибергигиены, так как в Министерстве финансов Эфиопии не проверили номера счетов, по которым осуществлялся перевод.

Вместе с тем, развитие цифровизации и цифровой экономики способно расширить поле кибервоздействий и спродуцировать рост как количества, так и качества кибервоздействий в отношении Эфиопии. В частности, возможны угрозы применения ИИ для осуществления целевых кибератак и отправки таргетированных фишинговых сообщений. Анализ больших данных позволит адаптировать кибератаки к конкретной цели – например, конкретной организации, производству, системе и т.д. Фишинг с применением ИИ, алгоритмов машинного обучения и анализа данных злоумышленники позволит генерировать персонализированные текстовые сообщения для конкретных «особо важных» лиц, каковыми могут быть первые лица компании, государственного института и т.д. (Bahnsen et al., 2018; Goldman, 2022; Guembe et al., 2022, pp. 84-85, 89, 96-97, 102; Seymour & Tully, 2016; Zouave et al., 2020, p. 22-23). Посредством таких кибервоздействий в условиях Эфиопии можно легко разжечь межэтнический конфликт и спровоцировать новые волны социальной напряжённости. Например, существует реальная перспектива массовой рассылки фишинговых сообщений от имени предполагаемого лидера оппозиционно-настроенной этнической группировки с призывами начать войну против другого народа или даже федерального правительства. Можно представить фишинговые письма, содержащие информацию о сборе средств на организацию ополчения.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Третий уровень угроз связан непосредственно с конкретными технологиями ЗИИИ, которые в условиях Эфиопии могут привести к тяжелым последствиям и дестабилизацией военно-политической, социальной и общественной обстановки. Реальный сценарий был продемонстрирован, когда была широко растиражирован дипфейк с применеием ИИ о том, что глава Партии процветания в регионе Амхара Гирма Ешитила был убит именно боевиками этнонационалистического ополчения Амхара ФАНО (Addis Insight, 2023). В 2023 г. мошенники пытались выдать себя за председателя комиссии Африканского союза (штаб-квартира которого расположена в Аддис-Абебе) Мусу Факи, используя созданный дипфейк для совершения видеозвонков нескольким европейским лидерам (Adjetey, 2023).

Конкретный деструктивный социально-политический эффект могут оказать чат-боты, обученные на определённой информации и запрограммированные в соответствии с определёнными идеологическими, политическими и прочими ценностными установками (Mihajlenok & Malysheva, 2020). К примеру, чат-боты могут распространять на специальных цифровых платформах среди народа амхара информацию о том, что спорные территории на самом деле исторически принадлежат народу тиграи, или наоборот – это может спровоцировать выражение несогласия посредством локальных столкновений и привести к дестабилизации общественного порядка.

Множество негативных сценариев в условиях Эфиопии позволит реализовать использование прогностических возможностей ИИ в качестве «прогностического оружия». Проанализировав определённые данные – уровень социальной стабильности, политические предпочтения, степень лояльности к федеральному правительству – ИИ, использованный для прогнозирования уровня социальной стабильности в той или иной провинции, может предсказать, что, к примеру, в районе Сомали произойдет социальный взрыв. Если ИИ выявит, что в провинции амхара растёт число лиц мужского пола, то он может сделать выводы, что через несколько лет этот народ попытается отвоевать спорные территории. Все это может еще больше дестабилизировать ситуацию в провинциях.

Широкое поле для применения в условиях Эфиопии получит технология целевого автоматизированного профилирования на основе ИИ, которая позволяет составлять психологические портреты и классифицировать целевых интернет-пользователей на основе анализа открытых (как правило) данных социальных сетей, интернет-ресурсов, поисковых запросов и т.д. для выявления их психологических особенностей, эмоционального фона и даже прогнозирования будущих психологических состояний с целью последующего оказания необходимого воздействия и мотивирования их к осуществлению определённых действий (Bilal et al., 2019; Guembe et al., 2022, p. 95; Zouave et al., 2020, p. 19). Например, ИИ может быть применён для анализа данных о больших целевых группах, которыми в данном случае могут оказаться различные народы Эфиопии – амхара, оромо, тиграи, омета, иробы и т.д. – чтобы затем с помощью алгоритма можно было составлять своего рода составлять их «социально-психологические карты» - анализировать поведенческие паттерны и особенности, выявлять массовые политические предпочтения отдельных народов, определять степень лояльности федеральному правительству и выделять раздражающие факторы с целью последующего оказания воздействия на массовое сознание, настройки нужной политической повестки, подталкивания к определённым действиям и т.д.

Помимо этого, множество негативных для информационно-психологической, политической и социальной стабильности Эфиопии может быть связано с самим внедрением специализированных алгоритмов для решения широкого спектра конкретных задач.

Заинтересованные акторы – некоторые государства и крупные корпорации - которые внедряют различные ИИ-платформы и цифровые экосистемы – могут быть заинтересованы в сложившейся в стране нестабильной ситуации с целью реализации собственных интересов, не всегда совместимых с интересами Эфиопии. Главное, что здесь нужно понимать – специальные платформы и приложения, в которые будут встроены ИИ-алгоритмы как для решения конкретных общественных задач, так и для развлечений, будут собирать огромное количество информации о гражданах Эфиопии. Внешние акторы, имеющие доступ к этим массивам данных, смогут не просто манипулировать этой информацией, но использовать её в своих целях с применением большого спектра специальных ИИ-технологий для достижения различных эффектов.

Далее это позволит применять конкретные ИИ-технологии для оказания таргетированного воздействия на массовое сознание населения Эфиопии (или конкретных целевых аудиторий) с целью достижения широкого спектра эффектов. Так, с помощью конкретных различных ИИ-технологий (обученных чат-ботов, ассистентов, дипфейков и т.д.) можно создать и массово распространить фейк о том, что спорные территории, на которые одновременно претендуют амхара и тиграи, отходят (или должны отойти) той или иной стороне.

Таким образом, с помощью конкретных ИИ-технологий посредством воздействия на массовое сознание те или иные заинтересованные акторы могут формировать нужную политическую повестку и даже вызывать необходимые общественно-политические процессы, создавать очаги напряжённости и провоцировать новые конфликты, чего в условиях Эфиопии – перманентной нестабильности и недоверия властям со стороны некоторых народов – добиться очень легко. Например, 31 мая 2021 г. интернет-изданием Kello Media опубликовало фейковую аудиозапись, в которой премьер-министр Абий Ахмед в ходе собрания Партии процветания якобы заявил, что они выиграли выборы и никто в течение ближайших 10 лет сможет сформировать альтернативное правительство (Addis Insight, 2023).

Корпорации, владеющие платформами и соответствующими технологиями ИИ, аффилированные со своим правительством, могут воспользоваться своим «цифровым могуществом» и «насаждать» в сознание жителей Эфиопии необходимость определённых преобразований – продвижение коммерческих, инфраструктурных, инвестиционных, и пр. проектов, даже если это

не в истинных интересах жителей Эфиопии, что может привести к очередной волне недовольства и конфликтам.

Важно обратить внимание, что алгоритмы, интегрированные в цифровые платформы, анализируют пользовательские данные с целью персонализации услуг и контента, и таким образом настраивают контент для конкретной личности (аудитории) в зависимости от предпочтений. Вместе с тем, подобная персонализация контента в условиях Эфиопии может привести к негативным последствиям, таким как еще большая поляризация народов по ключевым вопросам политики, принадлежности территорий, религиозного выбора и т.д. Если пользователям нравится контент определённого ценностного содержания, ИИ начинает всё чаще показывать аналогичную информацию, помещая их в своего рода «информационный пузырь», тем самым создавая впечатление, что их убеждения единственно верные и негласно определяя мировоззренческие взгляды. Так, народу оромо на этих цифровых платформах может демонстрироваться контент одного содержания – например, о необходимости создания собственного синода, а народу амхара – противоположный контент о недопустимости церковного раскола[9]. Это может происходить как в силу естественного обучения алгоритмов на платформах – оромийцы отмечают в качестве приоритетного свой контент, а амхара – свой, и ИИ просто начинает рекомендовать схожий контент, так и в силу искусственных причин – алгоритм специально так настроили. Подобными механизмами можно ещё больше поляризовать и без того дезинтегрированное и раздираемое противоречиями на этнической почве общество Эфиопии, а также цензурировать информацию в масштабах целого государства.

Реальный сценарий подобного деструктивного воздействия алгоритмов был продемонстрирован в ходе вооружённого конфликта между Тыграем и Федеральным правительством (2020-2022 гг.). Согласно Amnesty International, алгоритмы Facebook негласно поспособствовали распространению в Эфиопии деструктивного контента, в котором содержались призывы к насилию в отношении народа тиграи (Amnesty International, 2023). Системы модерации и цензурирования контента Facebook не распознали данные деструктивные сообщения в силу того, что такие языки, как амхарский и оромо не являлись языками, приоритетными для модерации (Allen, 2022).

Наконец, вопрос репрезентативности и объективности данных, а также то, насколько правильно и корректно будет обучен ИИ в целях решения определённых задач, приобретает особенную остроту в условиях сохраняющегося конфликтного потенциала в Эфиопии – наличия острых межэтнических противоречий и территориальных споров на этой почве, крайне нестабильной политической и социально-экономической ситуации в провинциях Эфиопии. Применение «предвзятого» ИИ в определенных сферах, обученного на нерепрезентативных данных, может быть чревато очень серьёзными последствиями, ещё сильнее дестабилизировать обстановку и привести к очередному «взрыву». Конкретные ИИ-технологии на основе машинного обучения, внедряемые для решения определённых социальных, общественных и политических задач как зарубежными компаниями, так и самими государственными структурами Эфиопии, могут дискриминировать те или иные группы людей, в т.ч. по этническому признаку. Например, любой социальный алгоритм, используемый для конкретных общественных задач – анализа результатов выпускных экзаменов, анализа резюме кандидатов при приёме не работу т.д. – может предоставить больше прав и возможностей народу оромо, негласно дискриминируя тиграйцев – если в конкретной компании работает больше оромо, обученный на этих данных ИИ и дальше будет отдавать приоритет представителям этого народа. Или, алгоритм, применяемый в политических процессах, может дать больше политических прав одним, маргинализируя других – скажем ИИ для определения состава правящей политической партии, обучившись на нерепрезентативных

[9] Напомним, что в феврале 2023 г. на этой почве в районах, населённых оромо, начались столкновения.

данных, теоретически может предоставить больше мест амхарцам, чем тиграйцам. В подобных сценариях дискриминации особенно будут подвержены представители малочисленных народов.

Нетрудно представить, что применение ИИ с целью воздействия на массовое сознание жителей Эфиопии может стать катализатором очередной волны насилия и привести к кровопролитным конфликтам на почве многочисленных этнических противоречий. И этим могут воспользоваться заинтересованные акторы – как внешние, так и внутренние, которые не согласны с политикой федерального правительства.

Заключение

Сегодня Эфиопия сталкивается как с целым комплексом внутренних противоречий, так и испытает давление со стороны внешних акторов. Подобные условия делают Эфиопию потенциальной мишенью для ЗИИИ с целью дестабилизации общественной и социально-политической ситуации посредством воздействия на массовое сознание со стороны заинтересованных акторов. Более того, сам факт применения социально- и политически-ориентированных алгоритмов в условиях Эфиопии может привести к дискриминации отдельных групп населения в том случае, если ИИ будет обучен на нерепрезентативных данных.

Вместе с тем, экзистенциальные вызовы могут быть спродуцированы распространением в Эфиопии импортных систем ИИ и цифровых платформ с внедрёнными в них алгоритмами. Потенциально, это может привести к «ИИ-неоколониализму», сделать Эфиопию излишне зависимой от зарубежной цифровой инфраструктуры и внедрению импортных ИИ-систем, не учитывающих местные интересы, специфику, культуру, менталитет. Предполагается, что при этом будут игнорироваться этические, правовые и социокультурные последствия применения таких систем. Более того, с помощью цифровых платформ корпорации могут получить доступ к огромному количеству данных о жителях Эфиопии, что в последующем позволит им манипулировать этими данными с целью реализации собственных интересов.

Литература

Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews,* 1-2 (46), 176-197. https://doi.org/10.1080/03080188.2020.1840225

Ade-Ibijola, A., & Okonkwo, C. (2023). Artificial Intelligence in Africa: Emerging Challenges. In D. O. Eke, K. Wakunuma, & S. Akintoye (Eds.), *Responsible AI in Africa: Challenges and Opportunities* (pp. 101-117). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-08215-3

Afriyie, F. A., Ayangbah, S., & Effah, K. O. (2023). Diagnosing Ethiopia's Tigray War: Reverberations in the Horn of Africa. Insight on Africa, 15 (2), 139-151. https://doi.org/10.1177/09750878231170177

Bilal, M., Gani, A., Lali, M. I. U., Marjani, M., Malik, N. (2019). Social Profiling: A Review, Taxonomy, and Challenges. *Cyberpsychology, Behavior and Social Networking,* 22 (7), 433-450. https://doi.org/10.1089/cyber.2018.0670

Birhane, A. (2023). Algorithmic Colonization of Africa. In S. Cave, & K. Dihal (Eds.), *Imagining AI: How the World Sees Intelligent Machines* (pp. 247-260). Oxford University Press. https://doi.org/10.1093/oso/9780192865366.003.0016

Blackwell, A. F., Damena, A., & Tegegne, T. (2021). Inventing Artificial Intelligence in Ethiopia. *Interdisciplinary Science Reviews,* 3 (46), 363-385. https://doi.org/10.1080/03080188.2020.1830234

Center for Preventive Action. (2023). Conflict in Ethiopia. Council on Foreign Relations, December 19. Retrieved January 5, 2024, from https://www.cfr.org/global-conflict-tracker/conflict/conflict-ethiopia

Eke, D. O., Wakunuma, K., & Akintoye, S. (2023a). Introducing Responsible AI in Africa. In D. O. Eke, K. Wakunuma, & S. Akintoye (Eds.), *Responsible AI in Africa: Challenges and Opportunities* (pp. 1-12). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-08215-3

Eke, D. O., Wakunuma, K., & Akintoye, S. (Eds.). (2023b). *Responsible AI in Africa: Challenges and Opportunities.* Palgrave Macmillan. https://doi.org/10.1007/978-3-031-08215-3

Ena. (2023a). Ethiopia Fends Off Exponential Spike in Cyber Attacks, Thwarts over 96 Percent of Attack. *Ena,* October 24. Retrieved January 8, 2024, from https://www.ena.et/web/eng/w/eng_3120234

Ena. (2023b). INSA Foils 6768 Cyber-attacks in Concluded Fiscal Year. *Ena,* July 24. Retrieved January 8, 2024, from https://www.ena.et/web/eng/w/eng_3120234

Ethiopian Monitor. (2023). In 12 months, INSA Foils Over 6,700 Cyberattacks on Ethiopia. *Ethiopian Monitor,* July 24. Retrieved January 24, 2024, from https://ethiopianmonitor.com/2023/07/24/insa-foils-over-6700-cyberattack-attempts/

Federal Democratic Republic of Ethiopia. (2020). Digital Ethiopia 2025 – A Strategy for Ethiopia Inclusive Prosperity. Ethiopian Legal Information Portal. Retrieved January 15, 2024, from https://www.lawethiopia.com/images/Policy_documents/Digital-Ethiopia-2025-Strategy-english.pdf

Gadzala, A. (2018). Coming to Life: Artificial Intelligence in Africa. The Atlantic Council, November 14. Retrieved January 10, 2024, from https://www.atlanticcouncil.org/wp-content/uploads/2019/09/Coming-to-Life-Artificial-Intelligence-in-Africa.pdf

Girmay, F. G. (2019). Artificial intelligence for Ethiopia: Opportunities and Challenges. The Information Technologist: An International Journal of Information and Communication Technology (ICT), 1 (16), 157-180.

Guembe, B., Azeta, A., Misra, S., Victor Chukwudi Osamor, V. S., Luis Fernandez-Sanz, L., & Pospelova, V. (2022). The Emerging Threat of Ai-driven Cyber Attacks: A Review. Applied Artificial Intelligence. An International Journal, 36 (1), 1-34. https://doi.org/10.1080/08839514.2022.2037254

Kemp, S. (2023). Digital 2023: Ethiopia. Datareportal, Febirary 13. Retrieved January 27, 2024, from https://datareportal.com/reports/digital-2023-ethiopia#:~:text=The%20state%20of%20digital%20in%20Ethiopia%20in%202023&text=There%20were%2020.86%20million%20internet,percent%20of%20the%20total%20population.

Microsoft. (2023). Microsoft Security Intelligence Report. Microsoft, 24. Retrieved January 27, 2024, from https://info.microsoft.com/SIRv24Report.html

Михайленок О. М., Малышева Г. А. (2020) Роботизация социальных сетей и ее политические последствия. Власть. 1 (28), 85-92. https://cyberleninka.ru/article/n/robotizatsiya-sotsialnyh-setey-i-ee-politicheskie-posledstviya/viewer. Дата обращения: 21.07.2022

Okolo, C. T., Aruleba, K., & Obaido, G. (2023). Responsible AI in Africa – Challenges and Opportunities. In D. O. Eke, K. Wakunuma, & S. Akintoye (Eds.), *Responsible AI in Africa: Challenges and Opportunities* (pp. 35-64). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-08215-3

Reqiq Staff. (2023). A Daunting Digital Frontier: The State Of Cybersecurity In Ethiopia. *Reqiq Insights,* November 3. Retrieved January 25, 2024, from https://reqiq.co/a-daunting-digital-frontier-the-state-of-cybersecurity-in-ethiopia/

Teshome, M. (2023). Cyber attacks bombard Ethiopia. *Capital Ethiopia,* June 12. Retrieved January 25, 2024, from https://www.capitalethiopia.com/2023/06/12/cyber-attacks-bombard-ethiopia/

Tessema, B. (2023). Increasing cyber-attacks target Ethiopia. *Abren,* June 13. Retrieved January 26, 2024, from https://abren.org/increasing-cyber-attacks-target-ethiopia/

The Conversation. (2023). Whose job will AI replace? Here's why a clerk in Ethiopia has more to fear than one in California. *The Conversation,* November 2. Retrieved January 25, 2024, from https://theconversation.com/whose-job-will-ai-replace-heres-why-a-clerk-in-ethiopia-has-more-to-fear-than-one-in-california-216735

The Economist. (2023). The world's deadliest war last year wasn't in Ukraine. *The Economist,* April 17. Retrieved January 8, 2024, from https://www.economist.com/international/2023/04/17/the-worlds-deadliest-war-last-year-wasnt-in-ukraine

United Nations. (2023). With AI, jobs are changing but no mass unemployment expected - UN labour experts. Department of Economic and Social Affairs. Retrieved January 25, 2024, from https://www.un.org/ru/desa/ai-jobs-are-changing-no-mass-unemployment-expected-un-labour-experts

Zouave E., Bruce M, Colde K, Jaitner M, Rodhe I, Gustafsson T. (2020). Artificially intelligent cyberattacks. Totalförsvarets forskningsinstitut FOI. Retrieved January 15, 2024, from https://www.statsvet.uu.se/digitalAssets/769/c_769530-l_3-k_rapport-foi-vt20.pdf

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Федеративной Республике Бразилия

Е. Н. Пашенцев, Д. Ю. Базаркина

Введение

В январе 2024 г. Президент Бразилии Л. И. Лула да Силва обнародовал план промышленной политики страны на ближайшие десять лет, направленный на стимулирование промышленного роста с предоставлением государственных кредитов и субсидий. В число приоритетных направлений, изложенных в плане, входит и цифровая трансформация: цель состоит в том, чтобы перевести работу 90% всех предприятий промышленного сектора Бразилии в цифровой формат (в настоящее время доля этих компаний составляет около 23,5%). Это потребует инвестиций в проект «Индустрия 4.0» с внедрением интеллектуальных цифровых технологий в производство и промышленные процессы, а также стимулирования национального производства полупроводников (Mari, 2024). Объем рынка систем ИИ в Бразилии достигнет $4,37 млрд. в 2024 г. Для сравнения – самый большой объем рынка ожидается в США ($106,50 млрд в 2024 г.). Предполагается, что объем рынка ИИ будет ежегодно расти на 17,65%, в результате чего к 2030 г. составит $11,59 млрд (Statista 2023a).

Компания, работающая в сфере создания визуального контента и маркетинга Getty Images, провела исследование под названием «VisualGPS», в котором приняли участие более 7000 респондентов из 25 стран. Результаты были следующими: четверо из шести бразильцев верят, что ИИ может оказать положительное влияние на их жизнь. Данный уровень технооптимизма выше, чем в среднем по миру, где только половина опрошенных разделила это менение. Согласно исследованию, бразильские пользователи в среднем на 15% больше заинтересованы в использовании систем ИИ по сравнению с остальным миром. В отличие от таких стран, как Соединенные Штаты, Канада, Франция, Великобритания и Австралия, всего лишь менее 34% бразильцев чувствуют угрозу со стороны внедрения и развития этой технологии (Mari, 2023a). В то же время, ЗИИИ уже продуцирует реальные угрозы информационно-психологической безопасности Бразилии, которые во многом обусловлены наличием острых социально-политических противоречий в стране. Ближайшее будущее также вызывает тревогу у бразильцев.

Главным приоритетом Лулы да Силва остается помощь той части из 71 млн. бразильцев (33% населения), которые борются с бедностью. Но МВФ, возможно, придерживается на этот счёт пессимистичного прогноза, согласно которому рост ВВП Бразилии составит всего 2% в год до 2028 г. Эти скромные темпы роста не смогут обеспечить эффективное сокращение бедности (Martin 2024, p. 2). По состоянию на январь 2024 г. почти половина бразильцев (49%) опасаются, что в ближайшие шесть месяцев 2024 г. их доходы скорее снизятся, чем увеличатся (36%), хотя они и ожидают улучшения ситуации на рынке труда (Reuters 2024). За первые шесть месяцев 2023 г. в Бразилии было зарегистрировано 1790 убийств по сравнению с 1526 в период с января по июнь 2022 г. (Instituto Sou da Paz, 2023). Муниципальные выборы в октябре 2024 г. станут политическим испытанием для президента Лулы да Силвы, так как они сулят новое столкновение с правыми силами во главе с Жаиром Болсонару. Дальнейшее развитие и распространение технологий ИИ на фоне растущих социально-политических проблем и противоречий естественным образом приведет к росту ЗИИИ в Бразилии.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Позитивное восприятие внедрения ИИ бразильцами и тех выгод, которые это принесет в будущем, имеет как свои плюсы, так и минусы. Высокий уровень технооптимизма способствует разработке и распространению высоких технологий, что в целом повышает уровень благосостояния общества. Однако в тех случаях, когда люди недооценивают риски для собственной безопасности, продуцируемые этими новыми технологиями, запоздалое осознание угроз может спровоцировать усиленную негативную реакцию общественности. В условиях складывающейся сегодня трудной внутренней ситуации и все более усиливающейся турбулентности на международной арене любые трудности, дисбалансы или ошибки при разработке систем ИИ могут быть усугублены злонамеренными действиями внутренних и внешних акторов. Многое зависит от способности правительства не только вовремя разработать и принять правовую базу в области развития ИИ в стране (работа в этом направлении продолжается), но и донести до широкой общественности тот факт, что развитие этих технологий предоставляет как огромные возможности, так и продуцирует серьезные вызовы.

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Согласно отчету, опубликованному компанией Trend Micro, Бразилия является второй по счету страной в мире, наиболее уязвимой к кибератакам (Mari 2023b).

В 2021 г. Бразилия и Эквадор являлись двумя латиноамериканскими странами с наибольшей долей пользователей, подвергшихся фишинговым атакам - 12,39% и 10,73% соответственно (Bianchi 2021). В 2020 г. Бразилия установила мировой рекорд по количеству фишинговых атак: каждый пятый пользователь Интернета в стране хотя бы раз подвергался фишингу (Mari 2022b). По данным «Лаборатории Касперского», в 2021 г. в Бразилии было совершено 25 млн. попыток таких атак, а в 2022 г. было зафиксировано 134 млн. попыток осуществления фишинга (Folha Vitória 2023).

Более того, киберпреступления набирают обороты в сфере электронной коммерции и в социальных сетях, так как здесь используются адреса электронной почты для аутентификации личностей пользователей. Как правило, люди используют одни и те же адреса электронной почты и одинаковые пароли на различных онлайн-сервисах. Таким образом, когда мошенники узнают пароль электронной почты жертвы, как правило, они получают доступ и к ее банковскому счету.

В феврале 2022 г. Бразилия была включена Spamhaus – международной организацией, отслеживающей спамеров и активность, связанную со спамом – в список стран, где спам-боты получили наибольшее распространение. Большинство из этих ботов используются для рассылки спама, фишинга, DDoS-атак (распределенные атаки типа «отказ в обслуживании») и других вредоносных воздействий. Аналитики связывают обилие ботов в цифровом пространстве Бразилии с техническими, политическими и социально-экономическими факторами (The Spamhaus Project 2022). Как и во всех других странах, рост кибератак в Бразилии сопровождается общим увеличением использования технологий ИИ.

Изощренность методов кибервоздействия представляет серьезные угрозы для предприятий и организационной инфраструктуры, поскольку кибератаки способны прерывать корпоративные операции, уничтожать критически важные данные и наносить ущерб репутации. Киберпреступники смогут осуществлять целенаправленные атаки с беспрецедентной скоростью и невиданного масштаба, обходя при этом традиционные системы обнаружения (Guembe et al.

2022). Из-за общего отставания Бразилии в области обеспечения кибербезопасности можно предположить, что уязвимость систем к подобным воздействиям высока.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Уже на протяжении нескольких лет для Бразилии остаются актуальными угрозы ЗИИИ третьего уровня, связанные прежде всего с использованием ботов в социальных сетях в ходе избирательных кампаний. Например, в случае с президентскими выборами 2022 г. количество поддельных аккаунтов в Twitter значительно выросло за несколько месяцев после подсчета голосов, причем большинство из этих аккаунтов атаковали лидера левых сил Л. И. Лулу да Силву. Полученные данные указывают на значительный всплеск предполагаемой активности ботов в Интернете (Lima 2023).

Национальные выборы 2022 г. в Бразилии характеризовались злонамеренным использованием дипфейков на основе технологий ИИ, когда сфабрикованные изображения и видеоролики использовались для распространения дезинформации, согласно которой ведущие кандидаты были вовлечены в различные скандальные и компрометирующие ситуации. Использование продвинутых систем ИИ и сложных инструментов редактирования для точной имитации голосов и мимики кандидатов позволяют создавать убедительный результат, что часто приводит к дестабилизации общественного доверия и искаженному восприятию хода избирательной кампании (Ünver 2023).

5 августа 2022 г. в социальных сетях был распространён политический видео-дипфейк, направленный против нынешнего президента Лулы да Силвы, на котором популярная бразильская телеведущая Р. Васконселлос якобы сообщает ложную информацию о результатах голосования на президентских выборах. Голос Р. Васконселлос был подвергнут обработке и небольшим изменениям с целью убедить зрителей в том, что, согласно опросу, результаты которого были опубликованы 15 августа 2022 г., Ж. Болсонару обходит Лула да Силву по предполагаемому количеству голосов. В реальности ситуация была иной – за Лулу были намерены проголосовать 44% избирателей, а за Ж. Болсонару 32% опрошенных респондентов. Видео-дипфейк (пусть и в течении непродолжительного времени) был размещен не только на платформе YouTube, но и также распространялся в группах WhatsApp и социальных сетях (Pacheco 2023).

По словам Х. Ринкона, регионального директора по Латинской Америке компании Avast, «для борьбы с фейковыми новостями необходим высокий уровень общественной осведомленности, поскольку те, кто начинает использовать новостные сайты, содержащие дезинформацию, могут все чаще попадать в водоворот фейковых новостей. Исследование, проведенное командой Avast AI, показывает, что более 17% веб-сайтов, распространяющих дезинформацию, содержат ссылки на другие сайты с фейковыми новостями. Это может в короткие сроки запустить цепочку потребления поддельных новостей» (Bento 2022).

Манипулирование электоратом всегда является опаснейшей формой воздействия, и оно может быть кратно усилено с применением технологий ИИ. В условиях Бразилии ЗИИИ в ходе электоральных процессов позволит осуществлять воздействие на сознание с особой эффективностью, подрывая доверие и волю избирателей к участию в политических процессах (Resende 2023).

Подрыв доверия к СМИ также может иметь серьезные последствия, особенно в условиях нестабильной политической обстановки. С. Грегори, программный директор некоммерческой правозащитной организации Witness, помогающей людям фиксировать нарушения прав человека, приводит пример, согласно которому в Бразилии, имеющей в своей истории прецеденты

полицейского насилия, граждане и активисты опасаются, что любое снятое ими видео, на котором офицер убивает гражданское лицо, больше не будет достаточным основанием для начала расследования. С. Грегори сообщает, что страх претворения в жизнь ситуации, когда реальные доказательства могут быть отвергнуты как фальшивые, стал постоянной темой на его семинарах (Hao 2019).

Проблема заключается не только в самом разоблачении практики создания и использования дипфейков, но и в их эффективном выявлении. Ситуация усугубляется тем, что власти пытаются представить в качестве дипфейка любую справедливую критику, основанную на реальных доказательствах – фотографиях, видео, аудио или письменных источниках. Профессор Х. Ли из Университета Южной Калифорнии полагает, что риск, возникающий из-за деструктивного использования дипфейков кроется в возможности их применения в качестве инструмента дискредитации подлинных видео-доказательств: «Даже если есть видео, на котором вы осуществляете какие-либо действия и выражаете мнение, ничто не помешает вам заявить, что это был дипфейк, и доказать обратное будет очень трудно». Политиков по всему миру обвиняли в такого рода уловках, включая мэра Сан-Паулу Ж. Дориа. В 2018 г. политик, будучи женатым, утверждал, что видео с его участием, содержащее непристойные сцены группового характера, было дипфейком, и никто не смог доказать обратное (Thomas 2020). Таким образом, злонамеренное использование дипфейков в Бразилии может представлять собой угрозу ЗИИИ одновременно как первого, так и третьего уровней.

Дипфейки могут использоваться как мощный инструмент дезинформации, подрывающий общественное сознание (Pinheiro de Resende, 2021). Применение дипфейков позволяет довольно легко вводить людей в заблуждение, так как они склонны верить той информации, которую непосредственно наблюдают, считая, что то или иное событие произошло на самом деле. Аудиоматериал также может быть сгенерирован искусственным путем. Журналистка М. Прадо в своей книге «Фейковые новости и ИИ: как алгоритмы могут применяться для борьбы с дезинформацией» подчеркивает, что дипфейк-аудиоматериалы могут быть легко распространены на таких платформах, как WhatsApp, который широко используется в Бразилии. С помощью доступного программного обеспечения, которое постоянно совершенствуется, можно легко имитировать человеческий голос. Жертвами в основном становятся общественные деятели, чьи речи легко найти в открытом доступе. Аудио-дипфейки также могут быть использованы для осуществления финансового мошенничества. «В одном случае сотрудник технологической компании получил голосовое сообщение от топ-менеджера с просьбой перевести ему немного денег. У него возникли подозрения, и сообщение было проанализировано компанией в области обеспечения безопасности, которая подтвердила, что сообщение было создано с использованием технологий ИИ» (Schmidt 2022).

Чем изощреннее дипфейки и чем более антидемократичный режим действует в государстве, чем больше уровень абсентеизма в обществе. Чем ниже уровень цифровой грамотности у населения, тем больше рисков дипфейки продуцируют для социальной стабильности. Чтобы нейтрализовать дипфейки – и фейки в целом – власти должны на системной и комплексной основе выявлять подделки, несущие угрозу социальной безопасности, и маркировать их, одновременно предоставляя факты. В то же время, без сознательного внимания и участия граждан к данной проблеме власти вряд ли захотят или смогут эффективно противодействовать злонамеренному использованию дипфейков. Отделить зёрна от плевел и показать, где правда, а где ложь, будет чрезвычайно сложно без основательных и необходимых на то политических предпосылок.

В Бразилии (как и в ряде других стран) становятся популярными виртуальные инфлюенсеры – цифровые «лидеры мнений». Выглядящие как реальные люди, виртуальные инфлюенсеры со-

зданы с помощью цифровых технологий и способны осуществлять определённый спектр действий, присущих живым людям – разговаривать, танцевать, играть. Кроме того, они могут демонстрировать приверженность определенным убеждениям, что делает их популярными в социальных сетях, и, возможно, именно поэтому у них так много подписчиков (Little Black Book 2022). В 2016 г. в социальной сети Инстаграм начала набирать популярность 19-летняя бразильско-американская звезда социальных сетей Л. Микела (известная также как М. Соуза), собравшая более 2,5 миллионов подписчиков на платформе и регулярно публикующая спонсорский контент в партнерстве с такими брендами, как BMW и Pacsun. Однако это вовсе не реальная девушка, переехавшая в Лос-Анджелес – это сугубо виртуальный персонаж, созданный с применением цифровых технологий. Однажды Микела была вынуждена признаться своим поклонникам, что ее «взломал» ее заклятый враг – тролль Бермуда, аффилированный с Д. Трампом (Petrarca 2018). Очарованные человекоподобной внешностью Микелы, многие начали задаваться вопросом: была ли она маркетинговым ходом, реальным человеком или чем-то совершенно иным. Наконец, в 2018 г. правда была раскрыта, когда ее создатели – Т. Макфедрис и С. Деку, сотрудники компании Brud (разрабатывает программные продукты в области робототехники, машинного обучения и ИИ), объявили, что они являются создателями Л. Микелы (а также тролля Бермуды) (Sheena 2023).

Обладая 31,2 миллионами подписчиков в социальных сетях, бразильская звезда социальных сетей Л. ду Магалу является крупнейшим виртуальным лидером мнений в Интернете (Petrarca 2018). Она была создана Ф. Траяно, генеральным директором Magazine Luiza, диверсифицированного бразильского маркетплейса, ориентированного на потребителя, включая Magalu – одного из крупнейших в стране ретейлеров с более чем 1300 физическими магазинами по всей стране. Лу была создана в далеком 2003 г. – как раз в то время, когда в сфере электронной коммерции начали проявляться признаки того, что в конечном итоге данный формат торговли может стать жизнеспособным вариантом традиционных розничных продаж. «Мы несем ответственность за гуманизацию Лу», – отметила менеджер по контенту и социальным сетям в Magalu А. Изо в интервью изданию Observer (Wierson 2021). Изо, которая руководит командой 3D-дизайнеров, программистов и маркетологов, тщательно следящих за каждым аспектом этого востребованного инфлюенсера, утверждает, что «у Лу миллионы поклонников, и когда она делает какие-либо заявления, занимая определенную позицию – например, по таким вопросам, как проблема домашнего насилия или защита прав ЛГБТ[10] – люди обращают на это внимание. Здесь, в Бразилии, Лу – это не просто маркетинговый ход, она оказывает влияние в прямом смысле этого слова и может освещать и воздействовать на важные проблемы общества» (Wierson 2021).

Виртуальный инфлюенсер – это цифровая «личность», которая публикует посты в социальных сетях, чтобы собрать многочисленную аудиторию преданных поклонников подобно человеку, способному оказывать влияние (по крайней мере, так это представляется). Очевидно, что такие «ИИ-агенты влияния» могут доносить контент определенного политического содержания, и вопрос о том, в какой степени это может являться манипуляцией в отношении не только взрослой, но и детской и подростковой аудиторией, остается открытым. По словам Якова Барта, доцента Северо-Восточного университета (маркетинг) и члена руководящего комитета Института экспериментального искусственного интеллекта в Северо-Восточном институте, «в определенном контексте использование “виртуальных” лидеров мнений более полезно, если учитывать их эффективность в сравнении с издержками с точки зрения изменений в мышлении потребителей после взаимодействия с ним» (Contreras 2024). М. Герлих из Швейцарской бизнес-школы SBS приходит к выводу, согласно которому виртуальные инфлюенсеры могут создавать более глубокие и долговременные связи с клиентами, и их гибкость при программировании и обучении

[10] Движение ЛГБТ объявлено в России экстремистским.

движков ИИ позволяет им адаптироваться к изменяющемуся поведению клиентов. Виртуальные инфлюенсеры могут обладать более высоким авторитетом, чем люди, и поэтому являться перспективными с точки зрения маркетингового влияния, а также способны эффективно стимулировать человека к совершению покупки и повысить общую узнаваемость бренда для компаний (Gerlich 2023, p. 19).

М. Мрад, профессор маркетинга из Американского университета Шарджи в Объединенных Арабских Эмиратах, признает, что концепция, согласно которой живые люди могут легко попасть под влияние цифровых лидеров мнений, может показаться неправдоподобной, однако на самом деле поколение Z – тех, кто родился в конце 1990-х - начале 2010-х гг. – активно взаимодействуют с ними. «Это поколение, похоже, испытывает определенную связь с подобными лидерами мнений в той мере, в какой они формируют с ними эмоциональные и ментальные отношения… Они проявляют к ним чувства любви и даже привязанности». В некоторых случаях они даже считают виртуальных кумиров более надежными, чем людей (Kugler 2023). По состоянию на май 2020 г. Бразилия, насчитывающая почти 9,2 миллиона пользователей Instagram, была латиноамериканской страной с наибольшим количеством цифровых лидеров мнений. Несмотря на значительное отставание, Аргентина заняла второе место, зарегистрировав более 1,1 миллиона таких виртуальных знаменитостей (Statista 2023b). Этот факт делает Бразилию особенно уязвимой для воздействий посредством данной технологии.

Весьма примечательно, насколько избирательно работает в Бразилии прогностическая аналитика, основанная на ИИ. 8 января 2023 г., ровно через неделю после третьей инаугурации Л. Инасиу Лулы да Силвы в качестве президента Бразилии, сторонники крайне правого экс-президента Жаира Болсонару начали штурмовать правительственные здания в Бразилии. «Все видели, что в Бразилии зреет насилие. За исключением крупнейших социальных сетей», - написал Марк Скотт из издания Politico на следующий день после восстания, утверждая, что «гиганты Силиконовой долины снова были застигнуты врасплох...” (Digital Action 2023). За несколько месяцев до атаки эксперты предупреждали, что ультраправые использовали платформы с шифрованием сообщений, такие как WhatsApp и Telegram для организации и распространения дезинформации и провоцирования мятежей, и призывали сохранять связанные с выборами меры безопасности в первые недели работы нового правительства. Многие обращались с подобными запросами, которые компании игнорировали (Digital Action 2023). Не только не были учтены голоса экспертов, но и не использовались ресурсы прогнозной аналитики, основанные на ИИ. С другой стороны, по словам Э. Сарайвы, главы ассоциации политических маркетологов Бразилии, ИИ позволит людям прогнозировать результаты муниципальных выборов 2024 г. заблаговременно (Silva 2023), что свидетельствует о потенциальной возможности использования технологий ИИ для получения предварительных результатов выборов, что может повлиять на решения избирателей о том, стоит ли идти ли на выборы и как следует голосовать. Можно сделать вывод, что в Бразилии (и не только) уже складываются условия, при которых прогностическая аналитика, основанная на ИИ, может быть использована в качестве прогностического оружия.

Дипфейки, виртуальные инфлюенсеры, чат-боты, прогностическая аналитика и другие средства влияния на общественное мнение с использованием технологий ИИ могут оказать самое деструктивное влияние на политические процессы в стране. Высший избирательный суд (The Superior Electoral Court, TSE) провел в течение первого квартала 2024 г. дебаты по регулированию использования технологий ИИ на следующих выборах. А. де Мораес, президент Высшего избирательного суда, полагает, что алгоритмы позволят, например, изменять видеозаписи кандидатов-оппонентов, где они делают заявления, которых никогда не делали. «Представьте, сколько людей могут попасть под воздействие подобной дезинформации, но в правдивости ко-

торой не будет сомнений. Конфликтный потенциал очень велик, и он, особенно с использованием ИИ, действительно может изменить или исказить результаты выборов, на которых наблюдается резкая полярность конкурирующих сил» (Tocarnia 2023).

Примечательно, что меры, направленные на подавление пропаганды и манипуляций в Интернете, предложенные правительством Лулы, встречают резкое сопротивление со стороны ведущих транснациональных технологических компаний США. Сейчас на рассмотрении Бразильского Конгресса находится Законопроект № 2630 (Senado Federal 2020). Официальное название – Закон Бразилии о свободе, ответственности и транспарентности в Интернете, который бразильские СМИ окрестили «Законопроектом о фейковых новостях», а его противники - законопроектом о цензуре. Законопроект призван бороться с распространением дезинформации, созданной с применением технологий ИИ. Необходимость регулирования в этой области давно назрела. Согласно исследованию Avast, четверо из пяти (79%) бразильцев попадали на фейковые новости о выборах 2022 г. в социальных сетях, и большинство (57%) не верят (или не до конца уверены), что социальные сети являются надежным источником информации. Кроме того, 86% бразильцев считают, что СМИ должны взять на себя ответственность за удаление фейковых новостей в своих сетях (Bento 2022). Вопрос о принятии срочных мер по контролю над этой ситуацией очевиден для подавляющего большинства бразильцев и самого бразильского правительства, но не для IT корпораций из США.

В начале мая 2023 г., когда законопроект вот-вот должен был быть одобрен, Google и Telegram использовали собственные платформы, чтобы продемонстрировать свое несогласие с законопроектом бразильским пользователям. 1 мая бразильцы были удивлены, увидев в знакомом поле поиска на домашней странице Google ссылку с надписью: «Законопроект о фейковых новостях может ухудшить работу вашего Интернета». Тот, кто переходил по ссылке, попадал на блог Google, в котором выражалась критика Законопроект № 2630, голосование по которому в бразильском конгрессе должно было состояться на следующий день. Домашняя страница поиска, которой пользуются более 90% из 160 миллионов интернет-пользователей в Бразилии, также содержала ссылку с информацией, согласно которой «законопроект о фейковых новостях может создать путаницу в отношении того, что в Бразилии является правдой, а что ложью» (Viana 2023). Стратегия Google также включала в себя рассылку электронных писем Youtube-блогерам, в которых сообщалось, что в их каналы будет инвестироваться меньше денег, и в качестве решения им предлагалось выступить с обращением к своему Конгрессу. Согласно исследованию Федерального университета Рио-де-Жанейро (Viana 2023), технологический гигант также изменил настройки результатов поиска, на первый план выдвигая свой собственный пост в блоге и другие статьи, содержащие критику законопроекта. «Бразилия собирается принять закон, который положит конец свободе слова», – говорится в сообщении мессенджера Telegram, отправленном в мае пользователям по поводу «Законопроекта 2630», который прошел Сенат и ожидал голосования в нижней палате Конгресса (France 24 2023).

Таким образом, есть все основания полагать, что ведущие IT компании США сознательно тормозят принятие эффективных мер против распространения дезинформации, создание и распространение которой все чаще обеспечивается технологиями ИИ. За этим кроются как корыстные финансовые интересы компаний (нежелание брать на себя юридические обязательства по борьбе с неосознанными или сознательными нарушителями общественных норм в Интернете), так и желание оказать давление на своих оппонентов, вплоть до провоцирования массовых беспорядков и свержения неугодного правительства, намеренного посягнуть на сверхприбыли. Дополнительным, но важным фактором является давление правительства США, которое стремится использовать американские компании как для получения конфиденциальной информации, так и для оказания давления на «неугодные» правительства, что в целом скорее объединяет, чем

разобщает позиции Вашингтона и ведущих IT-компаний, но не исключает некоторых острых противоречий между ними.

Заключение

Угрозы информационно-психологической безопасности посредством ЗИИИ в настоящее время имеют четко выраженный характер на втором и третьем уровнях, что не исключает их яркого проявления на первом уровне по мере развития технологий ИИ, масштабов их капитализации и сохранения острых социально-политических противоречий в Бразилии. Угрозы ЗИИИ растут как в количественном, так и в качественном отношении, и становятся более изощренными и разнообразными. Нахождение Бразилии в БРИКС открывает как новые возможности для страны, так и сулит значительные риски, в первую очередь со стороны тех сил, которые отрицают проведение страной своего независимого политического курса. В будущем возможен рост ЗИИИ, поскольку уровень кибербезопасности в Бразилии остается довольно низким, законодательные инициативы не соответствуют реальной применяемой практике злоумышленников, а системный подход к противодействию ЗИИИ в области информационно-психологической безопасности отсутствует. В то же время, четкое стремление правительства Лулы противостоять информационно-психологическому воздействию может стать исходной основой для разработки такого системного подхода в будущем.

Литература

Bento G (2022) Fake news: 79% dos brasileiros encontraram mentiras sobre as eleições 2022 na internet. In: Olhar Digital. https://olhardigital.com.br/2022/10/04/pro/fake-news-79-dos-brasileiros-encontraram-mentiras-sobre-as-eleicoes-2022-na-internet/. Accessed 02 Feb 2024

Bianchi T (2021) Latin American & the Caribbean countries most targeted by phishing attacks in 2021. In: Statista. https://www.statista.com/statistics/997956/phishing-attack-user-share-latin-america-country/. Accessed 8 Dec 2023

Contreras C (2024) Is AI killing the social media star? How companies are cashing in on virtual influencers. In: Phys.Org. https://phys.org/news/2024-01-ai-social-media-star-companies.html. Accessed 02 Feb 2024

Digital Action (2023) Brazil municipal elections: Have Big Tech companies learnt anything from the January 8th attacks? In: Year of Democracy. https://yearofdemocracy.org/case-study/brazil-municipal-elections-have-big-tech-companies-learnt-anything-from-the-january-8th-attacks/. Accessed 02 Feb 2024

Folha Vitória (2023) Brasil teve 134 milhões de tentativas de phishing em um ano. https://www.folhavitoria.com.br/geral/noticia/09/2023/brasil-teve-134-milhoes-de-tentativas-de-phishing-em-um-ano. Accessed 02 Feb 2024

France 24 (2023) US tech giant Telegram calls Brazil disinformation law 'attack on democracy'. https://www.france24.com/en/americas/20230509-messaging-app-telegram-calls-brazil-disinformation-law-attack-on-democracy. Accessed 02 Feb 2024

Gerlich M (2023) The Power of Virtual Influencers: Impact on Consumer Behaviour and Attitudes in the Age of AI. *Administrative Sciences*. Issue 13(8): 178. https://doi.org/10.3390/admsci13080178.

Guembe B, Azeta A, Misra S, Chukwudi Osamor V, Fernandez-Sanz l, Pospelova V (2022) The Emerging Threat of Ai-driven Cyber Attacks: A Review. In: Applied Artificial Intelligence, issue 1.

Hao K (2019) The biggest threat of deepfakes isn't the deepfakes themselves. In: MIT Technology Review. https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/. Accessed 21 Jun 2022

Instituto Sou da Paz (2023) G1. Monitor da Violência: RJ registra 10 assassinatos por dia e tem 2ª maior alta do país no 1º semester. https://soudapaz.org/noticias/g1-monitor-da-violencia-rj-registra-10-assassinatos-por-dia-e-tem-2a-maior-alta-do-pais-no-1o-semestre/. Accessed 02 Feb 2024

Kugler L (2023) Virtual Influencers in the Real World. Communications of the ACM. March, Volume 66, Issue 3, p. 23-25. https://doi.org/10.1145/3579635.

Lima C (2023) Fake Twitter accounts denying election surged in Brazil, analysis finds. In: The Washington Post. https://www.washingtonpost.com/politics/2023/01/19/fake-twitter-accounts-denying-election-surged-brazil-analysis-finds/. Accessed 10 Dec 2023

Little Black Book (2022) How Lu from Magalu Became the Biggest Virtual Influencer in the World. https://www.lbbonline.com/news/how-lu-from-magalu-became-the-biggest-virtual-influencer-in-the-world. Accessed 02 Feb 2024

Mari A (2022) Brazil stagnant in tech investments and innovation. In: ZDNet. https://www.zdnet.com/article/brazil-stagnant-in-tech-investments-and-innovation/. Accessed 21 Jun 2022

Mari A (2023) Brazil Among Most Optimistic Countries About AI, Study Says. In: Forbes. https://www.forbes.com/sites/angelicamarideoliveira/2023/11/03/brazil-among-most-optimistic-countries-about-ai-study-says/?sh=673b14532daa. Accessed 30 Jan 2024

Mari A (2023) Brazil Is The World's Second Most Vulnerable Country To Cyberattacks. In: Forbes. https://www.forbes.com/sites/angelicamarideoliveira/2023/09/27/brazil-is-the-worlds-second-most-vulnerable-country-to-cyberattacks/?sh=699e7f0a27a4. Accessed 12 Dec 2023

Mari A (2024) Technology Takes Center Stage In Brazil's New Industrial Policy. In: Forbes. https://www.forbes.com/sites/angelicamarideoliveira/2024/01/25/technology-takes-center-stage-in-brazils-new-industrial-policy/?sh=18514b5524d3. Accessed 30 Jan 2024

Martin J-L (2024) First Year of Lula: Overview of the Political Situation in Brazil. IFRI Memos. 11 Jan. https://www.ifri.org/sites/default/files/atoms/files/ifri_martin_brazil_first_year_lula_2024.pdf. Accessed 02 Feb 2024

Pacheco V (2023) 1ª deepfake das eleições mostra números falsos em pesquisa para presidente. In: Showmetech. https://www.showmetech.com.br/deepfake-das-eleicoes-mostra-pesquisa-falsa/. Accessed 13 Dec 2023

Petrarca E (2018) Body Con Job. Miquela Sousa has over 1 million followers on Instagram and was recently hacked by a Trump troll. But she isn't real. https://www.thecut.com/2018/05/lil-miquela-digital-avatar-instagram-influencer.html. Accessed 02 Feb 2024

Pinheiro de Resende S M (2021) The effects of deepfakes on politics and on data justice issues – a perspective from Brazil and the United States. https://arno.uvt.nl/show.cgi?fid=156499. Accessed 13 Dec 2023

Resende F (2023) Ameaça Inquietante: O Uso Malicioso da IA. In: Tribuna entorno. https://www.tribunadoentorno.com.br/2023/06/ameaca-inquietante-o-uso-malicioso-da-ia.html?m=1. Accessed 10 Dec 2023

Reuters (2024) Lula's approval ratings inch up ahead of Brazil's local elections – poll. https://www.reuters.com/world/americas/lulas-approval-ratings-inch-up-ahead-brazils-local-elections-poll-2024-01-23/. Accessed 02 Feb 2024

Schmidt S (2022) Deepfake. In: Pesquina Fapesp. https://revistapesquisa.fapesp.br/en/deepfake/. Accessed 13 Dec 2023

Senado Federal (2020) Projeto de Lei n° 2630, de 2020 (Lei das Fake News). https://www25.senado.leg.br/web/atividade/materias/-/materia/141944. Accessed 02 Feb 2024

Sheena J (2023) Brands are still figuring out virtual influencers. In: Marketing Brew. https://www.marketingbrew.com/stories/2023/09/12/brands-are-still-figuring-out-virtual-influencers. Accessed 02 Feb 2024

Silva C (2023) Electoral Use of AI Rattles. Brazil's Political World. In: The Brazilian Report. https://brazilian.report/power/2023/12/13/electoral-use-of-ai-rattles-political-world/. Accessed 02 Feb 2024

Statista (2023a) Artificial Intelligence – Brazil. https://www.statista.com/outlook/tmo/artificial-intelligence/brazil. Accessed 30 Jan 2024

Statista (2023b) Countries with most Instagram influencers in Latin America as of May 2020. https://www.statista.com/statistics/1126484/countries-most-social-media-influencers-latin-america/. Accessed 02 Feb 2024

Teixeira P S (2023) Brasil é líder global em golpe de link falso no WhatsApp, diz Kaspersky. In: Folha de S.Paulo. https://www1.folha.uol.com.br/tec/2023/03/brasil-e-lider-global-em-golpe-de-link-falso-no-whatsapp-diz-kaspersky.shtml. Accessed 8 Dec 2023

The Spamhaus Project (2022) The Top 10 Worst Botnet Countries. In: Spamhaus.org. https://www.spamhaus.org/statistics/botnet-cc/. Accessed 21 Jun 2022

Thomas D (2020) Deepfakes: A threat to democracy or just a bit of fun? In: BBC News. https://www.bbc.com/news/business-51204954. Accessed 21 Jun 2022

Tocarnia M (2023) TSE debaterá regulamentação da IA para eleições de 2024. In: Agência Brasil. https://agenciabrasil.ebc.com.br/justica/noticia/2023-12/tse-debatera-regulamentacao-da-ia-para-eleicoes-de-2024. Accessed 02 Feb 2024

Ünver H (2023) The role of technology: new methods of information manipulation and disinformation. In: ResearchGate. https://www.researchgate.net/publication/373445537_THE_ROLE_OF_TECHNOLOGY_NEW_METHODS_OF_INFORMATION_MANIPULATION_AND_DISINFORMATION. Accessed 12 Dec 2023

Viana N (2023) Why is Google stonewalling regulation in Brazil? In: The Guardian. https://www.theguardian.com/commentisfree/2023/may/09/us-tech-companies-regulations-brazil. Accessed 02 Feb 2024

Wierson A (2021) Meet Lu, The Non-Human Influencer With 25 Million Followers. In: Observer. https://observer.com/2021/05/meet-lu-the-non-human-influencer-with-25-million-followers/. Accessed 02 Feb 2024

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Королевстве Саудовская Аравия

В. А. Романовский

Стратегическим приоритетом наследного принца М. ибн Салмана является структурная перестройка экономики Королевства. В этих целях приоритет отдается стимулированию инноваций, концентрации ресурсов на НИОКР, развитию высокотехнологичных и наукоемких отраслей с высокой добавленной стоимостью, ускорению интеграции Королевства в глобальное экономическое пространство.

Саудовское руководство создало особую среду, стимулирующую развитие ИИ. Организованное в 2019 г. Управление данных и искусственного интеллекта Саудовской Аравии (Saudi Data and Artificial Intelligence Authority, SDAIA) определяет повестку развития ИИ в стране и ставит целью «продвижение Королевства в элитную лигу экономик, основанных на данных» (OECD 2024). Национальный центр ИИ (National Center for Artificial Intelligence, NCAI), филиал SDAIA, отвечает за инновации в области ИИ, наращивание потенциала и продвижение Национальной стратегии в области данных и ИИ. Эр-Рияд также активно продвигает вопросы этики в передовых технологиях благодаря недавно созданному Международному центру исследований ИИ и этики.

В то же время Королевство не остается в стороне от растущей глобальной тенденции пристального изучения рисков и проблем, связанных с ИИ. Среди лиц, принимающих решения, лидеров отрасли, экспертов и научных кругов Саудовской Аравии все более распространенным становится убеждение в необходимости соблюдения осторожного подхода при использовании огромного преобразующего потенциала ИИ. В этой главе предпринята попытка дать краткое описание области ЗИИИ и проблем информационно-психологической безопасности в Саудовской Аравии через призму трехуровневой модели данных угроз.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Исследования, посвященные аспектам информационно-психологической безопасности в связи с ЗИИИ, определяют «намеренно искаженные интерпретации обстоятельств и последствий развития ИИ в пользу антисоциальных групп» как первый уровень угроз информационно-психологической безопасности (Pashentsev 2023).

Около 70% целей проводимой Королевством реформы «Видение 2030» напрямую связаны со стратегией развития экономики данных и ИИ, цель которой — вывести Саудовскую Аравию в число 15-и ведущих стран в области ИИ к 2030 г. Недавним свидетельством того, что Эр-Рияд продолжает уделять внимание развитию своего потенциала в области ИИ, является индекс «2023 Global AI», согласно которому страна заняла первое место в мире по параметру «государственная стратегия» и 31-е место в целом (Alarabiya 2023). Более того, Саудовская Аравия входит в число мировых лидеров по внедрению технологий ИИ в сферу финансовых услуг. Согласно ежегодному исследованию Finastra «Financial Services: State of the Nation Survey 2023», 55% респондентов из Саудовской Аравии (самый высокий показатель в мире) внедрили или улучшили уже используемые технологические решения на основе ИИ за предшествующий опросу год (Asharq Al-Awsat 2023).

Высокий уровень внимания властей Саудовской Аравии и лично М. ибн Салмана к развитию ИИ в стране, а также политика жесткого регулирования информационной среды, объясняют отсутствие в медиасфере Королевства статистически значимых доказательств, которые могли бы подкрепить предположение о планируемой или продолжающейся информационной кампании по дискредитации усилий Эр-Рияда по развитию своего потенциала в сфере ИИ.

Вместе с тем существует возможность целенаправленной дискредитации в информационном пространстве развития ИИ в Саудовской Аравии, особенно из-за усиления конкуренции между США и Китаем в сфере передовых технологий и ожидаемого стремления Эр-Рияда найти разумный баланс между Вашингтоном и Пекином. Например, в недавней публикации Financial Times обращалось внимание, что саудовско-китайское сотрудничество в области ИИ, в частности, в сфере трансфера технологий, может поставить под угрозу доступ Университета науки и технологий имени короля Абдаллы к передовым чипам американского производства (Kerr et al. 2023). Это важный сигнал для Эр-Рияда, учитывая стремление Королевства стать региональным лидером в разработке ИИ, способным создавать суперкомпьютеры и внедрять большие языковые модели (Ibid.).

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Целенаправленное использование ИИ, например, при кибератаках на критически важную инфраструктуру, представляет собой второй уровень угроз информационно-психологической безопасности в связи с ЗИИИ (Pashentsev 2023).

Действительно, недавние достижения Саудовской Аравии в сфере кибербезопасности свидетельствуют о растущем внимании Эр-Рияда к этому вопросу. Кибербезопасность стала для Эр-Рияда одним из основополагающих элементов стратегии экономического развития (Arabian Business 2022). Ярким примером является последний на момент написания материала рейтинг Глобального индекса кибербезопасности МСЭ за 2020 г., в котором Королевство по уровню обеспечения кибербезопасности заняло первое место в регионе арабских государств и второе место в мировом рейтинге с таким же баллом, как и у Великобритании (ITU 2020). Примечательно, что столь значительных результатов Эр-Рияд добился менее чем за 10 лет (Tsukanov 2024).

Исследование, проведенное Tenable (компанией, работающей в сфере обеспечения кибербезопасности) в декабре 2023 г., в ходе которого были опрошены 50 руководителей в области кибербезопасности и ИТ из Саудовской Аравии, показало, что 40% кибератак на организации в Королевстве за последние два года были успешными (Clewlow 2023). Другое исследование, проведенное в 2023 г. SOCRadar (также работет в сфере кибербезопаности) показывает, что наиболее пострадавшими секторами от целевых кибератак являются розничная торговля, электронная коммерция, информационные услуги, телекоммуникации, финансы, страхование, коммерческие банковские операции, государственное управление, т.е. ключевые отрасли государственной инфраструктуры, с которыми жители страны имеют дело ежедневно (SOCRadar 2023). В том же исследовании подчеркивается, что Королевство, являющееся региональной политической и глобальной экономической державой с одними из крупнейших запасов нефти в мире, «особенно подвержено риску кибератак, нацеленных на критически важную инфраструктуру, такую как нефтяные и газовые месторождения, электростанции и транспортные узлы» (Ibid.). Представляется логичным предположить, что целевые кибератаки с использованием ИИ, даже если они частично успешны, могут представлять угрозу информационно-психологической безопасности жителей отдельного региона, если затронута общественная инфраструктура, или даже всей страны, если пострадала критически важная инфраструктура.

Риски, связанные с использованием технологий ИИ для критической инфраструктуры, уже стоят на повестке дня в Саудовской Аравии. В своем выступлении на Глобальном форуме по кибербезопасности в Эр-Рияде в ноябре 2023 г. генеральный директор Aramco А. Хасан Нассер подчеркнул необходимость выявления рисков и уязвимостей, связанных с генеративным ИИ, «который меняет правила игры для многих отраслей, включая энергетику», и предупредил, что энергетический сектор является мишенью для кибератак с использованием новых технологий (Barakati 2023).

На сегодняшний день, вместе с тем, неочевидно, сможет ли Королевство (или ему будет дано «разрешение») в ближайшем будущем разработать свои собственные системы киберзащиты на базе ИИ для обеспечения безопасности собственной инфраструктуры от целевых кибератак с поддержкой ИИ. В настоящее время Эр-Рияд по-прежнему предпочитает закупать зарубежные готовые пакеты кибербезопасности, а не концентрироваться на разработке собственных решений (Tsukanov 2024).

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

К третьему уровню угроз информационно-психологической безопасности посредством ЗИИИ относятся те, которые прежде всего направлены на причинение ущерба информационно-психологической стабильности или установление контроля над общественным сознанием (Pashentsev 2023).

После подписания Декларации Блетчли 1 ноября 2023 г. внимание Саудовской Аравии к этому уровню угроз возросло. В этом политическом документе, подписанном 28 государствами и Европейским союзом, отмечаются непредвиденные риски, связанные со способностью ИИ манипулировать контентом или генерировать вводящий в заблуждение контент (Gov.uk 2023). Более того, было отмечено, что «передовые системы ИИ могут усиливать такие риски, как дезинформация», и подчеркнули «потенциал серьезного, даже катастрофического вреда, как преднамеренного, так и непреднамеренного, вытекающего из наиболее значительных возможностей этих моделей ИИ» (Ibid.).

На самом деле внимание Саудовской Аравии к этому типу угроз вполне естественно, поскольку фейковые новости, генерируемые ИИ, стали глобальным явлением, и Королевство не остается в стороне от этой тенденции (Jones 2019; Fusco 2022; Sumsub 2023). Итальянский исследователь Федерико Фуско (Fusco 2022) подчеркивает, что «распространенность и влияние фейковых новостей, генерируемых ИИ, в Саудовской Аравии требуют внимания, поскольку они потенциально могут нанести значительный вред отдельным людям и обществу в целом, включая подстрекательство к насилию, распространение ненависти и подрыв доверия к учреждениям».

Другая проблема заключается в растущем спросе на использование арабских чат-ботов на Ближнем Востоке, и в Саудовской Аравии в частности. Несмотря на то, что морфологические особенности арабского языка, его сложность и возможность сбить с толку чат-бота при незначительном изменении одного арабского слова препятствуют широкому использованию чат-ботов, например, в исследовательских институтах Саудовской Аравии (Almurayh 2021), проблема целенаправленного «отравления» данных чат-ботов на арабском языке, скорее всего, проявит себя как высоко вероятностный риск в кратко- и среднесрочной перспективе, если использование чат-ботов на арабском языке станет общей тенденцией в региональных СМИ.

Быстрое распространение фейковых новостей и дезинформации, созданных ИИ, представляет собой серьезную проблему для правительства Саудовской Аравии, которое борется с рас-

пространением такого контента и испытывает нехватку времени для разработки соответствующей политики, обеспечивающей надежность официальных источников новостей для локальной и зарубежной аудитории. На данный момент правительство, похоже, делает выбор в пользу балансирующей политики между запретительными нормами и развитием правовой базы в целях поддержания избранной стратегии развития. Тем не менее, учитывая растущее внимание к рискам и проблемам, связанным с ИИ, представляется естественным ожидать от правительства более строгого регулирования, направленного на медиаконтент, генерируемый ИИ.

Заключение

Несмотря на активную реализацию масштабной стратегии развития ИИ в стране, саудовские лидеры, похоже, уделяют повышенное внимание связанным с ИИ рискам. В отличие от ОАЭ, репутационные активы которых правительство в Абу-Даби считает чрезвычайно важными для поддержания инвестиционной привлекательности сектора высоких технологий, для Саудовской Аравии одним из ключевых приоритетов с точки зрения информационно-психологической сферы национальной безопасности является сохранение репутации государства, способного обеспечить безопасность своих технологических активов. Этот подход направлен в том числе на сохранение эффективности курса руководства Королевства в реализации стратегического плана «Vision 2030».

Технологическая безопасность и безопасность технологических активов, в частности ИИ, имеют важное значение для государственных приоритетов Эр-Рияда в межгосударственных объединениях, в частности, в рамках БРИКС, которые предполагают активизацию двусторонней торговли и использование новых механизмов взаиморасчетов. Последнее приобретает особую важность, учитывая заинтересованность Королевства в достижении лидирующих позиций на региональном рынке финансовых технологий (Al-Baity 2023).

Предварительные результаты исследования показывают, что второй и третий уровни угроз ЗИИИ представляют собой наиболее значительный риск для информационно-психологической безопасности КСА с возможностью возникновения угроз первого уровня. Более полная и детальная модель угроз информационно-психологической безопасности в Саудовской Аравии, связанных с ИИ, может быть создана после дополнительных исследований.

Литература

Alarabiya (2023) السعودية الأولى عالمياً في مؤشر الاستراتيجية الحكومية للذكاء الاصطناعي [Saudi Arabia ranks first in the world in the government strategy index for AI]. https://shorturl.at/pBLS6. Accessed 02 Feb 2024

Al-Baity HH (2023) The Artificial Intelligence Revolution in Digital Finance in Saudi Arabia: A Comprehensive Review and Proposed Framework. Sustainability. Issue *15*(18), 13725. https://doi.org/10.3390/su151813725

Almurayh A (2023) The Challenges of Using Arabic Chatbots in Saudi Universities. IAENG International Journal of Computer Science. https://www.iaeng.org/IJCS/issues_v48/issue_1/IJCS_48_1_21.pdf. Accessed 02 Feb 2024

Arabian Business (2022) Resecurity drives AI-powered cybersecurity in Saudi Arabia with new R&D centre. https://www.arabianbusiness.com/industries/technology/resecurity-drives-ai-powered-cybersecurity-in-saudi-arabia-with-new-rd-centre. Accessed 02 Feb 2024

Asharq Al-Awsat (2023) السعودية تحتل الصدارة بتبني تقنية الذكاء الاصطناعي في الخدمات المالية [Saudi Arabia is at the forefront in adopting artificial intelligence technology in financial services]. https://shorturl.at/fghGW. Accessed 02 Feb 2024

Barakati M (2023) Aramco chief calls ‘innovation’ backed by cybersecurity regime. In: Arab News. https://www.arabnews.com/node/2401461/business-economy. Accessed 02 Feb 2024

Clewlow A (2023) Tenable study reveals 40% of cyberattacks breach Saudi Arabian organisations’ defences. In: Intelligentico. https://www.intelligentcio.com/me/2023/12/13/tenable-study-reveals-40-of-cyberattacks-breach-saudi-arabian-organisations-defences/. Accessed 02 Feb 2024

Fusco F (2022) Artificial Intelligence and Fake News: Criminal Aspect in Pakistan and Saudi Arabia. Pakistan Journal of Criminology. https://faculty.alfaisal.edu/ffusco/publications/artificial-intelligence-and-fake-news%3A-criminal-aspects-in-pakistan-and-saudi-arabia. Accessed 02 Feb 2024

Gov.uk (2023) The Bletchley Declaration by Countries Attending the AI Safety Summit, 102 November 2023. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023. Accessed 02 Feb 2024

ITU (2020) Global Cybersecurity Index 2020. https://www.itu.int/epublications/publication/D-STR-GCI.01-2021-HTM-E. Accessed 02 Feb 2024

Jones M (2019) Propaganda, Fake news, and Fake Trends: The Weaponization of Twitter Bots in the Gulf Crisis. International Journal of Communication. Issue 13. https://ijoc.org/index.php/ijoc/article/viewFile/8994/2604. Accessed 02 Feb 2024

Kerr S, Al-Atrush S, Liu Q, Murgia M (2023) Saudi-China collaboration raises concerns about access to AI chips. In: Financial Times. https://www.ft.com/content/2a636cee-b0d2-45c2-a815-11ca32371763. Accessed 02 Feb 2024

OECD (2024) AI Policies in Saudi Arabia. https://oecd.ai/en/dashboards/countries/SaudiArabia. Accessed 02 Feb 2024

Pashentsev E (2023). General Content and Possible Threat Classifications of the Malicious Use of Artificial Intelligence to Psychological Security. In: Pashentsev E (ed) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-22552-9_2.

Saudi Broadcasting Authority (2023) المملكة تؤكد التزامها بتسخير القوة التحويلية للذكاء الاصطناعي لخير البشرية [The Kingdom confirms its commitment to harness the transformative power of AI for the good of humanity]. https://www.sba.sa/Stories-MainCaption-10435. Accessed 02 Feb 2024

SOCRadar (2023) Saudi Arabia (KSA) Threat Landscape Report 2023. https://socradar.io/saudi-arabia-threat-landscape-report/. Accessed 02 Feb 2024

Sumsub (2023) Identity Fraud Report 2023. https://sumsub.com/newsroom/sumsub-research-global-deepfake-incidents-surge-tenfold-from-2022-to-2023/. Accessed 02 Feb 2024

Tsukanov L (2024) «По обе стороны Персидского залива»: развитие высокотехнологичного бизнеса в регионе и интересы России [“On both sides of the Persian Gulf”: development of high-tech business in the region and Russian interests]. In: PIR Center. https://shorturl.at/ckFJO. Accessed 02 Feb 2024

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Китайской народной республике[11]

Е. Н. Пашенцев, Д. Ю. Базаркина, Е. А. Михалевич, Н.С. Вонг

Введение

Кейс Китая представляет интерес по нескольким причинам. Во-первых, растущее лидерство страны в области активного развития и внедрения ИИ не только породило больше угроз ЗИИИ в отношении Китая, но также создало больше возможностей для противодействия им. Во-вторых, плановый характер китайской экономики – развитая система государственно-частного партнерства, в том числе в области кибербезопасности – и самое большое по численности население в мире, генерирующее колоссальные объемы больших данных для обучения ИИ, делают опыт Китая в борьбе с ЗИИИ поистине уникальным (Bazarkina et al. 2023). На сегодняшний день Китай внес весомый вклад в исследования и разработки в области развития ИИ. Согласно отчету Стэнфордского университета AI Index Report за 2022 г., на Китай приходится 27,6% всех мировых публикаций по итогам конференций по ИИ, в то время как на Соединенные Штаты – 16,9% в 2021 г. (Stanford University Human-Centered Artificial Intelligence 2022). Китай также является лидером по по поданным заявкам на патенты в области ИИ. Так, Китай подал 51,69% от общемирового количества заявок. Однако, пока, получил только около 6% патентов, в то время как доли ЕС, Великобритании и США составляют 3,89% и 16,92% соответственно (Zhang et al 2022). Китай лидирует и по внедрению систем ИИ: 58% компаний внедряют системы ИИ, и 30% рассматривают такую возможность. Для сравнения, в США уровень внедрения ниже: 25% компаний используют ИИ, а 43% изучают потенциал его применения.

В китайском обществе, которое сейчас достаточно инклюзивно, наблюдается консенсус относительно того, что страна по-прежнему должна развивать все новейшие достижения научно-технологического прогресса, включая применение ИИ. Вместе с тем, предполагается, что злонамеренное использование технологий ИИ должно тщательным образом регулироваться в целях защиты безопасности страны и обеспечения соблюдения неприкосновенности частной жизни людей. Китайский народ в основном испытывает веру в собственное правительство и в то, что оно предпримет своевременные и необходимые усилия для предотвращения злонамеренного использования ИИ – это, хочется верить, позволит Китаю опередить многие страны в достижении тонкого баланса одновременно между стимулированием динамичного развития ИИ и обеспечением безопасности и социальной стабильности страны (Xu 2022). Дальнейшее (и весьма вероятное) ухудшение международной обстановки, растущие попытки империалистического давления в контексте перехода на качественно более высокий уровень социально-экономического и научно-технического развития создадут поле для расширения угроз информационно-психологической стабильности, а также всей системе национальной безопасности Китая путем ЗИИИ.

[11] Авторы данной главы выражают свою благодарность за сбор и обработку материалов по теме этого исследования Бо Пенгу, программному директору и научному сотруднику Шанхайского центра стратегических и международных исследований RimPac; Серене Чен, младшему научному сотруднику Шанхайского центра стратегических и международных исследований RimPac; и Никите Тарасову, аспиранту факультета прикладной математики – процессов управления Санкт-Петербургского государственного университета.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Постепенный выход Китая на лидирующие позиции в мире, в том числе и в сфере развития высоких технологий, стал причиной продолжающихся попыток нивелировать и затормозить развитие ИИ в Китае в целом. Так, известный миллиардер Дж. Сорос на Всемирном экономическом форуме в Давосе в 2019 г. резко раскритиковал председателя КНР С. Цзиньпина, заявив, что инклюзивные общества сталкиваются со «смертельной опасностью» со стороны высокотехнологичных авторитарных режимов: «Китай не единственный авторитарный режим в мире, но он, несомненно, самый богатый, сильный и самый развитый в области машинного обучения и ИИ. Это делает С. Цзиньпина самым опасным врагом тех, кто верит в ценности открытого общества» (Watts 2019). Это обвинение вполне можно считать проявлением рисков и угроз информационно-психологической безопасности первого уровня (когда само государство однозначно обвиняется в ЗИИИ в отношении собственных граждан без какой-либо серьезной аргументации).

Эта общая политика по дискредитации развития и внедрения систем ИИ в Китае имеет широкую вариативность в конкретных вопросах. Так, в США и других западных странах появилось множество публикаций, обвиняющих Китай в тотальной слежке и преследовании этнических меньшинств, включай уйгуров, с использованием специальных технологий ИИ (Taddonio 2019; Bhuiyan 2022). Между тем, ИИ в Китае широко применяется в рамках общей системы превентивной деятельности полиции (широко практикуемой в США) для раскрытия прошлых или предотвращения будущих преступлений вне зависимости от национальности преступника (Mantello, 2017). Однако, есть и более “объективные” публикации, которые не фокусируются на уйгурах, а заявляют о тотальной угрозе ИИ в руках «авторитарной (коммунистической и т. д.) диктатуры” (Singman 2023; Kasperowicz 2023; Lanum 2023; Raasch and Sahakian 2023; Hauf 2023). В США тотальная слежка, основанная на технологиях ИИ, развита не меньше, чем в Китае, и их использование вызывает даже бóльшие вопросы. По словам В. Доэллгаст, профессора в области трудовых отношений Корнеллского университета, “работники находятся под постоянным наблюдением, и инструменты мониторинга на основе ИИ могут допускать ошибки, способные привести к несправедливому уменьшению заработной платы или сокращениям. Работники часто не знают, какие инструменты мониторинга используются, какие данные эти инструменты собирают, или как эти данные применяются для оценки их эффективности” (Greenhouse 2023). Однако вместо профессионального обсуждения действительно реальных преимуществ и рисков внедрения высокотехнологичных систем видеонаблюдения (они существуют во всем мире), США выбирают путь пропагандистской войны с целью нанести ущерб как индустрии ИИ в Китае (в совокупности с многочисленными санкциями против китайских компаний в области ИИ), так и интересам социально-экономического развития государства, даже если это приведет к ущербу всему миру.

Угрозы ЗИИИ первого уровня также включают широко распространенные попытки ведущих западных СМИ: 1. подорвать веру в способность Китая разрабатывать технологии ИИ в условиях санкций; 2. убедить китайских разработчиков ИИ в том, что невозможно успешно работать в условиях нахождения у власти Коммунистической партии Китая (КПК); 3. посеять сомнения среди покупателей в качестве/безопасности продуктов ИИ, производимых в Китае и т.д.

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Угрозы информационно-психологической безопасности посредством ЗИИИ второго уровня также получают распространение в Китае. В частности, наблюдается быстрое внедрение технологий ИИ в системы управления (Wang 2023). Многочисленные объекты инфраструктуры, такие

как роботизированные самообучающиеся транспортные системы с централизованным управлением на основе ИИ могут стать отличной мишенью для высокотехнологичных террористических атак. Если, например, злонамеренные акторы захватят контроль над системой управления транспортом крупного города (или другой критически важной инфраструктурой - электростанцией, железнодорожными линиями, телекоммуникационными системами и т.д.), это может привести к значительным деструктивным эффектам - многочисленным авариям и жертвам, панике, созданию негативной информационно-психологической среды, способствующей дальнейшим враждебным действиям (Bazarkina and Pashentsev 2019).

Примерами реализации «пограничных» угроз информационно-психологической безопасности посредством ЗИИИ между первым и вторым уровнями могут являться фишинг и социальная инженерия. Согласно совместному отчету компаний Group-IB и Bridewell за 2023 г., хакерская группа SideWinder использует новую инфраструктуру и возможности для нанесения таргетированных кибератак по целям в Пакистане и Китае. По данным исследователей, хакеры зарегистрировали 55 доменов, имитирующих сервисы различных организаций из таких сфер, как СМИ, государственное управление, телекоммуникация и финансовый сектор. Вышеупомянутые домены, созданные злоумышленниками, имитируют сервисы правительственных организаций в Пакистане, Китае и Индии. Многие из них содержали «документы-ловушки» о деятельности китайского правительства, которые несли вредоносную нагрузку и были предназначены для установки специальной программы на целевое устройство (SecurityLab 2023).

В январе 2020 г. руководитель гонконгского филиала японской компании санкционировал переводы на сумму 35 миллионов долларов на различные банковские счета. Он стал жертвой особо крупного ограбления с применением аудио-дипфейка – злоумышленники, которые в том числе использовали поддельные электронные письма для подтверждения финансовых операций (всего в мошенничестве участвовали 17 человек), подделали голос директора «материнского» бизнеса и убедили директора филиала осуществить перевод денежных средств, так как якобы это было необходимо для проведения операций по поглощению одной компанией другой (Brewster 2022).

Аналогичный случай снова произошел в Гонконге в январе 2024 г., когда финансовый работник транснациональной компании был введен в заблуждение и перевел мошенникам 25 миллионов долларов. Однако в данном случае была применена технология дипфейк, с помощью которой в ходе видеоконференции с обманутым сотрудником мошенники выдали себя за финансового директора компании и прочих ее сотрудников (Chen and Magramo 2024). В интервью службе общественного вещания «Радио и телевидение Гонконга» Б. Ч. Шун-Чинг, исполняющий обязанности суперинтенданта Бюро кибербезопасности, технологий и борьбы с преступностью в Гонконге, заявил: «Я предполагаю, что мошенник заранее загрузил видео, а затем использовал технологии ИИ, чтобы добавить синтезированные голоса для использования в видеоконференции» (Sharma 2024). Все началось с того, что сотрудник получил якобы официальное сообщение от финансового директора. В сообщении содержалось приглашение на конфиденциальный видеозвонок для обсуждения осуществления важных транзакций. Сначала это показалось ему подозрительным, поскольку сообщалось, что необходимо провести секретную транзакцию, но когда он связался со своими коллегами, которые выглядели вполне реалистично, у него не осталось сомнений. Люди, с которыми связался сотрудник, выглядели и говорили в точности как его коллеги. Только после того, как у него состоялся разговор с головным офисом компании, стало понятно, что он стал жертвой высокотехнологичной мошеннической схемы.

Данный прецедент ярко демонстрирует, что мошенники могут использовать технологии ИИ для создания и применения дипфейков в режиме реального времени прямо в ходе видеозвон-

ков, поэтому необходимо быть бдительным даже на встречах с большим количеством участников (Sharma 2024). Этот пример является одним из нескольких недавних эпизодов, в которых мошенники, как предполагается, использовали технологию дипфейков для модификации общедоступных видео и других видеоматериалов с целью выманить у людей деньги посредством обмана. Полиция Гонконга уже произвела шесть арестов в связи с подобными актами мошенничества. По данным полиции, по меньшей мере в 20 случаях фейки использовались для обмана программ распознавания лиц путем имитации людей, изображенных на удостоверениях личности (Chen and Magramo 2024). Фишинг и претекстинг[12] – два разных, но распространенных метода социальной инженерии, используемые вместе, образуют синергетический потенциал и создают более убедительный и опасный продукт манипулирования, основанный на использовании технологий ИИ.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Третий уровень угроз информационно-психологической безопасности, реализуемых посредством ЗИИИ, в Китае в первую очередь характеризуется применением дипфейков, которые считаются здесь одной из самых серьёзных угроз. Дипфейк представляет из себя спектр определенных технологий ИИ, применяемых для создания или изменения видео-, аудио-, фото- контента и машинно-сгенерированных текстов.

В мае 2023 г. Бюро кибербезопасности Министерства общественной безопасности в провинции Ганьсу на Северо-западе Китая раскрыло случай использования технологии ИИ для создания и распространения фейковой информации, задержав подозреваемого в совершении преступления. По данным полиции, он изменял и редактировал собранные новостные материалы, используя популярную нейросеть ChatGPT (что требовало от него обхода «Великого китайского файрвола»[13]), а затем использовал программное обеспечение «Seal Technology» для загрузки своих «новостей» в учетную запись Baijia, которую он приобрел для получения незаконной прибыли. Информация, содержащаяся в сфабрикованной статье «Этим утром поезд сбил дорожного строителя в Ганьсу, в результате чего погибли 9 человек», была заведомо фейковой и не соответствовала действительности. Полиция интернет-безопасности Китая обнаружила, что в общей сложности 21 аккаунт Baidu опубликовал статью, которая за короткий промежуток времени набрала 15 000 просмотров (Gansu Public Security Bureau 2023). Это одно из первых правоприменительных действий, попадающее под принятые законы Китая, регулирующие использование дипфейков.

Созданные ИИ фейковые видеоролики с участием знаменитостей стали обыденным явлением. Однако поддельное видео, распространенное в индонезийских социальных сетях с участием президента Индонезии Джоко Видодо, произносящего речь на беглом китайском, заслуживает особого внимания. Когда президент на видео говорит по-китайски, движения его губ и мимика в основном синхронны, а его голос и интонация также соответствуют реальному голосу. Президент выглядит довольно реалистично, и фоновый звук на видео сопровождается смехом

[12] Претекстинг – форма тактики социальной инженерии, используемая злоумышленниками для получения доступа к информации, системам или сервисам путем создания ложных историй, которые повышают вероятность успеха атаки на жертву.

[13] «Великий китайский файрвол» – система из нескольких десятков инструментов, используемых правительством Китая для блокировки зарубежных сервисов и контроля над внутренними веб-сайтами и приложениями.

аудитории. Недавно правительство Индонезии разъяснило данную ситуацию, указав, что видео было создано с применением технологии дипфейк (Zhang 2023).

Дипфейки – идеальные инструменты для проведения кампаний по дезинформации, поскольку они позволяют генерировать довольно убедительные и реалистичные фейковые новости, для идентификации и разоблачения которых требуется время. Вместе с тем, ущерб, наносимый фейковыми новостями (особенно теми, которые влияют на репутацию людей), часто действует системно, является долгосрочным и необратимым. Все это приводит к отсутствию доверия и нигилизму, когда общество настолько привыкло к дезинформации и постоянному информационному обману, что пытается фильтровать всю получаемую информацию и не доверять даже официальным источникам. В контексте серьезных внутренних и внешних проблем это может представлять возрастающую угрозу социально-политической стабильности в Китае.

Злонамеренное использование дипфейков может стать дополнительным триггером роста угроз международнойбезопасности. Так в апреле 2024 г. на фоне роста напряженности в отношениях Китая и Филиппин из-за разногласий по спорным территориальным вопросам, отдел коммуникаций президента Филиппин Ф. Маркоса-младшего, обратил внимание на получившее широкое распространение редактированное видео или дипфейк, в котором президент Филиппин Фердинанд Маркос-младший, казалось бы, призывает вооруженные силы страны к действиям против “другой нации”. Правительство расследует распространение поддельного видео и возбудит дела против виновных, говорилось в заявлении отдела коммуникаций. Фейки, которые, по-видимому, призывают вооруженные силы действовать против другой страны, с тех пор были удалены, сообщили в отделе, не упомянув Китай (Presidential Communications Office 2024). Данный эпизод возможного применением дипфейка в качестве инструмента обострения отношений Китая с Филиппинами (получившей соответствующее антикитайское освещение в ведущих СМИ США), весьма красноречивое предупреждение о возможности расширения подобной опасной практики в ближайшем будущем в широком спектре национальных и международных конфликтов.

Чат-боты становятся еще одной угрозой информационно-психологической безопасности посредством ЗИИИ. Китайские граждане уже давно стали активными пользователями чат-ботов, и, когда 30 ноября 2022 г. OpenAI (одним из основателей которой является И. Маск) при поддержке Microsoft запустила ChatGPT с более продвинутыми возможностями, это вызвало бурный отклик в Китае. В марте 2023 г. на рынок вышла еще более продвинутая мультимодальная модель GPT4 из семейства языковых моделей GPT (Generative Pre-trained Transformer).

Первый из аналогичных китайских сервисов был представлен поисковой системой Baidu в марте 2023 г. и получил название ERNIE (Enhanced Representation through Knowledge Integration). Он содержит 550 миллиардов различных фактов. По своим возможностям ERNIE близок к нейросети GPT4, представленной OpenAI несколькими днями ранее, а в некоторых отношениях даже превосходит ее. Китайские IT-гиганты Tencent и Alibaba больше сосредоточились на продуктах ИИ для бизнес-партнеров, но оба предлагают чат-ботов широкой китайской публике (Cheng 2023). Системная карта GPT4 признает, что модель даже способна конкурировать с людьми в осуществлении пропаганды во многих областях, особенно в сочетании с человеком-редактором (OpenAI 2023). Новостное издание The China Daily, принадлежащее китайскому правительству, предупредило, что ChatGPT может «усилить пропагандистские кампании, начатые и проводимые США» (Schuman 2023).

Чат-боты уже способны продуцировать реальные угрозы информационно-психологической безопасности и политической стабильности, что объясняет оперативные меры, предпринятые китайским руководством для регулирования их использования. Блокируя ChatGPT и других ботов, поддерживаемых Microsoft, китайские регуляторы оказывают отечественным компаниям

поддержку в разработке соответствующих технологий. В Белой книге о развитии индустрии ИИ в Пекине, опубликованной Пекинским муниципальным бюро экономики и информационных технологий 13 февраля 2023 г., декларируется цель создания прочных основ для развития индустрии ИИ к 2023 г., в том числе путем поддержки ведущих предприятий в создании технологий, подобных ChatGPT (The People's Government of Beijing Municipality 2022). Таким образом, Китай не стоит в стороне от глобального технического прогресса, вместе с тем стремясь предотвратить антиобщественное использование его достижений.

NewsGuard, североамериканская компания, работающая в сфере отслеживания и изучения дезинформации в Интернете, обнаружила, что инструменты ИИ активно используются по всему миру для создания и наполнения так называемых “контент-ферм” (сайты, содержащие, как правило, сомнительную информацию), которые создают огромное количество статей с кликбейтом для оптимизации доходов от рекламы. В апреле 2023 г. NewsGuard выявила 49 веб-сайтов на семи языках - китайском, чешском, английском, французском, португальском, тагальском и тайском – которые, по-видимому, полностью созданы языковыми моделями ИИ, предназначенными для имитации человеческого общения – в этом случае языковые шаблоны замаскированы под типичные новостные сайты (Sadeghi and Arvanitis 2023). То, что обнаружила NewsGuard, скорее всего, является лишь верхушкой айсберга. Мощь все более продвинутых языковых генеративных моделей может превратить такие сайты в эффективное и относительно доступное средство пропаганды.

В настоящее время Китай борется с возрастающим количеством фейковых новостных аккаунтов и постов, созданных ИИ. В середине мая 2023 г. Управление по вопросам киберпротсранства КНР (англ. Cyberspace Administration of China, CAC) заявило, что очистило более 107 000 аккаунтов, публикующих фейковые новости и удалило 835 000 публикаций ложной информации, а также призвало граждан сообщать о фейковых новостных аккаунтах и контенте (Frank 2023). Речь идет не только о текстовых сообщениях, но и о виртуальных ведущих, фейковых студийных сценах, которые позволяют имитировать дизайн существующих зарегистрированных сайтов и, используя различные методы, направленные на получение эмоционального отклика от пользователей сети, увеличивать трафик. Такие действия также могут носить явно злонамеренный характер (Dobberstein 2023). Информационный хаос и неопределенность вследствие спонтанного использования растущих возможностей генеративного ИИ отдельными пользователями и небольшими фирмами, несомненно, применяется для прикрытия крупномасштабных кампаний информационно-психологического воздействия с решающей ролью ЗИИИ в этом процессе, включая злонамеренные действия отдельных крупных государственных и негосударственных субъектов. Из-за острых геополитических и экономических противоречий с другими акторами на мировой арене Китай, очевидно, может стать объективной целью подобных кампаний.

Некоторые угрозы ЗИИИ все еще находятся на ранней стадии своего развития. Например, концепция метавселенной в перспективе может предоставить целый спектр новых возможностей для экономического и социального развития Китая в ближайшем будущем. Китайские IT компании начали «вступать» в эру развития метавселенных, разрабатывая соответсвтующие приложения и инвестируя в технологии, связанные с метавселенной. Метавселенная – это виртуальный мир, существующий одновременоно с физическим миром. Метавселенная делает возможной полноценную конвергенцию цифровой и физической реальностей (в таких общественных сферах, как работа, общение и развлечения), чему способствует обширный спектр передовых технологий (включая ИИ), которые могут сформировать следующую итерацию Интернета. Шесть ведущих технологических гигантов Китая, включая Baidu Inc, Alibaba Group Holding Ltd и Tencent Holdings Ltd (совместно известные как BAT), вошли в число 10 крупнейших компаний мира, подавших наибольшее количество патентов в области разработки критически важных тех-

нологий метавселенной (Interesse 2022). Такие технологии, как дополненная реальность и метавселенная, представят пространство событий в полноценном голографическом и визуальном виде. В метавселенную можно будет погрузиться и взаимодействовать с ней – объектами, контентом и опытом – а аудитория будет более восприимчива к влиянию логики восприятия при распознавании истинности событии. Вместе с тем, это также создает проблемы в виде более высокой угрозы злонамеренного воздействия на массовое сознание, считает эксперт Ч. Цзивэй (Zhang 2022).

Лидируя в области развития цифровых технологий, Китаю предстоит многое сделать с точки зрения обеспечения безопасности в метавселенных, учитывая появление антисоциальных субъектов, которые, несомненно, попытаются использовать возможности это нового пространства в своих собственных интересах. Следует подчеркнуть, что развитие технологий, основанных на ИИ, и возможность их использования в злонамеренных целях антиобщественными субъектами диктуют необходимость более тщательного изучения потенциала этих технологий и разработки системного подхода к задачам их нейтрализации.

Заключение

В Китае происходит постепенный переход от стратегии догоняющего развития к стратегии опережающего развития (не в последнюю очередь в области ИИ), которая предполагает не только появление больших возможностей, но и развитие серьезных рисков для принятия инновационных решений. В сложной и нестабильной среде угрозы информационно-психологической безопасности чрезвычайно велики, особенно если обладание новыми технологиями открывает новые возможности для злонамеренных акторов – начиная от преступных организаций и коррумпированных элементов управленческого аппарата до недружественных агрессивно настроенных государств.

В настоящее время Китай в основном сталкивается со случаями первого (многочисленные попытки дискредитировать китайский ИИ) и третьего (злонамеренное использование дипфейков, чат-ботов, новостных ферм и т.д.) уровней угроз в отношении информационно-психологической безопасности посредством ЗИИИ. Пограничные угрозы (между вторым и третьим уровнями) принимают вид фишинга и социальной инженерии. Угрозы второго уровня в отношении объектов инфраструктуры и систем управления пока не привели к серьезным инцидентам в результате ЗИИИ с соответствующими негативными психологическими и социальными эффектами, но нельзя исключать таких последствий в будущем из-за деятельности «высокотехнологичных» злоумышленников, действующих как внутри государства, так и из-за рубежа. Развитие метавселенных, совершенствование эмоционального ИИ, прогресс в области развития общего ИИ, гуманитарных наук и их прикладного применения, естественно, поставят Китай перед лицом еще более сложных вызовов в области информационно-психологической безопасности.

ЗИИИ сегодня стало значительным, но пока не основным источником угроз для развития ИИ в Китае, уступающим по эффективности общеизвестному и изученному набору форм и методов традиционной пропаганды (хотя сегодня они все реже могут обойтись без применения ИИ в качестве вспомогательного инструмента). Однако в ближайшем будущем, в связи с количественным и качественным ростом возможностей ИИ и его дальнейшим внедрением в различные сферы общественной жизни, ЗИИИ может выйти на первый план.

Все вышесказанное требует диалектического подхода к оценке роли ИИ. Адекватное использование технологий ИИ дает мощный импульс всему общественному развитию, что наглядно демонстрирует современный Китай. Однако с развитием этих технологий растут и

угрозы ЗИИИ, иногда даже опережая возможности полезного применения алгоритмов. Нейтрализация негативных форм применения ИИ требует не столько технологических или административных мер, сколько социальных. Уникальные технологии, которые направлены на частичную или полную замену человека в различных формах познавательной деятельности и эмоционально-психологической активности (например, эмоциональный ИИ при работе с инвалидами), монотонной или вредной физической работе (ИИ-роботы), усилят вызовы, вставшие перед человечеством, если последнее направит свои растущие возможности в области развития ИИ не на построение более гармоничной социальной системы и развитие личности с качественно новым уровнем интеллектуальных и психофизических возможностей, более высоким уровнем социальной ответственности, а для усиления политической, социальной, национальной или расовой поляризации. Китай обладает огромным научным, техническим, экономическим и, самое главное, человеческим потенциалом для предотвращения негативных сценариев вследствие использования ИИ в интересах не только своего национального развития, но и в целях прогресса всего человечества. Вопрос в том, насколько эффективно он будет использовать этот потенциал. Данный вопрос остается открытым, а более глубокое осознание основных возможностей и рисков применения ИИ придет в будущем.

Литература

Bazarkina D, Mikhalevich E, Pashentsev E, Matyashova D (2023) The Threats and Current Practices of Malicious Use of Artificial Intelligence in Psychological Security in China. In: Pashentsev E (ed) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-22552-9_13

Bazarkina D, Pashentsev E (2019) Artificial Intelligence and New Threats to International Psychological Security. Russia in Global Affairs. doi: 10.31278/1810-6374-2019-17-1-147-170

Bazarkina D, Pashentsev E (2020) Malicious Use of Artificial Intelligence: New Psychological Security Risks in BRICS Countries. *Russia in Global Affairs*. doi: 10.31278/1810-6374-2020-18-4-154-177

Bhuiyan J (2022) ‘There’s cameras everywhere’: testimonies detail far-reaching surveillance of Uyghurs in China. In: The Guardian. https://www.theguardian.com/world/2021/sep/30/uyghur-tribunal-testimony-surveillance-china. Accessed 02 Jan 2024

Chen H, Magramo K (2024) Finance worker pays out $25 million after video call with deepfake ‘chief financial officer’. In: CNN. https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html. Accessed 06 Feb 2024

Cheng E (2023) Baidu says its ChatGPT rival Ernie bot now has more than 100 million users. In: CNBC. https://www.cnbc.com/2023/12/29/baidu-says-its-chatgpt-rival-ernie-bot-has-more-than-100-million-users.html. Accessed 02 Jan 2024

Dobberstein L (2023) China cracks down on AI-generated news anchors. In: The Register. https://www.theregister.com/2023/05/16/china_crackdown_on_ai_generated_news/. Accessed 02 Jan 2024

Frank J (2023) China is Deleting Hundreds of Thousands of AI-Generated News Accounts and Postshttps. In: Business2Community. www.business2community.com/tech-news/china-is-deleting-hundreds-of-thousands-of-ai-generated-news-accounts-and-posts-02692962. Accessed 02 Jan 2024

Gansu Public Security Bureau (2023) Gansu public security cracked the first case of using AI artificial intelligence technology to concoct false information [Gansu gong'an zhenpo shou li liyong AI rengong zhineng jishu paozhi xujia xinxi an]. https://mp.weixin.qq.com/s/_Wfe-EV13O6uBM65jZDzdg. Accessed 02 Jan 2024

Greenhouse S (2024) 'Constantly monitored': the pushback against AI surveillance at work. In: The Guardian. https://www.theguardian.com/technology/2024/jan/07/artificial-intelligence-surveillance-workers. Accessed 02 Jan 2024

Hauf P (2023) China aiming for 'chaos and confusion' by weaponizing AI, warns GOP senator. In: Fox News. https://www.foxnews.com/politics/china-aiming-chaos-confusion-weaponizing-ai-warns-gop-senator. Accessed 02 Jan 2024

Interesse G (2022) China's Debut in the Metaverse: Trends to Watch (Updated). In: China Briefing. https://www.china-briefing.com/news/metaverse-in-china-trends/. Accessed 02 Jan 2024

Kasperowicz P (2023) China using tech to 'oppress its own people,' warns lawmaker looking to restrict AI exports. In: Fox News. https://www.foxnews.com/politics/china-using-tech-oppress-own-people-warns-lawmaker-restrict-ai-exports. Accessed 02 Jan 2024

Lanum N (2023) McCaul says China's AI, quantum investments are a race for military and economic 'domination of the world.' In: Fox News. https://www.foxnews.com/media/mccaul-china-ai-quantum-investments-race-military-economic-domination-world. Accessed 02 Jan 2024

Mantello P (2016) The Machine that Ate Bad People: The ontopolitics of the pre-crime assemblage. Big Data and Society, July-December, Vol.3-2 p.1-22.

OpenAI (2023) GPT-4 System Card. https://cdn.openai.com/papers/gpt-4-system-card.pdf. Accessed 02 Jan 2024

Presidential Communications Office (2024) Gov't vows legal action vs deep fake video creators, spreaders. https://pco.gov.ph/news_releases/govt-vows-legal-action-vs-deep-fake-video-creators-spreaders/. Accessed 02 Jun 2024

Raasch JM, Sahakian T (2023) AI's threat to humanity will be far greater if China masters it first: Gordon Chang. In: Fox News. https://www.foxnews.com/world/ai-threat-humanity-far-greater-china-masters-first-gordon-chang. Accessed 02 Jan 2024

Sadeghi M, Arvanitis L (2023) Rise of the Newsbots: AI-Generated News Websites Proliferating Online. In: NewsGuard. https://www.newsguardtech.com/special-reports/newsbots-ai-generated-news-websites-proliferating/. Accessed 02 Jan 2024

Schuman M (2023) Why Chatbot AI Is a Problem for China. In: The Atlantic. https://www.theatlantic.com/international/archive/2023/04/chatbot-ai-problem-china/673754/. Accessed 02 Jan 2024

SecurityLab (2023) SideWinder militantly masquerades as Pakistani and Chinese government agencies in their latest attacks. https://www.securitylab.ru/news/538242.php. Accessed 02 Jan 2024

Sharma S (2024) $25 million swindled as deep fake CFO tricks finance worker. In: Interesting Engineering. https://interestingengineering.com/culture/25-million-swindled-as-deep-fake-cfo-tricks-finance-worker. Accessed 06 Feb 2024

Singman B (2023) US intel community warns of 'complex' threats from China, Russia, North Korea. In: Fox News. https://www.foxnews.com/politics/us-intel-community-warns-complex-threats-china-russia-north-korea. Accessed 02 Jan 2024

Stanford University Human-Centered Artificial Intelligence (2022) Artificial Intelligence Index Report 2022. https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf. Accessed 02 Jan 2024

Taddonio P (2019) How China's Government Is Using AI on Its Uighur Muslim Population. In: PBS. https://www.pbs.org/wgbh/frontline/article/how-chinas-government-is-using-ai-on-its-uighur-muslim-population/. Accessed 02 Jan 2024

The People's Government of Beijing Municipality (2023) White Paper on the Development of Beijing Artificial Intelligence Industry in 2022 is released ["2022 Nian beijing rengong zhineng chanye fazhan baipishu" zhong bang fabu]. http://www.beijing.gov.cn/ywdt/gzdt/202302/t20230214_2916514.html . Accessed 02 Jan 2024

Wang D (2022) Threats and Countermeasures of Malicious Use of Artificial Intelligence [Rengong zhineng eyi shiyong weixie yu yingdui]. https://www.cnki.com.cn/Article/CJFDTOTAL-CINS201908008.htm. Accessed 02 Jan 2024

Watts W (2019) Soros blasts China's Xi as 'most dangerous opponent' of open societies. In: MarketWatch. https://www.marketwatch.com/story/george-soros-blasts-chinas-xi-as-most-dangerous-opponent-of-open-societies-2019-01-24?siteid=yhoof2&yptr=yahoo. Accessed 02 Jan 2024

Veldkamp D (2022) Cyber Awareness 2022: Consider Deepfakes, NFTs, and More. In: InfoSystems. https://infosystems.biz/cybersecurity/cyber-awareness-2022-consider-deepfakes-nfts-and-more/. Accessed 6 Jul 2022

Xu C (2022) Artificial Intelligence and National Political Security [Ren gong zhi neng yu guo jia zheng zhi an quan]. http://finance.people.com.cn/n1/2022/0626/c1004-32456635.html. Accessed 02 Jan 2024

Zhang D, Maslej N, Brynjolfsson E, Etchemendy J, Lyons T, Manyika J, Ngo H, Niebles JC, Sellitto M, Sakhaee E, Shoham Y, Clark J, Perrault R (2022) The AI Index 2022 Annual Report. https://doi.org/10.48550/arXiv.2205.03468.

Zhang J (2023) "Artificial Intelligence Replacing" Hidden Risks, How to Supervise the Abuse of AI Technology ["Rengong zhineng zui ti" ancang fengxian,AI jishu lanyong gai ruhe jianguan]. https://www.163.com/dy/article/IJ7G5SFT0514R9L4.html. Accessed 02 Jan 2024

Zhang Z (2022) Cognitive Domain Operations from the Perspective of Intelligence: Emotional Conflict Becomes a Prominent Attribute of Cognitive Domain Operations [Zhineng hua shi yu xia de ren zhi yu zuozhan: Qinggan chongtu chengwei ren zhi yu zuozhan tuchu shuxing]. http://www.81.cn/yw_208727/10204158.html. Accessed 02 Jan 2024

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Республике Индия[14]

Е. Н. Пашенцев, Д. Ю. Базаркина

Введение

По прогнозам, объем рынка ИИ в Индии значительно увеличится и, по оценкам, достигнет $6 358,8 млн к 2025 г. (Srivastava 2023). В феврале 2018 г. правительство Индии объявило, что Национальный институт трансформации Индии (National Institution for Transforming India, NITI Aayog), правительственный аналитический центр, возглавит национальную исследовательскую программу в области ИИ. В 2017 г. Министерство торговли и промышленности Индии сформировало Целевую группу по ИИ с целью проведения экономических преобразований в Индии (Faggella 2019). Таким образом, ИИ способствует глобальным изменениям в индийской экономике. Индийские стартапы разрабатывают решения с использованием технологий ИИ в сфере образования, здравоохранения и финансовых услуг. Развитие ИИ в Индии также включает в себя работу над цифровыми помощниками (которые позволяют организациям выстраивать эффективную коммуникацию с клиентами), системами принятия решений на основе ИИ, а также подразумевает использование ИИ и блокчейна в торговле (Chakraborty 2022). Ожидается, что к 2025 г. внутренний валовой продукт Индии вследствие внедрения ИИ в экономику увеличится до $500 млрд, а к 2035 г. – до $967 млрд. Инвестиции Индии в ИИ ежегодно увеличиваются на 30,8% (по состоянию на июнь 2023 г.). “Глобальные тенденции прагматизма, развитие моделей с открытым исходным кодом, ориентация на регулирование, трансформация рабочих мест и повышение эффективности создают предпосылки для преобразующих изменений в ИИ. В то же время специфический рост Индии в секторах информационных технологий и ИИ придает этому направлению уникальное измерение” (Srivastava 2023). Конечно, такое стремительное развитие сферы ИИ в Индии продуцирует как преимущества (в том числе для индийской экономики), так и может привести к негативным последствиям для информационно-психологической стабильности в стране.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

В настоящее время одной из актуальных угроз первого уровня ЗИИИ в Индии остается "игра" на страхах экономически активного населения Индии, касающаяся потери работы по причине автоматизации. Компания Microsoft опубликовала в Индии результаты своего флагманского отчета "Work Trend Index 2023", в котором говорится, что около 74% индийских работников опасаются, что ИИ заменит их на рабочих местах. Новый отчет посвящен теме “Сможет ли ИИ изменить работу?”, так как проблема автоматизации рабочих мест постепенно становится все более актуальной вследствие внедрения таких ИИ-инструментов, как ChatGPT, Google Bard и Microsoft Bing Chat. Более 70% индийцев опасаются сокращения рабочих мест. 78% индийских работников говорят, что в настоящее время у них нет необходимых квалификации и умений для выполнения своей работы (Sengupta 2023). Это предоставляет злоумышленникам новые воз-

[14]Авторы настоящей главы выражают благодарность за ценные комментарии соучредителю Digital India Foundation Арвинду Гупте, а также научному сотруднику Digital India Foundation Аакашу Гуглани.

можности для манипулирования общественным мнением и позволит им критиковать и намеренно дискредитировать тех политиков, которые будут активно продвигать внедрение систем ИИ в различные сферы общественной жизни.

В социальных сетях уже велись жаркие дискуссии по вопросу сокращения рабочих мест из-за внедрения технологий ИИ. Генеральный директор одной из индийских компаний подвергся критике после того, как заявил, что его компания заменила 90% своего вспомогательного персонала чат-ботом на основе ИИ: Суумит Шах, основатель компании Dukaan, осуществляющей свою деятельность в области цифровой торговли, опубликовал пост в социальной сети Twitter (в настоящее время носит название “X”)[15], согласно которому чат-бот значительно сократил время первого ответа и более эффективно решал вопросы клиентов. Данный твит привел к волне негодований и возмущений в сети. Рассматриваемый инцидент произошел как раз в то время, когда развернулись наиболее активные обсуждения вопроса, сокращает ли ИИ количество рабочих мест (особенно в сфере услуг), и наблюдались возросшие опасения по этому вопросу (BBC 2023). Подобные примеры наглядно демонстрируют, что в индийском обществе существуют серьезные опасения по вопросу социальных последствий автоматизации. Рост рекламных ожиданий в отношении ИИ также является проблемой, но в меньшей степени, чем всего несколько лет назад, поскольку опасения и страхи по поводу внедрения и развития ИИ начинают усугубляться.

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Среди угроз второго уровня, актуальных для Индии, стоит выделить участившиеся случаи мошенничества в виртуальном пространстве с использованием технологий ИИ. Основным аспектом такого мошенничества является социальная инженерия, которая позволяет злоумышленникам при помощи таргетированного воздействия завоевать доверие жертвы, дезориентировать её и заставить действовать в своих интересах. Например, авторы фишинговых сообщений используют следующие психологические приемы:

- *Срочность*: фишинговое электронное письмо обычно требует, чтобы жертва осуществила какие-либо действия в ближайшее время, поскольку чем дольше человек будет думать, тем больше у него возникает сомнений в правдивости содержащейся в письме информации и законности требуемых действий;

- *Правдоподобность*: попытки фишинга будут основаны на реальных сценариях, согласно которым жертве необходимо оплатить счет или поделиться файлом;

- *Авторитетность*: наблюдается заметный рост числа случаев скрытого фишинга, когда атака, по крайней мере частично, нацелена на конкретного человека – часто утверждается, что она исходит от авторитетной личности, такой как генеральный директор компании или начальник службы безопасности;

- *Конфиденциальность*: требуемое действие относится конкретно к целевой персоне и должно выполняться ей полностью самостоятельно, поскольку вовлечение в этот процесс кого-либо еще увеличивает шансы обнаружения факта мошенничества (Egress 2021 г.).

[15] Социальная сеть X (Twitter), заблокирована на территории Российской Федерации по решению Роскомнадзора.

ИИ может значительно повысить “эффективность” фишинговых атак, что неизбежно скажется на статистике преступлений. Например, недавний отчет компании Group-IB, чья деятельность связана с обеспечением кибербезопасности, показал, что около 100 000 учетных записей ChatGPT были скомпрометированы, а их данные незаконно продаются через даркнет. При этом только Индия сообщила о 12 632 украденных учетных данных (по состоянию на 2023 г.). По мере того как ИИ становится всё дешевле и доступнее, риски кибербезопасности, связанные с этой технологией, будут возрастать. Согласно данным компании CheckPoint Research (также работающей в области обеспечения киберебзопасности), в первом квартале 2023 г. среднее количество еженедельных атак в Индии увеличились на 18% по сравнению с соответствующим периодом 2022 г., при этом каждая организация сталкивалась как минимум с 2 108 еженедельными атаками. Согласно отчету о состоянии безопасности приложений от компании Indusface (предоставляет программное обеспечение для обеспечения безопасности приложений), во втором квартале 2023 г. количество атак с применением ботов выросло на 48% по сравнению с первым, и девять из десяти веб-сайтов столкнулись с подобными воздействиями (Stanly 2023). Отрасль стремится защититься от таких угроз, разрабатывая различные решения для аудита качества программного обеспечения, которые могут выявлять и исправлять уязвимости. Однако злоумышленники обладают собственными способами эксплуатации таких технологий в злонамеренных целях. Например, Metasploit – проект в области обеспечения компьютерной безопасности, который предоставляет информацию об уязвимостях системы и как раз является одним из инструментов, ставшим популярным среди хакеров (Stanly 2023).

Мошенники все чаще используют технологическую продвинутость систем ИИ, чтобы совершать атаки на миллионы индийцев, которые тратят до 105 минут в неделю на просмотр, проверку или принятие решения о том, является ли сообщение, отправленное через мессенджеры, электронную почту или социальные сети, реальным или поддельным.

Согласно исследованию компании Mcafee (Doval 2023), 82% индийцев открывали поддельные сообщения или попадались на описанные в них мошеннические схемы. Среди изощренных способов обмана наиболее распространены такие эффективные формы воздействия на пользователей, которые включают в себя поддельные уведомления о приеме на работу или различные деловые предложения (64%), а также банковские оповещения (52%). “ИИ - излюбленный инструмент мошенников, помогающий киберпреступникам масштабировать фишинг и делать мошеннические сообщения, содержащие текстовую информацию, более изощренными. Скорость, с которой осуществляется мошеннические и фишинговые воздействия, содержащие текстовую информацию, растет – каждые 11 секунд создается новый фишинговый сайт. Это подчеркивает острую потребность в решениях в области обеспечения безопасности, способных изменить ситуацию в противостоянии с мошенниками, использующими технологии ИИ. Для 900 млн. интернет-пользователей страны вопрос об обеспечении собственной безопасности в Интернете никогда не был столь актуальным” (Doval 2023).

Онлайн-мошенничество имеет свои последствия психологического характера. Об этом свидетельствует возросший уровень стресса, с которым сталкиваются люди из-за увеличения количества мошеннических сообщений с применением технологий ИИ и ростом их профессионализма. Этот стресс может спровоцировать появление социальных потрясений, поскольку доверие к институтам, обеспечивающим безопасность, на фоне роста преступности, использующей для своих злонамеренных действий алгоритмы, снижается.

Киберпреступники теперь не боятся осуществлять свою злонамеренную деятельность с помощью таких сложных генеративных нейросетей, как FraudGPT и WormGPT, которые получили свою популярность в качестве “ботов без ограничений, правил и границ”. Эти чат-боты, недавно

появившиеся в даркнете, доступны всем, кто хочет создавать фишинговые электронные письма, вредоносное ПО или инструменты для взлома. Проблема использования подобных чат-ботов актуальна и в Индии. М. Тхакар, вице-президент по IT в компании Hitachi Hi-Rel, предупреждает: “Эти чат-боты основаны на популярной технологии ChatGPT-3, которая может генерировать реалистичные и связные тексты исходя из пользовательских запросов. С помощью этих инструментов хакеры могут создавать фишинговые электронные письма, чтобы обмануть ничего не подозревающих жертв, заставив их поверить, что они получили официальное деловое письмо, SMS-сообщение, или уведомление банка” (The Times of India 2023a). М. Тхакар продемонстрировал, что нейросеть FraudGPT[16] способна писать вредоносный код, создавать вирусы-невидимки или необнаруживаемое вредоносное ПО, находить ячейки, отличные от протокола VBV (Verified by Visa)[17], а также создавать фишинговые страницы и хакерские инструменты для проникновения в группы, сайты и онлайн-магазины. Программа даже может создавать мошеннические страницы или письма, находить утечки, уязвимости и получать доступ к активным банковским картам. Сотрудники уголовного розыска Гуджарата заявили, что поставщики мошеннического ПО довольно известны на подпольных веб-площадках, таких как Empire, WHM, Torrez, Alphabay и Versus (The Times of India 2023a). В будущем количество и совершенствование таких вредоносных чат-ботов, вероятно, увеличится, что может сделать методы социальной инженерии намного более деструктивными.

Наряду с преимуществами цифровизации индийской экономики, также увеличиваются и риски ЗИИИ в целях подрыва информационно-психологической безопасности. В 2022 г. в Индии планировалось введение цифровой валюты центрального банка (Central bank digital currency, CBDC) – цифровой рупии, которая основана на блокчейне. Хотя ряд политиков и финансистов выступают за усиление законодательного регулирования криптовалют, глава Резервного банка Индии Шактиканта Дас считает, что цифровая рупия также не является абсолютно безопасной и может быть подвержена проведению с ней цифровых мошеннических операций.

Несколько мошеннических инвестиционных порталов уже более года находятся в поле зрения индийских правоохранительных органов. Многие жители штата Керала были обмануты посредством несуществующей криптовалюты под» под названием Morris Coin, при этом правоохранительные органы оценили сумму мошеннических активов в $200 млн. Аналогичный инцидент произошел в штате Карнатака. Эксперты полагают, что сочетание мошенничества и злонамеренного использования дипфейков в сфере криптовалют выведет фишинговые воздействия на совершенно новый уровень (Biswas 2022), что позволит более эффективно убеждать пользователей в подлинности фейкового сайта, принимающего платежи в криптовалюте.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Уровень ЗИИИ в сфере голосового управления продолжает расти. Pindrop (компания, работающая в сфере обеспечения информационной безопасности) указывает, что мошенничество в этой области выросло более чем на 350% с 2013 по 2017 гг. (Pindrop 2020). По крайней мере, 39% населения Индии, имеющего доступ в Интернет, вероятно, пользуется какой-либо цифровым го-

[16] Генеративный ИИ злонамеренного назначения, который использует алгоритмы машинного обучения для создания дезинформации и обманного контента.

[17] Протокол Verified by Visa (VbV) добавляет дополнительный шаг авторизации пользователя при оплате покупки в интернет-магазине.

лосовым помощником. Эксперты прогнозируют, что в ближайшем будущем голосовая аутентификация станет предпочтительным методом транзакций для электронной коммерции, банковского дела и платежей (Bhatt 2018). В контексте возникновения будущих угроз в целях проведения междисциплинарных исследований целесообразным представляется рассмотреть риски не только взлома голосовых интерфейсов или чат-ботов, но и программ машинного перевода. Например, искажение содержания официальных документов посредством ЗИИИ может иметь такой же провокационный и дестабилизирующий характер, как и публикация дипфейка, способного спровоцировать конфликт между странами.

В Индии во время избирательных кампаний 2019 г. была зафиксирована волна активности ботов: с 9 по 10 февраля были массово развернуты бот-аккаунты в Twitter (ныне X), публиковавшие хэштеги как за, так и против действующего премьер-министра Нарендры Моди. В то же время, небольшие группы аккаунтов публиковали тысячи твитов в час. Основные политические партии расширили возможности цифровой коммуникации по сравнению с предыдущими выборами в 2014 г., но в 2019 г. влияние таких кампаний не отличалось масштабностью (Thaker 2019) по причине относительно небольшого числа пользователей Twitter среди избирателей. Однако массовое использование ботов, которые были задействованы в политических кампаниях, снизило качество онлайн-дебатов. 27 декабря 2021 г. известный лидер оппозиции направил письмо генеральному директору Twitter с жалобой на то, что, в отличие от других политических лидеров, число его подписчиков не увеличивалось с июля. На это Twitter заявил, что они удаляют миллионы ботов и вредоносных учетных записей с помощью инструментов машинного обучения (OpIndia 2022). Таким образом, активность ботов стала предметом активных политических споров, которые могли использоваться антисоциальными субъектами, что подрывало информационно-психологическую безопасность.

Помимо политических кампаний, чат-боты могут быть задействованы в религиозной деятельности, где они также способны представлять угрозу посредством ЗИИИИ третьего уровня. Опыт Индии демонстрирует рост использования чат-ботов для комментирования религиозных текстов, и эти комментарии (установить достоверность которых не представляется возможным) могут носить весьма провокационный характер. Многие люди в Индии отказываются от личного контакта с духовным гуру, интерпретирующим Бхагавадгиту, и обращаются к онлайн-чат-ботам, которые от имени индуистского бога Кришны дают ответы на животрепещущие вопросы о смысле жизни. По мнению экспертов, такая форма использования чат-ботов на основе ИИ-алгоритмов может нести угрозу, так как имеет способность действовать не по сценарию и потворствовать насилию. Некоторые боты постоянно дают ответ, что убивать кого-либо – это нормально, если это ваша дхарма или долг. Как заявляет Л. Юсуф, юрист из Мумбаи и соавтор книги об ИИ (Yusuf 2020), “это недопонимание и дезинформация, основанные на религиозных текстах”. “Текст имеет большую философскую ценность, а что делает бот? Бот дает вам буквальный ответ, и в этом опасность” (Shivji 2023). По меньшей мере пять чат-ботов Gita, созданных на базе языковой модели Generative Pre-trained Transformer 3 (GPT-3), появились в Сети в начале 2023 г. Они используют алгоритмы, имитирующие общение, и генерируют ответы на основе статистических вероятностных моделей. Согласно новостным сайтам, эти боты имеют миллионы пользователей. По словам Л. Юсуф, потенциальная опасность ответов таких чат-ботов состоит в том, что они оправдывают насилие, и эта проблема стоит более остро в такой стране, как Индия, где религия имеет огромное эмоциональное воздействие (Shivji 2023).

7 января 2023 г. активист М. Викрам Хегде в своем аккаунте в Твиттере опубликовал скриншот из ChatGPT для собственной аудитории, насчитывающей более чем 185 000 подписчиков. Представляется, что в посте продемонстрирован чат-бот на базе ИИ, способный генерировать

шутки в отношении Кришны. При этом чат-бот обучен быть осмотрительным, и, когда ему делают запрос пошутить о Христе или пророке Мухаммеде, ему следует отвечать “Прошу прощения, но я не запрограммирован шутить на религиозные темы или в отношении какого-либо из Богов”. Это ограничение, по-видимому, не распространяется на богов индуистского пантеона. OpenAI, которой принадлежит ChatGPT, не стала комментировать данную ситуацию. Твит широко распространился в индийских социальных сетях, был просмотрен более 400 000 раз и вызвал большой ажиотаж среди пользователей. В течение нескольких дней это превратилось в теорию заговора в социальных сетях и вещательных медиаканалах. Данный прецедент, когда неосмотрительность и неосторожность разработчиков ChatGPT привела к конфликту, можно рассматривать одновременно как проявление угроз первого уровня, при котором происходила дискредитация продуктов ИИ с привлечением лидеров общественного мнения, так и третьего уровня, когда сам чат-бот был использован для нагнетания социальной напряженности.

Однако наиболее показательными примерами ЗИИИ в целях подрыва информационно-психологической безопасности в условиях Индии являются дипфейки. Согласно отчету компании Deeptrace, в 2019 г. около 3% веб-сайтов в мире, содержащих дипфейковые видео порнографического содержания, были индийскими (Ajder et al. 2019, p. 2). Deeptrace также отмечает, что «значительный вклад в создание и использование синтетических медиа-инструментов внесли интернет-пользователи из Китая» (Ajder et al. 2019, foreword). Уже известны случаи использования дипфейков в Индии для нанесения репутационного ущерба. Фотографии и видео индийской журналистки были использованы для создания порнографического дипфейка с ее участием (Ajder 2019), что свидетельствует о принятии на вооружение технологий ИИ современными преступниками и об опасности развязывания широкомасштабных кампаний по дискриминации различных групп интересов в будущем.

Печальную известность в Индии приобрел прецедент использования дипфейков в ходе предвыборной кампании (который, возможно, допустимо считать первым подобным случаем в мире). В феврале 2020 г. Накануне выборов в Законодательное собрание Бхаратия Джаната Парти (БДП, Индийская народная партия) применила технологию дипфейк для создания двух видеороликов, в которых лидер партии М. Тивари обращается к избирателям на двух языках – английском и хинди (диалект хариани). Целью данной инициативы было донести послание соответственно до двух групп избирателей, говорящих на разных языках. Видеоролики, по словам представителей партии, были распространены среди примерно 5800 чатов в WhatsApp, и охватили около 15 млн. человек (Alavi and Achom 2020). В этих роликах Тивари также поздравил своих сторонников с принятием парламентом Индии поправок к Закону о гражданстве. В оригинальном видео послание Тивари транслируется на хинди, при этом выражение его лица, мимика и движения губ имитируют язык, смоделированный с помощью ИИ для диалекта хариани. Пресс-секретарь партии сообщил, что видео на диалекте хариани было воспринято позитивно, после чего и было решено создать англоязычную версию. Однако стало ясно, что события могут выйти из-под контроля: «Кто-то использовал видео нашего лидера... Маноджа Тивари в Facebook и прислал ролик с измененным содержанием на диалекте хариани», — сказал представитель по связям с общественностью делийского отделения БДП. «Для нас это стало шоком, поскольку оппозиция могла использовать это видео в злонамеренных целях… Мы решительно осуждаем использование данной технологии, которая находится в открытом доступе и использовалась без нашего согласия» (Mihindukulasuriya 2020).

Хотя в данном случае дипфейки использовались исключительно для преодоления языкового барьера и охвата более широкой аудитории, дальнейшая его модификация в злонамеренных целях — а затем и ответная реакция со стороны представителей БДП — вызвали оживленную

дискуссию в индийских СМИ. Очевидно, что это укрепляет опасения по поводу возможного использования дипфейков для распространения дезинформации в таких областях, как политические процессы, где данная технология позволит вкладывать противоречивые и неприемлемые заявления в речи политических оппонентов (Alavi and Achom 2020). Для Индии, как и для многих других стран, в нынешних условиях стремительной цифровизации и усиления кампаний по дезинформации может стать актуальной задача комбинации технологий ЗИИИ с различными угрозами информационно-психологической безопасности на трех уровнях.

В ноябре 2023 г. правительство Индии опубликовало предупреждение об “опасных и разрушительных” последствиях технологий ИИ после того, как фейковое видео с участием известной болливудской актрисы Р. Манданна, на котором она выходит из лифта, одетая в черную одежду, получило широкую популярность. Однако данный материал оказался дипфейком. Женщиной в шестисекундном видео на самом деле была инфлюенсер британо-индийского происхождения З. Патил, которая опубликовала оригинальный клип в своем Instagram-аккаунте. Журналист А. Кумар, сотрудник Alt News (индийское СМИ, занимающееся проверкой фактов в публикациях других СМИ), первым сообщил, что вирусное видео, на котором, по-видимому, изображена Манданна, было дипфейком (Bandara 2023). Звезда Болливуда Манданна выразила свое беспокойство из-за фейкового видео, распространяемого в Интернете, и заявила, что проблему создания и распространения технологий дипфейков необходимо решать “в срочном порядке”. Ветеран Болливуда, актер А. Баччан утверждает, что у Манданны “имелись веские основания для судебного иска” по поводу кражи личных данных, которые использовались в видео, сгенерированном ИИ (Bandara 2023). Фальшивые изображения индийских актрис К. Каиф и К. Девган также появлялись в стремительном порядке одно за другим, что указывает на тревожный факт распространения дипфейков.

Согласно заявлению Р. Чандрасекара, государственного министра по электронике и информационным технологиям, «феномен дипфейков, который на самом деле представляет собой синтез ИИ и индустрии дезинформации, безусловно, является для всех нас поводом для беспокойства, потому что данная технология действительно может оказаться очень эффективной для её применения против отдельных людей, обществ, сообществ и стран» (Sukumaran 2023). 17 ноября 2023 г. премьер-министр Индии Н. Моди, предупредив, что дипфейки могут привести к кризису и «разжечь огонь недовольства» в обществе, рассказал, что недавно он увидел дипфейк, на котором он сам исполнял гуджаратский танец «гарба» (который на самом деле не танцевал со школы). Например, активное применение дипфейков в качестве инструмента информационно-психологического воздействия наблюдалось после индо-китайского пограничного конфликта в Ладакхе в июне 2020 г. В мае 2023 г. Индия стала свидетелем примера манипуляции выражением лица, произошедшего в разгар пятимесячного протеста олимпийских женщин-борцов, требующих расследования дел о сексуальных домогательств в отношении них. Когда полиция Дели задержала протестующих, рестлерши Винеш и Фогат опубликовали селфи, на котором они в мрачном настроении сидят в полицейском фургоне вместе с полицейскими. Однако существовала и другая версия данных фотографий, на которых обе женщины улыбались на камеру – она стала вирусной, пока не была разоблачена как фейк (Sukumaran 2023). Все эти случаи ясно показывают существование угрозы информационно-психологической безопасности третьего уровня, которую представляет злонамеренное использование дипфейков.

В индийском обществе уже растет озабоченность возможными психологическими последствиями данной угрозы. А. Капур директор Современной государственной школы Шалимар Багх, говорит: “Для учащихся социальные сети – это больше, чем просто платформа для общения; это виртуальное пространство, где формируется их личность, завязываются дружеские отношения и

повышается самооценка. По мере того, как прецеденты с использованием фейков проникают в социальные сети, последствия для психического здоровья обучающихся становятся все более тревожными. Цифровое изменение контента, которое когда-то воспринималось как безобидная забава, теперь представляет серьезную угрозу психологическому благополучию молодых людей. Принуждение соответствовать измененным цифровыми технологиями стандартам красоты и поведения в сочетании со страхом стать жертвами злонамеренных манипуляций посредством применения дипфейков создает токсичную среду для студентов" (The Times of India 2023b). Журналисты предлагают нам представить сценарий, когда ребенок обнаруживает дипфейк, в котором он выставлен в непристойном виде, совершая действия, которые он никогда не делал. "Возможность нанесения ущерба чьей-либо репутации и последующая социальная изоляции могут нанести серьезный ущерб. Страх стать следующей мишенью вводит детей в состояние постоянной тревоги, подрывая их доверие к цифровому миру и усугубляя и без того сложный подростковый период" (The Times of India 2023b).

В апреле 2023 г. в южноиндийском штате Тамилнад произошел политический скандал, когда К. Аннамалай, глава партии Бхаратия Джаната (БДП) — правящей партии Индии, — опубликовал аудиозапись сомнительного происхождения, содержащей выступление П. Тиагараджана, депутата от партии Дравида Муннетра Кажагам (ДМК), который в настоящее время является министром информационных технологий и цифровых услуг и президентом Законодательного собрания штата Тамилнад. На 26-секундной аудиозаписи низкого качества можно было услышать, как Тиагараджан, бывший в то время министром финансов штата Тамилнад, обвинял членов своей партии в незаконном обогащений на $3,6 млрд. Тиагараджан категорически опроверг оригинальность записи, назвав ее «сфабрикованной» и «сгенерированной машиной». «Никогда не доверяйте аудиозаписи, если не указан источник ее происходения», — написал Тиагараджан в Твиттере 22 апреля 2023 г., заявив, что сейчас голоса очень легко подделать. 25 апреля К. Аннамалай опубликовал второй клип продолжительностью 56 секунд и с гораздо более качественным звуком, в котором Тиагараджан якобы пренебрежительно отзывался о своей партии и восхвалял БДП. На этот раз Тиагараджан назвал это отчаянной попыткой «банды шантажа» создать политический раскол внутри его собственной партии и заявил, что никто не заявил о своем авторстве в отношении этих записей. Мнения аналитиков по поводу подлинности первого клипа разделились: одни посчитали, что его качество слишком плохое, чтобы можно было сделать какие-либо выводы; другие решили, что клип «скорее всего, фейк». Однако второй клип был признан оригинальным (Christopher 2023). Непоследовательность и неоднозначность этого инцидента стала предметом спекуляций в Индии на тему того, что в эпоху тотального злонамеренного использования дипфейков недобросовестные политики могут объявить любые компроматы, даже подлинные, сфабрикованными. Так или иначе, подобные случаи обнажают другую сторону проблемы дипфейков – снижения доверия к источникам информации в целом.

Заключение

В настоящее время для Индии характерны угрозы информационной-психологической безопасности посредством ЗИИИ всех трёх уровней, однако угрозы второго и третьего уровней имеют наиболее ярко выраженный характер. Угрозы второго уровня могут привести к крупным техногенным катастрофам. В то же время прямое физическое воздействие подобных последствий на экономические, военные и политические структуры может быть дополнено информационно-психологическим воздействием (страх, паника и другие массовые эффекты), которое может нанести вторичный удар по этим структурам посредством ослабления воли, решимости и способности людей противостоять последствиям техногенных катастроф. Эти катастрофы могут

сопровождаются своеобразным остаточным “психологическим влиянием”, долговременными травматическими психологическими последствиями, которые также могут негативно влиять на жизнь общества в течение длительного времени. Следует учитывать, что многие угрозы ЗИИИ второго уровня получают более широкое распространение по мере увеличения доступности и удешевления технологий ИИ, на основе которых злоумышленники уже создают свои продукты злонамеренного воздействия для создания деструктивных эффектов (например, мошеннического контента). Психологические инструменты социальной инженерии значительно усовершенствованы технологиями, которые повышают опасность скрытого фишинга. Угрозы третьего уровня включают деструктивно-запрограммированных чат-ботов и быстрое распространение злонамеренного использования фейков. Системный характер угроз требует целостного и комплексного реагирования, потребность в котором в индийском обществе уже созрела.

Литература

Ajder H (2019) Social Engineering and Sabotage: Why Deepfakes Pose An Unprecedented Threat To Businesses. In: Deeptrace. https://deeptracelabs.com/social-engineering-and-sabotage-why-deepfakes-pose-an-unprecedented-threat-to-businesses/. Accessed 21 Jun 2022

Ajder H, Patrini G, Cavalli F, Cullen L (2019) The State of Deepfakes: Landscape, Threats, and Impact. In: The Register. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf. Accessed 21 Jun 2022.

Alavi M, Achom D (2020) BJP Shared Deepfake Video on WhatsApp During Delhi Campaign. In: NDTV. https://www.ndtv.com/india-news/in-bjps-deepfake-video-shared-on-whatsapp-manoj-tiwari-speaks-in-2-languages-2182923. Accessed 21 Jun 2022

Ali A, Sarwar N (2023) ChatGPT Has Been Sucked Into India’s Culture Wars. In: Wired. https://www.wired.com/story/chatgpt-has-been-sucked-into-indias-culture-wars/. Accessed 18 Jan 2024

Bandara P (2023) India is Rocked by Deepfake Video Scandal Featuring Bollywood Star. In: Petapixel. https://petapixel.com/2023/11/09/india-is-rocked-by-deepfake-video-scandal-featuring-bollywood-star/. Accessed 18 Jan 2024

BBC (2023) Indian CEO criticised for picking AI bot over human staff. In: MyjoyOnline.

Bhatt S (2018) How Indian startups gear up to take on the voice assistants of Apple, Amazon and Google. In: The Economic Times. https://economictimes.indiatimes.com/small-biz/startups/features/how-indian-startups-gear-up-to-take-on-the-voice-assistants-of-apple-amazon-and-google/articleshow/64044409.cms. Accessed 21 Jun 2022

Biswas P (2022) Deepfakes, Crypto Scams on the Rise in India 2022. In: Digit. https://www.digit.in/features/crypto/deepfakes-crypto-scams-on-the-rise-in-india-2022-63740.html. Accessed 23 Jun 2022

Bundhun R (2023) How artificial intelligence can transform India's economy. In: The National. https://www.thenationalnews.com/business/2023/04/03/how-artificial-intelligence-can-transform-indias-economy/. Accessed 18 Jan 2024

Chakraborty M (2022) Artificial Intelligence: Growth and Development in India. In: Analytics Insight. https://www.analyticsinsight.net/artificial-intelligence-growth-and-development-in-india/. Accessed 21 Jun 2022

Christopher N (2023) An Indian politician says scandalous audio clips are AI deepfakes. We had them tested. In: Rest of World. https://restofworld.org/2023/indian-politician-leaked-audio-ai-deep-fake/. Accessed 18 Jan 2024

Doval P (2023) AI new tool for online scammers as 82% Indians concede to clicking on or fall for fake messages: Survey. In: The Times of India. https://timesofindia.indiatimes.com/india/ai-new-tool-for-online-scammers-as-82-indians-concede-to-clicking-on-or-fall-for-fake-messages-survey/articleshowprint/105072681.cms. Accessed 19 Jan 2024

Egress (2021) The psychology of social engineering and phishing. https://www.egress.com/blog/phishing/psychology-social-engineering-phishing. Accessed 19 Jan 2024

Faggella D (2019) Artificial Intelligence in India—Opportunities, Risks, and Future Potential. In: Emerj Artificial Intelligence Research. https://emerj.com/ai-market-research/artificial-intelligence-in-india/. Accessed 21 Jun 2022

https://www.myjoyonline.com/indian-ceo-criticised-for-picking-ai-bot-over-human-staff/. Accessed 18 Jan 2024

Mihindukulasuriya R (2020) Why the Manoj Tiwari deepfakes should have India deeply worried. In: The Print. https://theprint.in/tech/why-the-manoj-tiwari-deepfakes-should-have-india-deeply-worried/372389/. Accessed 21 Jun 2022

OpIndia (2022) “Removed accounts for manipulation and spam,” Twitter says after Rahul Gandhi loses bot followers. In: OpIndia. https://www.opindia.com/2022/01/removed-accounts-for-manipulation-and-spam-twitter-says-after-rahul-gandhi-loses-followers/. Accessed 21 Jun 2022

Pindrop (2020) Voice Intelligence & Security Report. A review of fraud, the future of voice, and the impact to customer service channels. Revised for 2020 including updated data. Pindrop, Atlanta

Sengupta A (2023) Over 70 per cent Indian workers fear losing jobs to AI, new Microsoft survey reveals. In: India Today. https://www.indiatoday.in/technology/news/story/over-70-per-cent-indian-workers-fear-losing-job-ai-new-microsoft-survey-reveals-2387406-2023-06-01. Accessed 18 Jan 2024.

Shivji S (2023) India's religious chatbots condone violence using the voice of god. In: CBC. https://www.cbc.ca/news/world/india-religious-chatbots-1.6896628. Accessed 18 Jan 2024

Srivastava R (2023) AI Landscape for 2024: Navigating the Journey in India. In: Financial Express. https://www.financialexpress.com/business/industry-ai-landscape-for-2024-navigating-the-journey-in-india-3350643/. Accessed 18 Jan 2024

Stanly M (2023) AI in cybersecurity: How is India grappling with the risks of cyber-attack. In: IndiaAI. https://indiaai.gov.in/article/ai-in-cybersecurity-how-is-india-grappling-with-the-risks-of-cyber-attack. Accessed 19 Jan 2024

Sukumaran A (2023) Deepfakes: Clear and present danger. In: India Today. https://www.indiatoday.in/india-today-insight/story/deepfakes-clear-and-present-danger-2473187-2023-12-07. Accessed 18 Jan 2024

Thaker A (2019) Automated bots manipulated Twitter traffic before Narendra Modi’s visit to Tamil Nadu: US think-tank. In: Scroll.in. https://scroll.in/article/919445/automated-bots-manipulated-twitter-traffic-before-narendra-modis-visit-to-tamil-nadu-us-think-tank. Accessed 23 Jun 2022

The Times of India (2023a) Beware of FraudGPT, the rogue AI chatbot. https://timesofindia.indiatimes.com/city/ahmedabad/beware-of-fraudgpt-the-rogue-ai-chatbot/articleshow/102267830.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst. Accessed 18 Jan 2024

The Times of India (2023b) How social media scandals like deepfake impact minors and students' mental health. https://timesofindia.indiatimes.com/life-style/parenting/moments/how-social-media-scandals-like-deepfake-impact-minors-and-students-mental-health/articleshow/105168380.cms. Accessed 18 Jan 2024

Yusuf L (2020) A New Internet, Data Banks and Digital World War. In: The AI Book (eds. S. Chishti, I. Bartoletti, A. Leslie and S.M. Millie). https://doi.org/10.1002/9781119551966.ch6

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Южно-Африканской Республике

Е. Н. Пашенцев, Д. Ю. Базаркина

Введение

В ЮАР на государственном уровне уделяется большое внимание пользе, которую общество и экономика могут извлечь из масштабного внедрения систем ИИ. В 2019 г. президент Сирил Рамафоса организовал Президентскую комиссию по четвертой промышленной революции (The Presidential Commission on Fourth Industrial Revolution, PC4IR). Комиссия помогает правительству использовать возможности, предоставляемые цифровой промышленной революцией, включая ИИ и машинного обучения (The Presidency, Republic of South Africa 2019). Факт того, что «комиссия состоит из представителей технологических стартапов, академических кругов, специалистов по кибербезопасности, исследователей, социологов, профсоюзных активистов и других представителей ключевых секторов экономики» (Ramafosa 2020), свидетельствует о глубоком понимании влияния ИИ, изучаемого совместно специалистами в области технических и социальных наук, на все сферы жизни общества. По словам президента С. Рамафосы, Южная Африка стремится к максимальному использованию потенциала технологических инноваций для того, чтобы к 2030 г. обеспечить устойчивый рост экономики и повышение благосостояния людей (Ramafosa 2020). В октябре 2020 г. Комиссия по четвертой промышленной революции рекомендовала создать институт ИИ для содействия генерированию новых знаний и внедрению приложений ИИ в таких секторах, как здравоохранение, сельское хозяйство, образование, энергетика, производство, туризм и информационно-коммуникационные технологии, а также образование с целью обеспечения положительного социального воздействия (PC4IR 2020). 30 ноября 2022 г. Департамент коммуникаций и цифровых технологий (англ. Department Of Communications and Digital Technologies, DCDT) создал Южноафриканский институт искусственного интеллекта (англ. The Artificial Intelligence Institute of South Africa, AIISA) (AIISA 2023). Развитие ИИ в Южной Африке привело к созданию ряда известных компаний в этой области (GoodFirms 2022), а правительственные инициативы помогают создавать институциональные условия и инфраструктуру для дальнейшего внедрения ИИ.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Для ЮАР являются актуальными угрозы информационно-психологической безопасности первого уровня посредством ЗИИИ. Эти угрозы включают в себя возможность социальной дестабилизации из-за замещения все большего числа работников технологиями ИИ. Я. Голдин, профессор в области глобализации и развития Оксфордского университета, говорит, что в будущем ИИ может автоматизировать миллионы рабочих мест, нанеся ущерб экономическому росту в развивающихся регионах, таких как Африка (BBC News 2022). По данным консалтинговой компании McKinsey, многие рабочие места в Южной Африке уже автоматизированы, особенно в результате механизации в таком исторически трудоемком секторе, как горнодобывающий. Такие отрасли, как торговля и сфера услуг, также находятся в зоне риска. Например, сеть супермаркетов Pick n Pay Stores внедрила автоматизированную систему оформления заказов с поддержкой

ИИ, которая замещает кассиров (Van den Berg, 2018). На фоне высокого уровня безработицы степень общего понимания специфики использования ИИ в Южной Африке остаётся крайне низкой. Как следствие, связанная с внедрением систем ИИ тревога по поводу потери работы быстро растет среди экономически активного населения (Business Tech 2019).

Согласно исследованию Kaspersky Research, работающие люди в Южной Африке считают, что чем лучше роботы будут справляться с различными задачами, тем больше они займут рабочих мест. Большинство опрошенных местных сотрудников (74%) считают, что роботы должны более широко использоваться в различных отраслях, однако многие опасаются возможности их потенциального взлома. Сотрудники сообщили о повышении уровня роботизации в своих компаниях за последние два года. 33% сотрудников из ЮАР заявили, что их организации уже используют роботов, а 39% местных организаций планируют использовать их в ближайшем будущем (IT-Online 2022). Большинство опрошенных сотрудников в ЮАР (92%) считают, что роботы в конечном итоге заменят людей в их отрасли. Поскольку роботы внедряются во всех секторах экономики, людям необходимо постоянно адаптироваться, повышать квалификацию и профессионализм, получать новые компетенции, знания и навыки, чтобы не потерять работу по причине автоматизации. И они готовы к этому: среди тех, кто считает, что их рабочие места могут быть заняты роботами, большинство (75%) готовы освоить новые навыки или улучшить существующие. Еще один важный вывод, который необходимо сделать – роботизация существенно повышает риски для обеспечения кибербезопасности. Большинство местных респондентов (89%) считают, что роботы могут быть взломаны, а 53% знают о подобных инцидентах, произошедших в их или других местных компаниях. В своей оценке того, насколько обеспечена кибербезопасность роботов, мнения респондентов разделились: почти половина опрошенных сотрудников в ЮАР (42%) считают, что для обеспечения безопасности роботов в различных отраслях применяются недостаточные меры (IT-Online 2022).

Кроме того, угрозы первого уровня проявились, когда США попытались вытеснить с южноафриканского рынка китайскую телекоммуникационную компанию Huawei, поставляющую в страну интернет-технологии 5G, необходимые для развития ИИ. Huawei находится «на передовой» развития 5G, и в 2018 г. США запустили кампанию, направленную на то, чтобы помешать другим странам покупать оборудование Huawei. Президент Южной Африки С. Рамафоса в своей вступительной речи на конференции по цифровой экономике в Йоханнесбурге в июле 2019 г. заявил, что США «явно завидуют тому, что китайская компания Huawei опередила их, и поскольку их обогнали, они теперь должны наказать эту компанию» (Staden, 2019). В 2019 г. Huawei подписала контракт с ЮАР на создание первой коммерческой сети 5G на африканском континенте. По словам К. ван Стадена, исследователя китайско–африканских отношений в Южноафриканском институте международных отношений, Huawei уже построила около 70 % сетей 4G на континенте (EFE–EPA 2019).

26 февраля 2020 г. южноафриканская компания мобильной связи Rain объявила о планах по созданию подключенной к 5G транспортной сети с использованием таких решений Huawei, как оптическое кросс-соединение (или оптическая кросс-коммутация, OXC) и 200G, используя новейший полностью оптический коммутационный продукт Huawei — OXC (P32) – для создания оптической транспортной сети метрополитена. Rain сосредоточена на внедрении сетей мобильной широкополосной связи в Южной Африке и становится первым оператором, широко развернувшим сети 5G в стране (Huawei 2020).

12 марта 2020 г. на саммите по здравоохранению и правам человека в Тусоне (штат Аризона), доктор Т. Коуэн выдвинул гипотезу, согласно которой COVID-19, возможно, был вызван 5G-связью (Huawei 2020). Начав свою карьеру с преподавания садоводства в качестве волонтера

Корпуса мира в Свазиленде и Южной Африке, позже он занимал пост вице-президента Ассоциации врачей антропософской медицины. Заявления Коуэна, занимавшего пост члена-основателя Фонда У. А. Прайса, в большинстве своем были опровергнуты, но дискуссии о его гипотезах привлекли множество сторонников на онлайн-платформах. Это естественно в условиях нарастающей паники, которая активно нагнеталась средствами массовой информации и социальными сетями.

Оригинальная гипотеза Т. Коуэна вызвала бурную дискуссию в южноафриканских СМИ и социальных сетях, которая вышла за пределы страны. Более 4000 человек подписали петицию с требованием прекратить внедрение связи 5G в Кейптауне на платформе для сбора подписей Change.org (Independent 2020). Аналогичные петиции были распространены во всей ЮАР. Доступ к этим сайтам был ограничен правительством в апреле, хотя это не остановило от протестных акций сторонников идей Т. Коуэна.

Не углубляясь в суть аргументации Т. Коуэна, следует отметить, что его оценка взаимосвязи между COVID-19 и технологиями 5G стала инструментом в конкурентной борьбе. В Южной Африке Huawei являлась бесспорным лидером в области внедрения 5G, и активная кампания против распространения нового поколения мобильной связи в стране была направлена почти исключительно против этих инноваций, что облегчалось тяжелыми последствиями COVID-19. В конечном счете внедрение технологий 5G было продолжено и к середине 2023 г. доля мобильных подключений 5G в Южной Африке уже достигла 4%, что явилось самой высокой долей подключений 5G среди африканских стран (IT Online 2024). При этом Huawei сохраняет лидирующие позиции в технологиях 5G как на рынках ЮАР (MyBroadband 2023 and 2024, Huawei 2023), так и на глобальном уровне (Huawei 2024).

Вышеприведенный пример демонстрирует, по какому пути могут идти торговые войны. В данном случае, тесно связанные с ИИ технологии стали предметом информационно-психологического противоборства, в которую, наряду с бизнес-структурами, оказались вовлечены и политические элиты. Подобные инциденты вызывают тревогу, поскольку в этом виде конфронтации задействовано множество негосударственных субъектов, которые, руководствуясь ложными сообщениями и теориями заговора, могут нанести ущерб физической инфраструктуре, нагнетать панику в обществе или дискредитировать технологии, направленные на преодоление экономических проблем.

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Как и для многих других стран, для Южной Африки в одинаковой степени характерны угрозы осуществления фишинговых атак, число которых здесь растет и которые осуществляются в отношении как частных лиц, так и предприятий. Одной из причин эффективности фишинговых атак в стране является высокий уровень доверия людей к знакомым учреждениям. Киберпреступники эксплуатируют это доверие, используя местные бренды и соответствующие проблемы, чтобы повысить вероятность успеха своей аферы. Распространенные фишинговые ловушки в ЮАР могут включать в себя поддельные банковские электронные письма, мошенничество с возвратом налогов от Налоговой службы Южной Африки (SARS) и сообщения или новости о популярных событиях (Apex 2023). Например, согласно сообщению, которое было растиражировано в мессенджере WhatsApp, правительство ЮАР собирается предоставить «всем родителям ЮАР» пособие на содержание ребенка в размере 1100 рэндов (около $59) на шесть месяцев. Сообщение также было опубликовано в социальной сети Facebook в публичных группах с тысячами под-

писчиков, причем некоторые пользователи спрашивали, соответствует ли это утверждение действительности. В ЮАРе нет «министра по гуманитарным вопросам и борьбе с бедностью» (который был упомянут в сообщении), а за субсидии на содержание детей отвечает Южноафриканское агентство социального обеспечения (The South African Social Security Agency, SASSA), национальное агентство правительства Южной Африки. Некоммерческая организация Africa Check произвела поиск упоминаний об иске в аккаунтах SASSA в социальных сетях и обнаружила пост в Facebook, опубликованный агентством 12 января 2024 г., со скриншотом сообщения с пометкой «Фейк» (Khourie 2024).

С. Пауэлл, глава отдела криминалистики юридической компании ENSafrica, указывает, что кибермошенничество приносит ущерб экономике Южной Африки примерно в размере 2 миллиардов рандов в год – именно поэтому компании тратят миллионы на усиление своей кибербезопасности (Moodley 2023). «Генеративный ИИ предоставляет новые возможности по созданию удивительно убедительных фишинговых электронных писем, сообщений и веб-сайтов, которые дублируют официальный стиль и символику организаций, чтобы обманом заставить частных лиц предоставить платежные данные и разглашать конфиденциальную информацию» (Africa Business 2023), - сообщает Р. Мер, генеральный директор компании Eftsure Africa, специализирующейся на обеспечении безопасности онлайн-платежей и делающей обзор текущих угроз для ЮАР.

Адвокаты, ведущие дело в региональном суде Йоханнесбурга, в соответствии с судебным решением были привлечены к ответственности за использование фейковых ссылок, сгенерированных текстовой генеративной нейросетью ChatGPT. Согласно газете Sunday Times, судебное решение также наложило на клиента адвокатов штрафные санкции (Prior 2023). В рассматриваемом случае речь шла о женщине, которая подала в суд на корпорацию, в которой она работала, за клевету. Адвокат, защищающий интересы корпорации, утверждал, что иск неправомерен. В свою очередь, адвокат истца М. Паркер заявила, что ранее уже были судебные прецедентны, которыми необходимо руководствоваться при решении данного вопроса, а у адвоката и корпорации просто не было времени ознакомиться с ними. Мировой судья А. Чайтрам отложил рассмотрение дела, чтобы дать обеим сторонам достаточно времени для поиска информации, необходимой им для доказательства своей правоты. В последующие два месяца юристы компании, принимавшие участие в деле, пытались проверить информацию, на которую ссылались адвокаты истицы. Они обнаружили, что, хотя ChatGPT и ссылался на настоящие случаи и приводил реальные цитаты, это были совершенно иные ситуации и с точки зрения прецедентного права они совсем не подходили для решения вопросов о диффамации между юридическими и физическими лицами. Затем было установлено, что судебные решения были разработаны «при помощи ChatGPT».

А. Чайтрам постановил, что адвокаты вовсе не намеревались вводить в заблуждение суд — они «просто переусердствовали и одновременно проявили небрежность». Это означало, что в отношении адвокатов не было предпринято никаких дальнейших санкционных мер, кроме постановления о штрафных расходах. «Смущение, которое испытали адвокаты истца, связанное с этим инцидентом, вероятно, является достаточным наказанием для них», – заявил Чайтрам (Prior 2023). Данный пример совмещает угрозы второго и третьего уровней ЗИИИ, поскольку адвокаты сознательно не намеревались вводить суд в заблуждение, но данная дезинформация была непреднамеренно использована в ходе судебного разбирательства.

«В последние месяцы множество работодателей, включая глобальных технологических гигантов, стали жертвами «утечек разговорного ИИ» (утечки данных, в которых задействован чат-бот). При этом конфиденциальные данные/информация, к которым имеют доступ чат-боты, та-

кие как ChatGPT, непреднамеренно раскрываются. Информация отправляется на сторонний сервер и используется для дальнейшего обучения чат-бота. Это означает, что информация, внедряемая в чат-бота, может быть использована алгоритмом при формировании ответов в будущем (Boda et al 2023). Проблема усугубляется, когда чат-бот имеет доступ к конфиденциальной информации и использует ее. В период с марта 2022 по март 2023 г. средняя общемировая стоимость утечек данных достигла рекордно высокого уровня – $4,45 млн. $, а для ЮАР, в частности, она превысила 50 млн. рандов (Boda et al 2023). Этот пример описывает техническую ошибку, однако сам факт этого прецедента показывает, что у злоумышленников могут появится новые возможности для эксплуатации данной уязвимости.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

В контексте угроз информационно-психологической безопасности третьего уровня острой социальной проблемой в Южной Африке является буллинг. Буллинг, присутствующий сегодня в социальных сетях или мобильных приложениях, может быть усилен встроенными в эти цифровые платформы алгоритмами ИИ. Согласно отчету Независимой системы общественного опроса (Independent Polling System of Society), в 2020 г. в Южной Африке наблюдался самый высокий уровень кибербуллинга: 54% родителей заявили, что они как минимум знают детей из местного сообщества, которые подверглись травле (Kahla 2020). Благодаря способности алгоритмов в социальных сетях ранжировать контент, привлекающий большое внимание, информация сомнительного содержания может быть быстро распространена реальными пользователями, что приводит к повышению её рейтинга, и, соответственно, большему охвату аудитории сообщениями, содержащих язык ненависти.

В ЮАР произошла утечка фейковых аудиозаписей с заседаний руководства правящего Африканского национального конгресса (АНК) в преддверии выборной конференции партии 2022 г. Согласно заявлению заместителя генерального секретаря АНК Дж. Дуарте, которое он сделал в рамках радиоинтервью новостному агентству Eyewitness News 15 апреля 2021 г., «кто бы это ни сделал, он сделал это очень профессионально». «[Аудиозапись] была смонтирована очень качественно с целью создать особое впечатление весьма конструктивного разговора на встрече официальных лиц АНК» (Maree 2021). «Одна из утечек содержала предполагаемую запись выступления Дуарте на встрече шести высших должностных лица АНК с бывшим президентом Дж. Зумой в конце прошлого месяца. Партия пытается убедить Дж. Зуму дать показания перед комиссией по расследованию крупномасштабной коррупции во время его пребывания в должности, получившей название "захват государства", но он продолжает отказываться» (Maree 2021).

Это уже не первый случай, когда явно сфальсифицированная информация была опубликована во время внутренних баталий партии. Однако этот случай дезинформации, вероятно, предназначенный специально для того, чтобы продемонстрировать избирателям раскол в рядах АНК, был назван в СМИ дипфейком. По словам ректора Университета ООН, инженера в области ИИ Т. Марвала, новые и легкодоступные технологии упрощают манипулирование информацией и делают воздействие более эффективным. Он также отмечает, что в ЮАР уже были прецеденты изощренных киберманипуляций (Maree 2021), и что злонамеренное использование технологий в политическом соперничестве, таких как дипфейки и, в частности, подделка голоса, может угрожать демократии в стране.

Количество дипфейков резко возросло по всему миру, и ЮАР не исключение. Среди всех африканских стран наибольшее количество воздействий с применением дипфейков было зафик-

сировано именно в ЮАР (19,7%), а также в Нигерии (11,5%). «Южноафриканское население беспокоит беспрецедентный рост числа мошенничеств с применением дипфейков, который составил 1200%», – заметил Ханнес Безуйденхаут, вице-президент по продажам компании Sumsub, специализирующейся на обеспечении безопасности личных данных (Fraser 2023). «На фоне роста на 450% случаев мошенничества с личными данными в странах Ближнего Востока и Северной Африки, эти технологии представляют значительную угрозу и повод для беспокойства». Х. Безуйденхаут считает, что создавать дипфейки становится проще, поскольку мошенники используют подлинный материал с изображением человека и извлекают из него фотографию для создания 3D-образа. «Провайдеры, которые не предпринимают постоянных усилий по обновлению технологий обнаружения дипфейков, ставят под угрозу и свой бизнес, и пользователей» (Fraser 2023).

Существующие примеры также включают злонамеренное использование дипфейков с участием ведущих южноафриканского телевидения. Южноафриканская телерадиовещательная корпорация (The South African Broadcasting Corporation, SABC) 14 ноября 2023 г. была вынуждена выступить с заявлением, что их ведущие Б. Зване и Ф. Херд были показаны в дипфейк-видеороликах, циркулирующих в Интернете. Эти видеоролики, содержащие рекламу мошеннической инвестиционной схемы, привлекли значительное внимание, а одно из них – с «участием» Ф. Херд – набрало более 123 000 просмотров на YouTube с момента его опубликования 3 ноября 2023 г. (Women Press Freedom 2023). Дипфейки, которые включали логотип SABC и изображали поддельного И. Маска, рекламирующего мошенническую схему, вызвали серьезные опасения по поводу влияния данной технологии на доверие к СМИ и поставили на повестку дня вопрос о свободе прессы. В ответ на это Ф. Херд и Б. Зване публично опровергли какую-либо причастность к дипфейк-видеороликам, созданным ИИ. М. Монаре, исполнительный директор по новостям и текущим вопросам SABC, осудил акт распространения дипфейков, подчеркнув необходимость защиты репутации общественного вещателя и его журналистов. Инцидент не только подрывает доверие к отдельным журналистам, но и представляет более широкую угрозу СМИ, затрудняя возможности для общественности отличать реальный контент от фейкового (Women Press Freedom 2023).

В то время как чуть менее половины работающих людей, опрошенных в ЮАР (42%), заявили, что могут определить дипфейки, по словам руководителя компании «Лаборатория Касперского» Е. Касперского, в действительности в ходе специального теста лишь 21% смогли отличить реальное изображение от сгенерированного ИИ. Это означает, что организации уязвимы для подобных мошенничеств, поскольку киберпреступники используют изображения, сгенерированные ИИ для осуществления деструктивной деятельности. С помощью дипфейков они могут создавать фальшивые видео или изображения, которые затем могут быть использованы для введения в заблуждение отдельных лиц или даже организаций. Например, киберпреступники могут создать фейковое видео генерального директора, запрашивающего банковский перевод или санкционирующего платеж, которое может быть использовано для кражи корпоративных средств. Могут быть созданы также и компрометирующие видео или изображения физических лиц, которые затем будут использованы для вымогательства у них денег или информации. Киберпреступники также смогут использовать дипфейки для распространения ложной информации или манипуляции общественным мнением – 55% опрошенных работающих лиц в Южной Африке считают, что их компания может потерять деньги из-за дипфейков (IT-Online 2024). Таким образом, потенциал угроз третьего уровня посредством ЗИИИ в ЮАР продолжает расти.

Заключение

Проведенный анализ показывает, что в первую очередь для ЮАР характерны угрозы информационно-психологической безопасности первого уровня, вызванные ЗИИИ. Вместе с тем,

число случаев цифрового мошенничества, в которых может использоваться ИИ, также растет (как это происходит и в глобальном масштабе). ЮАР представляет собой перспективный рынок для технологических компаний из других стран, что иногда приводит к обострению информационно-психологического противоборства (как в случае с попытками вытеснить Huawei с этого рынка). Страна уже столкнулась с политическим использованием дипфейков, что на фоне слабой способности граждан распознавать поддельный медиа-контент усугубляет угрозы третьего уровня.

Литература

Africa Business (2023) Understanding generative AI and its impact on payment fraud in South Africa. https://africabusiness.com/2023/11/21/understanding-generative-ai-and-its-impact-on-payment-fraud-in-south-africa/. Accessed 19 Jan 2024

AIISA (2023) About Artificial Intelligence. https://aii-sa.co.za/. Accessed 18 Jan 2024

Apex (2023) Phishing Attacks: Recognizing and Avoiding Common Scams in South Africa. https://apexcybertechnologies.co.za/blog/phishing-attacks-recognizing-and-avoiding-common-scams-in-south-africa/. Accessed 18 Jan 2024

BBC News (2022) Will AI kill developing world growth? In: BBC News. https://www.bbc.com/news/business-47852589. Accessed 20 Jun 2022

Boda R, Salt L, Keil L, Powell A (2023) South Africa: Conversational AI Leaks: How Can Employers Mitigate The Risks Of Using ChatGPT In The Workplace? In: Mondaq. https://www.mondaq.com/southafrica/privacy-protection/1402174/conversational-ai-leaks-how-can-employers-mitigate-the-risks-of-using-chatgpt-in-the-workplace. Accessed 19 Jan 2024

Business Tech (2019) How AI is being used in South Africa. In: Business Tech. https://businesstech.co.za/news/enterprise/322505/how-ai-is-being-used-in-south-africa/. Accessed 21 Jun 2022

EFE-EPA (2019) South African president says USA jealous of Huawei. In: www.efe.com. https://www.efe.com/efe/english/business/south-african-president-says-usa-jealous-of-huawei/50000265-4016943. Accessed 21 Jun 2022

Fraser L (2023) The crime seeing 1,200% growth in South Africa. In: BusinessTech. https://businesstech.co.za/news/technology/735165/the-crime-seeing-1200-growth-in-south-africa/. Accessed 18 Jan 2024

GoodFirms (2022) Top Artificial Intelligence Companies in South Africa 2022. In: GoodFirms. https://www.goodfirms.co/artificial-intelligence/south-africa. Accessed 3 Jun 2022

Huawei (2020) South Africa's Rain and Huawei Build the First 5G Transport Networks Using OXC+200G Solution. In: Huawei. https://www.huawei.com/en/press-events/news/2020/2/5g-transport-networks-oxc-200g-solution. Accessed 22 Apr 2020

Huawei (2023) Huawei holds Africa 5G Summit at Africacom. https://www.huawei.com/en/news/2023/11/huaweiafrica-5gsummit. Accessed 06 Jun 2024

Huawei (2024) Huawei 5G Core Named "Leader" for the Sixth Consecutive Year by GlobalData, Gets Full Scores in All Dimensions for the First Time. https://www.huawei.com/en/news/2024/3/leader-5g-core. Accessed 06 Jun 2024

Independent (2020) Watch: Debate raging on link between 5G technology, coronavirus pandemic. In: Independent Online (IOL). https://www.iol.co.za/capetimes/news/watch-debate-raging-on-link-between-5g-technology-coronavirus-pandemic-45124913. Accessed 22 Apr 2020

IT Online (2024) The battle to defend the 5G attack surface. https://it-online.co.za/2024/04/18/the-battle-to-defend-the-5g-attack-surface/. Accessed 06 Jun 2024

IT-Online (2022) The robots are coming … and people are wary of job losses. https://it-online.co.za/2022/11/21/the-robots-are-coming-and-people-are-wary-of-job-losses/. Accessed 18 Jan 2024

Kahla C (2020) Social media platforms need to take a stand against cyberbullying. In: The South African. https://www.thesouthafrican.com/technology/social-media-stand-against-cyberbullying/. Accessed 3 Jun 2022

Khourie T (2024) Scam alert! South African government is not giving R1,100 child support grant to every parent through this link. In: Africa Check. https://africacheck.org/fact-checks/meta-programme-fact-checks/scam-alert-south-african-government-not-giving-r1100-child. Accessed 18 Jan 2024

Maree A (2021) South Africa: Leaks and deepfakes shaping the race for ANC presidency. In: The Africa Report. https://www.theafricareport.com/80648/south-africa-leaks-and-deepfakes-shaping-the-race-for-anc-presidency/. Accessed 3 Jun 2022

Moodley N (2023) Deepfakes, hackers and the man in the middle – the murky world of cyber fraud. In: Daily Maverick. https://www.dailymaverick.co.za/article/2023-10-23-deepfakes-hackers-and-the-man-in-the-middle-the-murky-world-of-cyber-fraud/. Accessed 18 Jan 2024

MyBroadband (2023) Huawei equipment offers the fastest 5G speeds in South Africa. https://mybroadband.co.za/news/5g/505784-huawei-equipment-offers-the-fastest-5g-speeds-in-south-africa.html. Accessed 06 Jun 2024

MyBroadband (2024) Why Huawei is South Africa's Best Mobile Network Vendor in 2024. https://mybroadband.co.za/news/industrynews/539123-why-huawei-is-south-africas-best-mobile-network-vendor-in-2024.html. Accessed 06 Jun 2024

PC4IR (2020) Report of the Presidential Commission on the 4th Industrial Revolution. In: South African Government. https://www.gov.za/documents/report-presidential-commission-4th-industrial-revolution-23-oct-2020-0000. Accessed 3 Jun 2022

Prior B (2023) South African lawyers use ChatGPT to argue case — get nailed after it makes up fake info. In: My Broadband. https://mybroadband.co.za/news/software/499465-south-african-lawyers-use-chatgpt-to-argue-case-get-nailed-after-it-makes-up-fake-info.html. Accessed 18 Jan 2024

Ramaphosa C (2020) A national strategy for harnessing the Fourth Industrial Revolution: The case of South Africa. In: Brookings. https://www.brookings.edu/blog/africa-in-focus/2020/01/10/a-national-strategy-for-harnessing-the-fourth-industrial-revolution-the-case-of-south-africa/. Accessed 3 Jun 2022

Staden C van (2019) How the US-China conflict over Huawei could play out in Africa. In: The Africa Report. https://www.theafricareport.com/15451/how-the-us-china-conflict-over-huawei-could-play-out-in-africa/

The Presidency, Republic of South Africa (2019) President appoints Commission on Fourth Industrial Revolution. In: The Presidency. https://thepresidency.gov.za/press-statements/president-appoints-commission-fourth-industrial-revolution. Accessed 3 Jun 2022

Van den Berg A (2018) Hoe kunsmatige intelligensie die werkwêreld gaan verander (How artificial intelligence will change the world of work). In: Jou Werk / Solidariteit Wêreld. https://jouwerk.solidariteit.co.za/hoe-kunsmatige-intelligensie-die-werkwereld-gaan-verander/. Accessed 21 Jun 2022

Women Press Freedom (2023) South Africa: Crypto Scam Features Deepfakes of TV Anchors Bongiwe Zwane and Francis Herd. https://www.womeninjournalism.org/threats-all/south-africa-crypto-scam-features-deepfakes-of-tv-anchors-bongiwe-zwane-and-francis-herd. Accessed 18 Jan 2024

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Российской Федерации

Д. Ю. Базаркина, Е. Н. Пашенцев

Введение

Основные направления научных исследований и разработок в области цифровых технологий в России включают машинное обучение, человеко-машинные интерфейсы, технологии промышленного интернета, использование пространственных данных (транспортные сети) и многое другое. 10 октября 2019 г. была принята Национальная стратегия развития искусственного интеллекта до 2030 года (Президент Российской Федерации, 2019а). Показательно, что сроки реализации стратегии охватывают период в десять лет, а основные принципы развития и использования технологий ИИ включают следующие аспекты безопасности: «…недопустимость использования искусственного интеллекта в целях умышленного причинения вреда гражданам и юридическим лицам, а также предупреждение и минимизация рисков возникновения негативных последствий использования технологий искусственного интеллекта» (Президент РФ, 2020). В России приоритетное внимание уделяется угрозам информационно-психологической безопасности, реализуемых посредством ЗИИИ второго уровня, хотя стратегия оставляет пространство для маневра в отношении регулирования угроз первого и третьего уровней. В январе 2024 г. президент В. В. Путин поручил профильным ведомствам проанализировать практику использования технологий ИИ при расследовании преступлений. Верховный суд, Генеральная прокуратура, Следственный комитет, Министерство внутренних дел и Министерство юстиции должны проработать этот вопрос до 1 июля. При необходимости, эти структуры должны предоставить предложения по совершенствованию технологии (Уварчев, 2024).

Среди других государственных структур можно выделить Фонд перспективных исследований (ФПИ), который поддерживает научные исследования и разработки в области противодействия ЗИИИ. При ФПИ функционирует Национальный центр развития технологий и базовых элементов робототехники. ФПИ поддержал конкурс, в результате которого была разработана технология преобразования труднораспознаваемой (по причине фонового шума или индивидуальных особенностей говорящего) русской речи в текст. ФПИ также поддерживает проекты по интерпретации изображений, полученных со спутников и беспилотных летательных аппаратов, основным подрядчиком которых является МФТИ. В рамках такого проекта МФТИ создает технологии, направленные на борьбу с террористической угрозой посредством выявления тайников с оружием и замаскированных баз террористов по снимкам с дронов. Также в рамках ФПИ ведется работа над проектом Государственного научно-исследовательского института авиационных систем (ГосНИИАС) по выявлению угроз в социальных сетях. ГосНИИАС создал технологию, позволяющую идентифицировать предполагаемых преступников в сложных условиях – в толпе, общественном транспорте и т.д.

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

В период с февраля по апрель 2020 г. Государственная Дума Российской Федерации рассматривала законопроект об установлении экспериментального правового режима, связанного

с разработкой и внедрением технологий ИИ в г. Москва. Этот законопроект вызвал несколько неоднозначную реакцию со стороны СМИ. Комментарии – как нейтральные (Интерфакс, 2020), так и негативные (РИА «Катюша» 2020) – касались в первую очередь рисков нарушения прав граждан на неприкосновенность частной жизни. 24 апреля 2020 г. законопроект был принят в качестве федерального закона (Президент Российской Федерации, 2019b). В данной ситуации особенно важно обратить внимание на угрозы информационно-психологической безопасности первого и второго уровней. Злоупотребление применением специальных систем ИИ со стороны бизнеса может иметь место в процессе сбора персональных данных, которые могут не только передаваться государственным органам (в соответствии с законодательством), но и, например, использоваться для рассылки агрессивной таргетированной рекламы. Последовавшие негативные реакции на закон сравнивали экспериментальный режим с «концентрационным лагерем», к тому же «китайского образца» (РИА "Катюша 2020"). Исходя из этого можно предположить, что кампании в СМИ против практики применения систем ИИ как в России, так и в Китае могут проходить по аналогичной схеме. Это говорит о необходимости обеспечения российскими государственными учреждениями более широкой информированности граждан об использовании систем ИИ, особенно с учетом того, что сервисы, в которые уже внедрены алгоритмы, начинают более активно влиять на жизнь россиян. Об этом стало широко известно, когда «робот» мобильного оператора Tele2 позвонил абоненту, но вместо него трубку взял бот по имени «Олег» Тинькофф Банка. Робот Tele2 предложил «Олегу» новый тариф, на что последний согласился без ведома владельца смартфона (Гаврилюк и Королев, 2022). Несмотря на желание бизнеса ограничить регулирование систем ИИ с помощью Кодекса этики ИИ, необходимость в адекватном законодательстве, соответствующем уровню развития технологий, очевидна.

Угрозы первого уровня также составляют манипуляции страхами потери работы из-за внедрения систем ИИ. Например, в Российской Федерации инструменты ИИ начинают активно внедряться на рынке электронных книг. Так, сервис «Строки» от компании МТС и сервис LitRes используют алгоритмы для дублирования аудиокниг, что уже вызвало недовольство среди профессиональных дикторов: их профсоюз предложил Государственной Думе установить регуляторные меры технологии синтеза голоса с использованием ИИ (Юрасова и др., 2023). Согласно исследованию образовательной платформы GeekBrains (по состоянию на конец 2022 г.), более половины россиян (около 60%) имеют представление об ИИ, но только 14% из них полностью доверяют этой технологии. В опросе приняли участие более 2000 респондентов в возрасте от 18 до 55 лет из разных регионов России. Аналитики выяснили, что большая часть опасений респондентов касательно ИИ связана со страхом сокращения рабочих мест из-за развития технологий. Таким образом, 58% респондентов опасаются, что ИИ может лишить их работы, 11% респондентов уверены, что под натиском алгоритмов их профессии могут полностью исчезнуть, а 46% из них не сомневаются в том, что ИИ сможет частично взять на себя их функции (Мамиконян, 2022). Это относительно новое явление в России, где люди, как правило, не рассматривают ИИ как равноценного конкурента.

Согласно другому исследованию, проведенному компанией Superjob в марте 2023 г. (было опрошено 1600 представителей экономически активного населения из всех округов страны), развития ИИ боятся 17% россиян, причём чаще всего они опасаются лишения рабочих мест в результате автоматизации. «На втором месте по количеству упоминаний — доводы о бесконтрольности и закрытости разработки мощных компьютерных систем, на третьем — боязнь деградации человечества», — говорится в данном исследовании. Не испытывают страха, связанного с развитием ИИ, 43% россиян. Большинство из них уверены, что нейросети никогда не смогут заменить человека (RT 2023).

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Угрозы ЗИИИ (существующие и перспективные), проявившиеся в других странах БРИКС, в одинаковой степени актуальны и для России. Это было продемонстрировано в период пандемии COVID-19, когда в России участились случаи фишинга. На фоне новостей о выплатах пособий семьям с детьми стали появляться поддельные сайты с предложением подать заявление на получение пособий. В зоне .ru в 2020 г. было обнаружено около 30 новых поддельных доменов. По словам А. Дрозда, руководителя отдела информационной безопасности SearchInform, многие сайты еще не были до конца доработаны и наполнены контентом, вероятно, по причине подготовки к «отзеркаливанию» дизайна оригинального сайта, архитектуру которого они копируют (Степанова, 2020). Согласно опросу, проведенному компанией Avast, в 2021 г. 45% россиян подверглись фишинговым атакам, что на 4% больше по сравнению с результатами 2020 г. Кроме того, в 2021 г. 72% респондентов столкнулись с фишинговыми звонками (в 2020 г. этот показатель был равен 56%), 60% получали вредоносные электронные письма, и 52% сталкивались с SMS-фишингом (ТАСС, 2021). Это свидетельствует о способности злоумышленников быстро осваивать технологии ИИ и проводить деструктивную деятельность в цифровой среде. По данным Роскомнадзора, в первом квартале 2023 г. в России было удалено и заблокировано более 7,2 тыс. фишинговых ресурсов; за аналогичный период 2022 г. количество заблокированных ресурсов не превысило 2 тыс. (Исакова, 2023). «Лаборатория Касперского» предупреждает, что киберпреступники в 2024 г. будут использовать более передовые технологии, включая алгоритмы ИИ, для запуска фишинговых кампаний (Lenta, 2023a).

П/ Коростелев, руководитель отдела продвижения продуктов компании «Код безопасности», предупреждает, что с помощью языковых моделей мошенники уже повысили правдоподобность фишинговых ссылок — адресов, переход на которые чреват потерей данных или денег. Пользователи с большей вероятностью перейдут по ссылке, ведущей на страницу с идеальным текстом и вредоносным программным обеспечением, объясняет эксперт (Юрьев, 2023). Эксперты «Лаборатории Касперского» раскрыли подробности новой мошеннической схемы в популярном мессенджере Telegram. Злоумышленники заманивают людей в чат-бот, который, предположительно, работает на основе кода ChatGPT 4.0. Авторы бота утверждают, что с его помощью можно искать «слитые» фотографии человека, имея ссылку на его профиль в социальных сетях или номер телефона. Если вы запустите чат-бота, появится сообщение с предложением отправить ссылку на профиль интересующего вас человека в одной из нескольких популярных социальных сетей. После этого сервис начнет имитировать процесс работы — сначала выведет сообщение «ведется поиск», а затем выдаст результат «страница найдена в базе данных» и «материал готовится к отправке». Владельцы бота указывают предполагаемую дату «утечки» материала, а также количество обнаруженных интимных снимков с человеком. На этом этапе человек увидит несколько скриншотов, но разобрать, что на них изображено, невозможно (контент скрыт). Чтобы получить все фотографии и видео, необходимо заплатить 399 руб. за разовый доступ к базе данных или 990 рублей за неограниченный доступ. Если осуществить перевод денег, обманутый пользователь не получает никаких материалов, а средства достаются злоумышленникам (Kaspersky, 2023). По данным «Лаборатории Касперского», злоумышленники выманивают деньги под видом обмена валюты в Telegram-боте. Сейчас с помощью чат-ботов хакеры уже создают вирусы-шифровальщики и плагины для браузеров, которые могут красть пароли и данные карт (Юрьев, 2023).

В начале 2024 г. россиян начали предупреждать о новом виде мошенничества, при котором злоумышленники начали применять нейронные сети для подделки голосовых сообщений в социальных сетях. «Сегодня это уже задача, которая решается практически по щелчку пальцев. По-

тому что существуют модели ИИ, которым требуется записанная речь человека продолжительность от 3 до 20 секунд чтобы сгенерировать любой текст в любой эмоциональной тональности независимо от того, была ли она в вашем разговоре. Этот голос будет практически не отличим от оригинала даже для близких людей этого человека», - говорит генеральный директор компании-разработчика ИИ Роман Душкин (MIR24TV, 2024). «Есть еще один широко распространенный вид мошенничества, когда руководитель какой-то организации якобы начинает писать всем сотрудникам. У нас в вузе это тоже распространено, ректор МИФИ якобы пишет всем сотрудникам вплоть до самого маленького ассистента кафедры и говорит, что сейчас с ним свяжется так называемый куратор. Я предполагаю, что скоро ректор будет не просто писать, а говорить голосовыми сообщениями и использоваться голос настоящего ректора, потому что это публичная персона, его голос доступен в публичном пространстве. Это будет еще сильнее оказывать давление на людей, которые будут получать такие сообщения» (MIR24TV, 2024). Ранее сообщалось о фишинге с использованием фейкового изображения банковской карты, созданного с помощью нейронной сети (ВТБ, 2023). Таким образом, в России возрастает угроза фишинга с применением алгоритмов.

В сентябре 2021 г. мошенники разместили дипфейк-рекламу с участием основателя Тинькофф Банка Олега Тинькова. На фальшивом видео миллиардер призывает людей инвестировать и получать бонусы, перейдя по указанной ссылке. Дипфейк был опубликован на поддельной странице Тинькофф бонус в социально сети Facebook. Изображение профиля напоминало логотип банка. Согласно Fakecheck, когда пользователи переходили по ссылке под видео, они перенаправлялись на целевую страницу с логотипом банка, где должны были ответить на несколько вопросов об инвестировании и заполнить регистрационную форму с указанием своего имени, электронной почты и номера телефона (Дульнева и Милюкова, 2021). Очевидно, что подобные мошеннические схемы могут легко спровоцировать стресс и панику у обманутых людей, особенно в критической ситуации. По мере дальнейшего совершенствования дипфейков и появления более эффективных схем манипулятивного информационно-психологического воздействия с применением данных технологий угрозы будут только возрастать.

Согласно исследованию, проведенному немецкой компанией, специализирующейся на рыночных и потребительских данных в 2021 г. (Statista, 2021), более 10% граждан России регулярно пользуются умными голосовыми помощниками в повседневной жизни. Для сравнения, в США, соперничающих с Китаем за лидерство в области ИИ, этот показатель в том же году достиг 30% (Edison Research, 2022). Существует реальная угроза злонамеренного использования голосового помощника в России. Взлом голосовых помощников может привести к аналогичным угрозам, характерным для кибервоздействий на чат-ботов. Кроме того, взлом системы «умный дом» или даже простое подключение к интеллектуальному громкоговорителю с помощью этой технологии могут позволить злоумышленникам вторгаться в частную жизнь людей и перехватывать управление устройствами в их домах, тем самым влияя на их психоэмоциональное состояние.

Как и в других странах, где разрабатываются роботы на базе ИИ, существует риск попадания данной технологии в руки злоумышленников, что обуславливает необходимость для России бороться с угрозами ЗИИИ второго уровня. Например, в России набирают популярность роботы-собаки. Компания «Интеллект машин» начала производство модели робота-собаки М-81 в конце 2021 г. (TV BRICS, 2022) на базе китайских технологий (IXBT.com, 2022). Российские СМИ задаются вопросом: что произойдет, если робот будет использован злоумышленниками? Даже упоминание о такой возможности снижает доверие общественности к ИИ и робототехнике, но в целом данные разработки находят позитивный отклик общественности.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

В России одними из лидеров в разработке и внедрении технологий ИИ являются крупные банковские компании. В частности, правительство России недавно подписало соглашение о развитии ИИ с одним из крупнейших банков страны «Сбербанком» (The Russian Government, 2023). Более того, цифровой банкинг в России признан одним из самых динамичном развивающихся во всем мире (Wodzicki et al., 2020, p. 8). Учитывая эти обстоятельства, в России использование банковских чат-ботов в деструктивных целях может стать довольно опасной формой ЗИИИ, направленной на доступ к персональным данным пользователей. Проблема злонамеренного и даже террористического использования ботов, созданных не для общения, давно обсуждается в академических и профессиональных кругах. Например, установлено, что такие боты используются для манипулирования общественным мнением и нанесения репутационного ущерба, в том числе в ходе электоральных процессов (Bazarkina and Pashentsev 2019, p. 155), вербовки новых членов в преступные организации и для координации их деятельности (Mikhalevich 2022). Между тем, злоумышленники используют популярных в России чат-ботов не по назначению: логические уязвимости позволяют использовать их для кражи данных клиентов банков (Ильина 2021). Очевидно, что чат-боты могут быть просто взломаны чтобы получать информацию непосредственно от пользователей. Стоит отметить, что эта технология также используется в российской единой онлайн-системе предоставления государственных услуг гражданам под названием «Госуслуги». Несмотря на то, что утечка данных через чат-бот «Госуслуг» невозможна, на самом пике пандемии COVID-19 он все же подвергся кибератаке: преступники использовали его для дезинформации населения о коронавирусе и угроз, согласно которым вакцинированные граждане умрут (Ушков и Балашова, 2021). Этот пример наглядно иллюстрирует, что чат-боты являются уязвимой технологией, и их использование злоумышленниками может как нанести психологический вред отдельному человеку, так и повлиять на информационно-психологическую безопасность в масштабах страны.

На международном уровне обеспечение информационно-психологической безопасности в России сталкивается с деструктивными действиями американских интернет-компаний. В 2017 г. Google объявила о своем намерении понизить поисковый рейтинг репортажей российских государственных изданий Russia Today (RT) и Sputnik. Председатель правления Alphabet (холдинговой компании, владеющей Google) Э. Шмидт заявил, что поисковому гиганту необходимо бороться с распространением дезинформации; тем временем спецслужбы США называли RT «государственной пропагандистской машиной России». Соответствующие публикации в СМИ назвали этот шаг одной из форм цензуры. Выступая на Галифакском международном форуме безопасности (Вашингтон, округ Колумбия), Э. Шмидт заявил: “Мы так не работаем. Я категорически против цензуры, и решительно выступаю за ранжирование информации”. Понижение рейтинга публикаций в поисковике происходит, когда Google меняет свои алгоритмы для обнаружения «информационного оружия», под которым в данном случае Шмидт понимает публикации российских государственных СМИ (BBC News, 2017). Эти комментарии вызвали естественную ответную реакцию со стороны RT и Sputnik, основанную на заявлении Google Конгрессу США. Маргарита Симоньян, главный редактор двух представленных российских изданий, объявила, что более раннее заявление Google подтверждает: не было обнаружено каких-либо манипуляций или нарушений со стороны RT (RT, 2017). Спецслужбы США обвинили Россию в попытке повлиять на президентские выборы в США в 2016 г. в пользу Д. Трампа путем распространения фейковых новостей и взлома ресурсов Демократической партии с целью подорвать авторитет его оппонента Х. Клинтон (BBC News, 2017). Это обвинение побудило Twitter запретить рекламу RT и Sputnik на своей платформе в октябре 2017 г. В ноябре 2017 г. Министерство юстиции США вынудило RT зарегистрироваться в качестве «иностранного агента».

Создание негативного образа России в западном информационном поле по отношению к российскому руководству и целевой аудитории (граждане США) может быть оценена как угроза ЗИИИ третьего уровня. В 2020 г. в социальных сетях были размещены два дипфейка – политические рекламные ролики с изображением президента России В.В. Путина и лидера Северной Кореи Ким Чен Ына. Посыл обоих видеороликов было одинаковым: России или Северной Корее нет необходимости вмешиваться в выборы в США, так как Соединенные Штаты сами разрушат свою собственную демократию. Видеоролики были созданы и распространены американской правозащитной группой RepresentUs с целью повышения осведомленности о правах избирателей и защиты этих прав на предстоящих президентских выборах в США. Видеозаписи были опубликованы на фоне резкой публичной критики голосования по почте тогдашним президентом Дональдом Трампом, и ходили слухи, что Трамп может отказаться уступить власть после выборов, если проиграет. Цель данной инициативы, согласно сообщениям СМИ (Hao 2020), состояла в том, чтобы «не только ввергнуть американцев в шок от понимания хрупкости демократии, но и подтолкнуть их к различным действиям, таким как проверка их регистрации в качестве избирателей и добровольное участие в выборах». RepresentUs сотрудничала с креативным агентством Mischief at No Fixed Address, у которого появилась идея «использовать диктаторов для передачи послания». В заявлении в конце ролика говорилось: «Видео поддельное, но угроза существует» (Hao, 2020). Стереотипный образ «диктатора» отражает, с одной стороны, реалии информационно-психологической войны между элитами на мировой арене; с другой стороны, он также может усилить стереотипы, вселяя в лидеров других стран тревогу по поводу экономического конкурента и политического оппонента. Примечательно, что американские медиа-сети не осмелились взять на себя такую ответственность – реклама должна была выйти в эфир на Fox News, CNN и MSNBC, но решение о её публикации было отменено в последний момент.

Случаи взлома российских информационных ресурсов и использования дипфейков с участием российского президента в информационно-психологической войне показывают, что третий уровень ЗИИИ может использоваться открыто и не только преступными субъектами. В настоящее время возможности российского государства донести до мирового сообщества свою точку зрения затруднены тем фактом, что компании, развивающие наиболее популярные англоязычные социальные сети, базируются в США и находятся под влиянием антироссийски настроенных сил. Это вынуждает Россию развивать альтернативные социальные сети, которые смогут обеспечить широкий охват аудитории за пределами национальных границ. Однако, если эту ситуацию оценивать не только как угрозу, но и как возможность, расширение аудитории российских социальных сетей значительно увеличит объем больших данных для обучения отечественного ИИ.

Оценка роста угроз ЗИИИ в России невозможна без учета внешних рисков в этой области. Крупные технологические компании США продемонстрировали, что они выступают в качестве мощного инструмента проведения кибероперациий, направленных против России. Б. Смит, президент Microsoft, недвусмысленно заявляет о роли упомянутой компании в конфликте на Украине. «Правительство Украины успешно поддержало свои гражданские и военные операции, быстро приняв меры по переводу своей цифровой инфраструктуры в публичное облако в центры обработки данных по всей Европе. Это потребовало срочных и экстраординарных шагов со стороны всего технологического сектора, в том числе со стороны Microsoft. Хотя работа технологического сектора была жизненно важной, также необходимо подумать о более долгосрочных уроках, которые можно извлечь из этих усилий» (Microsoft, 2022).

Генерал армии США и директор Агентства национальной безопасности (АНБ) П. Накасоне в своем интервью Sky News от июня 2022 г. подтвердил, что США проводили наступательные кибероперации в поддержку Украины: «Мы провели серию операций …, включая наступательные, защитные и информационные» (Martin, A., 2022).

Использование западных технологий ИИ в продолжающемся военном конфликте на Украине имеет большое значение. Американский стартап по распознаванию лиц «Clearview AI» оказал техническую поддержку Украине. Инструменты Clearview AI могут идентифицировать лица на видео, сравнивая их с базой данных компании, содержащей 20 миллиардов изображений из публичных сетей, и затем выявлять потенциальных шпионов и убитых людей. Инструменты ИИ также играют важную роль в пропагандистской войне на Украине и в обработке важной информации о конфликте. Программа от американской компании Primer способна распознавать речь, осуществлять транскрипцию и перевод. Она перехватывает и анализирует российские данные, включая разговоры между российскими солдатами на Украине. Швейцарский зашифрованный чат-сервис Threema позволяет украинским пользователям отправлять эти данные военным, при этом оставаясь анонимным (Global Times, 2022). 5 июня 2023 г. в российских регионах некоторые радиостанции и телеканалы транслировали фальшивое обращение от имени президента В.В. Путина, согласно которому в трех регионах вводится военное положение, а также объявляется мобилизация. Эти утверждения оказались ложными, а видео – дипфейком (Лента, 2023b), что, безусловно, открывает новый этап информационно-психологического противостояния в рамках украинского конфликта.

Заключение

Россия сталкивается с внутренними и внешними угрозами информационно-психологической безопасности посредством ЗИИИ. Причем последние явно усиливаются по мере роста международной напряженности, проведения США и их союзниками активной гибридной войны против России. Очевидно, что с развитием систем ИИ в различных государствах вероятность использования конкретных ИИ-технологий злонамеренного воздействия в незаконных целях возрастает. Таким образом, представляется целесообразным развивать международное сотрудничество с целью совместной разработки мер противодействия ЗИИИ, в том числе тех, при которых используются персональные данные, угрожающие безопасности всех стран. К сожалению, такое сотрудничество в крайне напряженной международной обстановке представляется возможным реализовать не на глобальном уровне, а в рамках влиятельных международных объединений и организаций в которых Россия играет важную роль , таких как БРИКС, ШОС, ЕАЭС и ОДКБ и др., не закрывая двери для соглашений в более широких форматах при более более благоприятной обстановке. Кроме того, межгосударственное сотрудничество также необходимо для определения взаимосвязи между персональными данными и ИИ и установления междисциплинарных стандартов. Более того, крайне важно не только определить случаи, в которых использование персональных данных для ИИ будет рассматриваться как нарушение прав, но и разработать меры по их защите, вплоть до ограничения использования и дальнейшего развития ИИ при определенных условиях.

Литература

Bazarkina D, Pashentsev E (2019) Artificial Intelligence and New Threats to International Psychological Security. Russia in Global Affairs. doi: 10.31278/1810-6374-2019-17-1-147-170

BBC News (2017) Google to 'derank' Russia Today and Sputnik. In: BBC News. https://www.bbc.com/news/technology-42065644. Accessed 21 Jun 2022

Edison Research (2022) The Smart Audio Report. In: NPM. https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/. Accessed 19 Jan 2024

Global Times (2022) From commercial satellites to social media, Western tech companies are deeply involved in the Russia-Ukraine conflict. In: Teller Report. https://www.tellerreport.com/news/2022-11-02-from-commercial-satellites-to-social-media--western-tech-companies-are-deeply-involved-in-the-russia-ukraine-conflict.HJSuXB1Bo.html. Accessed 19 Jan 2024

Hao K (2020) Deepfake Putin is here to warn Americans about their self-inflicted doom. In: MIT Technology Review. https://www.technologyreview.com/2020/09/29/1009098/ai-deepfake-putin-kim-jong-un-us-election/. Accessed 22 Jun 2022

IXBT.com (2022) В России представили робособаку с гранатомётом. IXBT.com. https://www.ixbt.com/news/2022/08/15/v-rossii-predstavili-robosobaku-s-granatometom.html. Дата обращения: 20.01.2024

Kaspersky (2023) Обмену и возврату не подлежит: злоумышленники выманивают деньги под видом обмена валюты в Telegram-боте. Kaspersky. https://www.kaspersky.ru/about/press-releases/2023_obmenu-i-vozvratu-ne-podlezhit-zloumyshlenniki-vymanivayut-dengi-pod-vidom-obmena-valyuty-v-telegram-bote?ysclid=lserd46yap191790756. Дата обращения: 10.02.2024

Martin A (2022) US military hackers conducting offensive operations in support of Ukraine, says head of Cyber Command. In: Sky News. https://news.sky.com/story/us-military-hackers-conducting-offensive-operations-in-support-of-ukraine-says-head-of-cyber-command-12625139. Accessed 19 Jan 2024

Microsoft (2022) Defending Ukraine: Early Lessons from the Cyber War. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE50KOK. Accessed 19 Jan 2024

Mihalevich E (2022) Malicious use of artificial intelligence was discussed at UNESCO Conference in Khanty-Mansiysk. In: RIAC. https://russiancouncil.ru/analytics-and-comments/columns/cybercolumn/zlonamerennoe-ispolzovanie-iskusstvennogo-intellekta-obsudili-na-konferentsii-yunesko-v-khanty-mansi/. Accessed 19 Jan 2024

RT (2017) Google will “de-rank” RT articles to make them harder to find—Eric Schmidt. In: RT International. https://www.rt.com/news/410444-google-alphabet-derank-rt/. Accessed 21 Jun 2022

RT (2023) Больше всего в развитии искусственного интеллекта россиян пугает риск потери работы. https://ru.rt.com/o875 Accessed 28 Nov 2023.

Statista Research Department (2021) Voice assistants home usage in Russia 2021. https://www.statista.com/statistics/1258819/voice-assistants-home-usage-russia/. Accessed 19 Jan 2024

TV BRICS (2022) В России развивается рынок робособачества. TV BRICS. https://tvbrics.com/news/v-rossii-razvivaetsya-rynok-robosobachestva/. Дата обращения: 23.01.2024

Wodzicki M, Majewski M, MacRae M (2020) Digital Banking Maturity 2020. In: Deloitte. https://www2.deloitte.com/content/dam/Deloitte/ce/Documents/financial-services/ce-digital-banking-maturity-2020.pdf. Accessed 19 Jan 2024

Гаврилюк А, Королев Н (2022*)* Граждан защитят от роботов. Коммерсантъ. https://www.kommersant.ru/doc/5173457. Дата обращения: 22.01.2024

Дульнева М, Милюкова Я (2021) «Всех обнял!»: образ Олега Тинькова использовали в дипфейк-рекламе. Forbes. https://www.forbes.ru/milliardery/439255-vseh-obnal-obraz-olega-tinkova-ispol-zovali-v-dipfejk-reklame. Дата обращения: 21.01.2024

Ильина Н (2021) Обман по переписке: уязвимости в банковских чат-ботах позволяют красть деньги. Известия. https://iz.ru/1214668/natalia-ilina/obman-po-perepiske-uiazvimosti-v-bankovskikh-chat-botakh-pozvoliaiut-krast-dengi. Дата обращения. 19.01.2024

Интерфакс (2020) Госдума одобрила для Москвы особый правовой режим для развития искусственного интеллекта. Интерфакс. https://www.interfax.ru/russia/704092. Дата обращения: 21.01.2024

Исакова, Т. (2023) Хакеры нащупали точку роста. Коммерсантъ. https://www.kommersant.ru/doc/5977758. Дата обращения: 20.01.2024

Коммерсантъ (2023) ВТБ: мошенники выманивают средства клиентов в Телеграме с помощью карт с поддельным дизайном. Коммерсантъ. https://www.kommersant.ru/doc/6111196. Дата обращения: 20.01.2024

Lenta (2023a) Россиян предупредили о самых опасных хакерских атаках в 2024 году. Лента. https://lenta.ru/news/2023/11/14/rossiyan-predupredili-o-samyh-opasnyh-hakerskih-atakah-v-2024-godu/?ysclid=lrm9958719396125920. Дата обращения: 20.01.2024

Lenta (2023b) «Обращение» Путина о мобилизации и военном положении оказалось дипфейком. Лента. https://lenta.ru/news/2023/06/05/fake_radio/?ysclid=lrm69q0hln874723542. Дата обращения: 20.01.2024

Мамиконян О (2022) 58% россиян боятся сокращения рабочих мест из-за развития искусственного интеллекта. Forbes. https://www.forbes.ru/forbeslife/481810-58-rossian-boatsa-sokrasenia-rabocih-mest-iz-za-razvitia-iskusstvennogo-intellekta. Дата обращения: 20.01.2024

МИР24 (2024) Как искусственный интеллект используют мошенники в России? МИР24. https://mir24.tv/articles/16577222/kak-iskusstvennyi-intellekt-ispolzuyut-moshenniki-v-rossii?ysclid=lrkoid1382919794190. Дата обращения: 20.01.2024

Правительство России (2023) В Правительстве подписан финальный пакет соглашений о сотрудничестве по развитию высокотехнологичных направлений. Правительство России. http://government.ru/news/47551/. Дата обращения: 19.01.2024

Президент Российской Федерации (2019a) Указ Президента Российской Федерации от 10.10.2019 № 490 "О развитии искусственного интеллекта в Российской Федерации". Портал правовой информации. http://publication.pravo.gov.ru/Document/View/0001201910110003. Дата обращения: 22.01.2024

Президент Российской Федерации (2019b) Федеральный закон № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона "О персональных данных"». Портал правовой информации. http://pravo.gov.ru/proxy/ips/?docbody=&prevDoc=102108261&backlink=1&&nd=102722375. Дата обращения: 21.01.2024

Президент Российской Федерации (2020) Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации». Портал правовой информации. http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=102608394. Дата обращения: 20.01.2024

РИА «Катюша» (2020) Здравствуй, электронный концлагерь китайского образца: Собянин хочет отдать москвичей под управление искусственному интеллекту. https://katyusha.org/oczifrovka/159866-zdravstvuj-elektronnyij-konczlager-kitajskogo-obrazcza-sobyanin-xochet-otdat-moskvichej-pod-upravlenie-iskusstvennomu-intellektu.html. Accessed 19 May 2024

Степанова Ю (2020) Родители повелись как дети. Коммерсантъ. https://www.kommersant.ru/doc/4343398. Дата обращения: 21.01.2024

ТАСС (2021) Эксперты: с фишинговыми атаками в 2021 году столкнулись 45% россиян. ТАСС. https://tass.ru/ekonomika/13105631. Дата обращения: 22.01.2024

Уварчев Л (2024) Путин поручил изучить применение искусственного интеллекта в расследования. Коммерсантъ. https://www.kommersant.ru/doc/6454973?ysclid=lrkljlwu1i347691345. Дата обращения: 20.01.2024

Углова Ю (2023) Россияне теряют деньги из-за «умного» чат-бота. Hi-Tech. https://hi-tech.mail.ru/news/100835-rossiyane-teryayut-dengi-iz-za-umnogo-chat-bota-podrobnosti/. Дата обращения: 10.02.2024

Юрасова Ю, Тишина Ю, Петрова В (2023) Горе от интеллекта. Коммерсантъ. https://www.kommersant.ru/doc/5928661. Дата обращения: 20.01.2024

Юрьев Д (2023) Не мытьем, так плагином: как мошенники используют чат-боты в России. Ferra.ru. https://www.ferra.ru/news/v-rossii/ne-mytem-tak-plaginom-kak-moshenniki-ispolzuyut-chat-boty-v-rossii-10-05-2023.htm?ysclid=lser17n5lg203176776. Дата обращения: 10.02.2024

Юшков М, Балашова А (2021) Власти заявили об атаке на «Госуслуги» после сообщений о боте-антиваксере. РБК. https://www.rbc.ru/technology_and_media/11/11/2021/618d42109a7947252fe7d448. Дата обращения: 19.01.2024

Злонамеренное использование искусственного интеллекта: вызовы информационно-психологической безопасности в Объединенных Арабских Эмиратах

Е. Н. Пашенцев, В. А. Чебыкина, Р. Т. Никифоров

Введение

На протяжении последних двух десятилетий Объединенные Арабские Эмираты начали активно развивать свои технологические возможности, включая разработки ИИ. Активный рост числа стартапов и высокий уровень инвестиций в эту область говорят о значительной заинтересованности граждан и правительства страны во внедрении ИИ в ключевые сферы жизнедеятельности государства. Согласно отчету компании Google, “Будущее искусственного интеллекта на Ближнем Востоке и Северной Африке”, ожидается, что ежегодный экономический рост, вызванный внедрением технологий ИИ, достигнет 20-34% в год во всех странах региона Ближнего Востока и Северной Африки, причем самые высокие показатели ожидаются в ОАЭ и Саудовской Аравии. В дальнейшем влияние ИИ на экономическое развитие в регионе, скорее всего, возрастет еще больше: согласно последнему исследованию The Economist Intelligence Unit (EIU), среди стран региона только Саудовская Аравия и ОАЭ получат $200 и 120 млрд. соответственно (The Economist Group 2022).

ОАЭ – первая страна на Ближнем Востоке, которая создала Министерство искусственного интеллекта в 2017 г. и приняла Национальную стратегию по искусственному интеллекту до 2031 г. Стратегия представляет собой план по обеспечению сотрудников компаний всеми необходимыми навыками, которые смогут помочь им ориентироваться в постоянно изменяющемся технологическом мире. В 2019 г. в столице ОАЭ состоялось открытие первого в мире Университета искусственного интеллекта (англ. Mohamed bin Zayed University of Artificial Intelligence, MBZUAI) с целью разработки необходимой экосистемы ИИ для использования его потенциала на всех уровнях (Zaatari 2019). Массовый приток технических специалистов с 2021 г. помогает стимулировать стремления страны Персидского залива в области развития ИИ. Как отмечал государственный министр ОАЭ по вопросам искусственного интеллекта, цифровой экономики и приложений для удаленной работы Омар аль-Олама в интервью агентству Bloomberg на Всемирном правительственном саммите в Дубае в феврале 2024 г., по состоянию на сентябрь 2023 г. в сфере ИИ или связанных с ним отраслях работало 120 000 человек (по сравнению с 30 000 двумя годами ранее) (Omar 2024). Более 50% работающих людей в ОАЭ применяют ИИ в процессе своей работы в таких сферах, как медиа, образование, медицина, банковский сектор и др. Правительство страны активно внедряет технологии ИИ повсеместно благодаря стратегическому государственно-частному партнерству.

Однако, наряду с положительными аспектами развития ИИ, растущие вызовы представляет практика его злонамеренного использования, которая набирает силу в стране, хотя и имеющей высокий уровень социально-политической стабильности, но располагающейся в регионе с высоким уровнем конфликтности и с потенциалом перерастания локальных конфликтов в большую ближневосточную войну или даже "Третью мировую".

Первый уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

С развитием и распространением технологий ИИ и внедрения автоматизированных систем в повседневную жизнь общества ОАЭ основной угрозой первого уровня являются целевые спекуляции на фоне роста опасений касательно потери миллионов рабочих мест. По данным отчета аудит-консалтинговой корпорации PwC “2023 Workforce Hopes and Fears Survey”, более трети опрошенных в ОАЭ отмечают положительное воздействие ИИ на производительность труда (PwC Press Release 2023). Однако, 52% участников опроса резонно полагают, что им потребуется прохождение дополнительных курсов повышения квалификации из-за того, что характер их работы значительно изменится в ближайшие пять лет или вовсе потеряет свою актуальность (Abbas W, 2023a).

В Национальной стратегии ОАЭ в качестве одного из положений декларируется стремление предоставить населению возможность пройти программы переквалификации, например, путем специализированного обучения или прохождения программ международных стажировок. Правительство ОАЭ финансирует программы STEM-образования для создания кадрового резерва в области инноваций, включая бесплатные курсы для желающих повысить свои компетенции в области ИИ (Alrahmah and Ahmed 2024). Вместе с тем, этого часто оказывается недостаточно для предотвращения роста опасений людей по поводу того, останутся ли они конкурентоспособными на рынке труда. В исследовании, проведенном консалтинговой компании по стратегическим коммуникациям Duke+mir в сотрудничестве с YouGov, респондентам был задан вопрос, как, по их мнению, ИИ повлияет на их жизнь. 55% выразили озабоченность тем, что к 2033 г. их рабочие места будут заняты ИИ или роботами. Примерно 24% опрошенных не были уверены, а 21% вовсе не беспокоились насчет того, что их обязанности будут выполнять технологии ИИ. Примечательно, что 66% людей в возрасте до 25 лет опасались, что автоматизация затронет их рабочие места в следующем десятилетии, по сравнению с 57% среди людей в возрастной группе от 25 до 44 лет, в то время как 43% в возрасте от 45 лет и старше были озабочены этим вопросом (Webster 2023a). Дж. Айвен-Дьюк, соучредитель и партнер Duke+mir, прокомментировал результаты опроса: “Учитывая такое пристальное внимание правительства ОАЭ к обеспечению и защите рабочих мест в стране как сейчас, так и в будущем, довольно неожиданно увидеть, что молодежь и жители ОАЭ больше всего обеспокоены будущими технологическими разработками» (Webster, N., 2023a).

По-видимому, основной причиной беспокойства о возможной потере работы в связи с внедрением технологий ИИ является их всеобъемлющий характер, который затрагивает (или затронет по мере совершенствования технологий) практически все профессии. И, чем лучше люди информированы о возможностях ИИ, тем выше они оценивают риски потери работы. Эксперты вносят свою весомую лепту в осознание этих рисков. По словам Шалини Вермы, генерального директора Pivot Technologies, “Искусственный интеллект в его генеративной версии полностью преобразит отдельные профессии, заместив почти все должности начального уровня. Любой, кто выйдет на рынок труда, начнет с гораздо более высокого уровня, чем ожидается – начиная от стажеров и младших руководителей сегодня. Однако замена человека машиной на этом не закончится. Если вы не являетесь кем-то исключительным, вы подвергаетесь серьезному риску быть замененным ботом. Если вы эксперт, то, скорее всего, будете выполнять свою работу совсем по-другому, поскольку основные задачи будут взяты на себя ИИ” (Abbas W 2023b). Сама формулировка “исключительный” заставляет большинство не «исключительных» специалистов со все большей тревогой смотреть в будущее.

Использование ИИ на национальном уровне в ОАЭ может стать средством манипулирования сознанием широких слоев населения, особенно молодежи, что способно привести к возможным негативным последствиям. Это может произойти в ОАЭ даже раньше, чем во многих других странах именно в силу общественно-необходимого, но и имеющего свои риски быстрого развития и внедрения технологий ИИ. Руководство ОАЭ признает всю сложность данной ситуации. Вышеупомянутый государственный министр ОАЭ по ИИ Омар аль-Олама, выступая на Дубайской ассамблее по вопросам генеративного ИИ, проходившей в октябре 2023 г., призвал граждан не бояться потери работы, а сосредоточить свое внимание на расширении возможностей положительных аспектов технологий ИИ (Awienat 2023). Использование ИИ в ОАЭ может принести пользу всем гражданам, включая и работников, и компании, при условии развития ответственного и, в чем-то, опережающего подхода к системе обучения и навыков в применении ИИ. В последующем это может открыть новые возможности совершенствования человека и развития все более сложных форм гибридного интеллекта.

Второй уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Второй уровень угроз представляет собой угрозы, реализуемые посредством технологий ИИ, направленные на объекты критической инфраструктуры, физическую безопасность человека и нанесение ущерба его имуществу и благосостоянию. Согласно составленному Международным союзом электросвязи рейтингу кибербезопасности, из 194 стран (Global Cybersecurity Index 2020) ОАЭ заняли 5 место, однако страна по-прежнему сильно страдает от кибератак: многие компании заплатили более $1,4 млн выкупа, 42% из них были вынуждены закрыться после произошедшего, а 90% подверглись повторным атакам. Правительство ОАЭ в свою очередь вкладывает значительные средства в кибербезопасность, но не всегда успевает равномерно реализовать все меры, что делает местные компании менее устойчивыми к масштабным киберинцидентам (Filatov 2021).

В 2022 г. было зарегистрировано рекордное количество уязвимостей – более 26 тыс., выявленных в соответствии с Национальной базой данных уязвимостей NIST (NVD). Согласно отчету Infoblox о глобальном состоянии кибербезопасности за 2023 г., две трети (66%) респондентов из ОАЭ сообщили об одном или нескольких нарушениях в их организации в результате кибератак. Фишинг был наиболее распространенным методом атаки на организации, подвергшиеся взлому, – на его долю пришлось 62% атак в прошлом году. За ним следуют APT-атаки[18] (53%) и программы-вымогатели (51%) (Bandyopadhyay S 2023). В среднем организации ОАЭ выявили больше проблем, возникающих в результате атак по электронной почте, и фишинга, чем любого другого типа.

Согласно отчету компании Proofpoint Inc, занимающейся вопросами информационной безопасности, за 2022 г. примерно 2/3 предприятий (64%) в ОАЭ подверглись атаке с использованием так называемых «схем-вымогателей» (Ryan P 2023). Так, в 2023 г. сотрудница компании в Дубае получила сообщение в WhatsApp по уже известной схеме – злоумышленник представился

[18] APT-атака (Advanced Persistent Threat) – это целевая продолжительная атака повышенной сложности, задача которой является обнаружение на устройстве пользователя секретной, конфиденциальной или любой ценной информации и использование ее в интересах киберпреступников. Хакер проникает в вашу компьютерную сеть и проводит в ней много времени, отслеживая перемещение данных, действия основных пользователей и важную информацию.

ее директором, загрузил в профиль соответствующую фотографию и тем самым выманил у сотрудницы почти $4000 «на покупку сертификатов для их клиента» (Nair D 2023). Сотрудницу не смутил чужой номер, она доверилась фотографии и рассказам злоумышленника о том, что телефон директора просто разрядился.

Одной из главных угроз на втором уровне является кибератака на объекты критической инфраструктуры – электростанции, системы водоснабжения или системы транспорта. В 2018 г. в стране произошли две серьезные утечки данных, в результате которых «были скомпрометированы 14 млн. записей» (Chandra, Sharma and Ali 2019). Две утечки данных были зафиксированы, в частности, в дубайской платформе для поездок на автомобиле Careem. Использование ИИ позволяет злоумышленникам автоматизировать и оптимизировать свои действия, что делает такие атаки более эффективными и опасными. Например, злоумышленники могут использовать ИИ для поиска и эксплуатации уязвимостей в системах управления электростанциями, что может привести к отключению электроэнергии в крупных городах.

Физическая безопасность человека также относится к угрозам второго уровня. Например, автономные дроны, оснащенные оружием и запрограммированные на атаку, могут представлять угрозу для массовых мероприятий или мест скопления людей. С учетом того, что применение дронов, оснащенных ИИ, растет, и они проходят проверку боем в горячих точках планеты, включая Ближний Восток, следует ожидать роста их применения злонамеренными акторами и в ОАЭ. По словам А. Аль-Хоори, старшего вице-президента по стратегии и передовому опыту оборонного конгломерата ОАЭ Edge, «...в конце концов, у пользователя будет система, которая может работать автономно, которая может принимать решения за конечного пользователя» в условиях постоянно меняющегося ландшафта автономных оборонных технологий под натиском развития генеративного ИИ (Combs 2024).

Помимо физической безопасности, ИИ может использоваться для нанесения ущерба имуществу и благосостоянию человека. Например, злоумышленники могут использовать ИИ для создания мошеннических финансовых схем, взлома систем онлайн-банкинга или кражи личных данных. Новые вредоносные троянские программы способствовали увеличению числа банковских атак в первом квартале 2023 г. по сравнению с первым кварталом 2022 г. В целом, в первом квартале 2023 г. в регионе Ближнего Востока также наблюдался рост числа троянских атак на банковские услуги. В ОАЭ наблюдался рост на 67% (El-Din 2023). Такие действия могут привести к значительным финансовым потерям и нарушению приватности человека.

Вышеназванные и прочие риски критической инфраструктуре, жизни и имуществу людей подразумевают и определенные психологические последствия – как непосредственные, так и те, которые наступят в перспективе, что может оказать контролируемое и неконтролируемое негативное воздействие на сознание людей. Так, например, атака дронов на основе ИИ может сопровождаться как спонтанной реакцией (чувства страха, ненависти, панические состояния и т.д.), которые на третьем уровне угроз могут дополняться формированием информационной повестки дня с помощью технологий ИИ и соответствующими информационными кампаниями.

Третий уровень угроз информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта

Несмотря на существование национальных законов, запрещающих кибербуллинг, создание дипфейков и публикацию контента, содержащего фейковые новости, ОАЭ, как и многие дру-

гие страны, сталкиваются со вспышкой распространения фейков, которые могут разрушить политическую систему государства, а также нанести ущерб частной жизни людей, оказать влияние на общественное сознание.

В отчете Sumsub Identity Fraud Report 2023 подтверждается существование этой новой опасности и указывается на то, что в период с 2022 по 2023 гг. глобальный объем создания дипфейков вырос в геометрической прогрессии (более чем в 10 раз). На Ближнем Востоке и в Африке рост составил 450%. Также отмечается, что паспорт ОАЭ стал самым подделываемым документом в мире, поэтому распространение дипфейков является актуальной проблемой для страны, требующей принятия незамедлительных действий (Wassi 2023).

В 2020 г. злоумышленники позвонили менеджеру филиала японской компании и представились директором, который возглавляет фирму в ОАЭ. Менеджер, не заподозрив неладное, совершил запрашиваемый денежный перевод размером в $35 млн (Brewster 2021). Мошенники использовали технологию “deep voice”, чтобы сымитировать речь директора. В ходе расследования было выяснено, что в данном преступлении принимали участие не менее 17 человек – украденные деньги были переведены на счета в разных точках мира. Эксперты уверены, что “deep voice” дает большие эффекты при манипулировании и гораздо доступнее в плане его создания, следовательно, количество таких преступлений будет расти с каждым днем, подвергая компании и обычных людей риску. Тем не менее, некоторые фирмы, такие как Pindrop, признают потенциал злонамеренного использования ИИ, и разрабатывают программы, которые могут идентифицировать синтезированные голоса и в дальнейшем предотвращать распространение дипфейков.

В 2023 г. другим громким делом в ОАЭ стал случай с телефонным звонком «друга» своему товарищу из Кералы (Индия). Мужчина использовал сфабрикованные аудио- и видеозвонки для того, чтобы заполучить тысячи дирхамов, ссылаясь на проблемы со здоровьем. 73-летний пострадавший не заподозрил фейки, так как злоумышленник применил тактику раскрытия подробностей из личной жизни своего товарища, тем самым пытаясь получить доверие со стороны жертвы (Sankar 2023). Этот случай также иллюстрирует, что злоумышленники используют и будут использовать самые различные методы с целью оказания информационно-психологического воздействия на свою жертву, будь то компания или отдельно взятый человек.

Другой немаловажной проблемой является использование генеративного ИИ для создания материалов непристойного характера и дальнейшего воздействия на сознание детей. По данным исследований компании WeProtect, работающей в сфере защиты детей от сексуального насилия, число сообщений о груминге с целью получения денежной выгоды увеличилось со 139 в 2021 г. до более чем 10 000 в 2022 г. (Webster 2023b). Рост числа преступлений такого характера эксперты связывают с популярностью социальных сетей и развитием технологий ИИ. Чаще всего жертвами виртуального сексуального насилия становятся мальчики-подростки, когда злоумышленники представляются молодыми девушками, рассылая фейковые фото- и видеозаписи непристойного характера, получая в ответ реальные материалы, а затем требуют деньги за молчание перед родителями.

Развитие ИИ привело к созданию и распространению религиозных чат-ботов. ОАЭ вошли в топ-10 стран, пользователи которых прибегают к чат-боту QuranGPT (Prabhakar A 2023). В данном вопросе существуют две проблемы: Первая, это предвзятость ИИ, который может неправильно интерпретировать запросы пользователей, что приведет как минимум к дезинформации, а как максимум – к этническим проблемам в обществе.Другая проблема в том, что религиозные чат-

боты могут попасть в руки злоумышленников, которые будут использовать их для распространения заведомо ложной информации или пропаганды, а также для разжигания межэтнических конфликтов. Для того, чтобы избежать этих проблем, разработчикам необходимо быть особо внимательными к подбору источников информации, правильному интерпретированию учений религии и обеспечении безопасности системы.

Распространение дипфейков порождает проблему снижения доверия к источникам информации внутри общества. Правительство ОАЭ хочет помочь общественности научиться распознавать дипфейки, опубликовав руководство, призванное повысить осведомленность о вредном и полезном применении дипфейк-технологий. Выявить дипфейк можно и самостоятельно, но, эксперты уверены, что по мере совершенствования технологий это будет все сложнее сделать. Со временем единственным способом определить подлинность информации станет сам ИИ. В руководстве говорится, что: "Наиболее точный подход к обнаружению фейковой информации – это систематическая проверка дипфейков с помощью инструментов на основе ИИ, которые необходимо регулярно обновлять" (The National 2021). Такие инструменты могут анализировать текст, изображения, аудио- и видеофайлы на предмет признаков манипуляции или искажения. Данный подход поможет повысить точность обнаружения фейков и защитить пользователей от их негативного воздействия.

Заключение

В ОАЭ существуют различные уровни угроз, связанные с технологиями ИИ. На первом уровне присутствуют риски намеренно искаженного толкования развития ИИ в интересах антисоциальных групп, что может привести к социально-экономическим конфликтам. Подобные интерпретации еще не стали значимым фактором общественной жизни в ОАЭ, но следует ожидать попыток злонамеренных акторов использовать негативные последствия, и даже важные достижения в развитии ИИ отрасли в своих целях, особенно, в условиях крайне сложной и опасной обстановки на Ближнем Востоке и в глобальном измерении. На втором уровне угрозы связаны с возможностью осуществления кибератак на объекты критической инфраструктуры и нанесения ущерба людям и их имуществу с сопутствующими информационно-психологическими последствиями. Число таких атак стремительно растет, как и роль технологий ИИ в их реализации. На третьем уровне существует опасность деструктивного информационно-психологического воздействия, включая использование дипфейков и чат-ботов для манипулирования общественным сознанием. Угрозы разных уровней требуют пристального внимания и принятия соответствующих мер для обеспечения информационно-психологической безопасности в ОАЭ.

Литература

Abbas W (2023 a) UAE: Will automation replace jobs or help employees gain new skills? In: Khaleej Times. https://www.khaleejtimes.com/jobs/uae-is-ai-a-threat-to-jobs-how-employees-can-use-tech-to-boost-hiring-chances-get-shortlisted-by. Accessed 24 Jan 2024

Abbas W (2023 b) Jobs in UAE: Entry level roles set to be wiped out by AI, automation and ChatGPT. In: Khaleej Times. https://www.khaleejtimes.com/jobs/jobs-in-uae-entry-level-roles-set-to-be-wiped-out-by-ai-automation-and-chatgpt. Accessed 24 Jan 2024

Alrahmah B, Ahmed M A (2024) The UAE's harnessing of AI at the national level can benefit everyone. In: The National News. https://www.thenationalnews.com/opinion/comment/2024/01/16/the-uaes-harnessing-of-ai-at-the-national-level-can-benefit-everyone/. Accessed 24 Jan 2024

Awienat D (2023) Generative AI should not be feared despite risks, says UAE minister of artificial intelligence. In: Arab news. https://www.arabnews.com/node/2390716/middle-east. Accessed 24 Jan 2024

Bandyopadhyay S (2023) UAE Cybersecurity: 26,000 vulnerabilities were reported in 2022. In: Khaleejtimes. https://uaetimes.ae/uae-cybersecurity-26000-vulnerabilities-reported-in-2022-news/. Accessed 2 Feb 2024

Brewster T (2021) Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find. In: Forbes. https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=442e5dd67559. Accessed 27 Jan 2024

Chandra G R, Sharma B K, Ali I (2019) UAE's Strategy Towards Most Cyber Resilient Nation. In: International Journal of Innovative Technology and Exploring Engineering. https://www.researchgate.net/publication/337146109_UAE%27s_Strategy_Towards_Most_Cyber_Resilient_Nation. Accessed 2 Feb 2024

Combs C (2024) AI has changed defence industry expectations, says Edge executive. In: The National News. https://www.thenationalnews.com/business/future/2024/01/25/uae-ai-edge-umex/. Accessed 2 Feb 2024

Devi A (2021) UAE ranks 5th in UN's 2020 Global Cybersecurity Index. In: Security Middle East&Africa. https://securitymea.com/2021/07/01/uae-ranks-5th-in-uns-2020-global-cybersecurity-index/. Accessed 2 Feb 2024

El-Din M A (2023) 49% increase in phishing attacks in Egypt during 1Q 2023: Kaspersky. In: Daily News Egypt. https://www.dailynewsegypt.com/2023/05/07/49-increase-in-phishing-attacks-in-egypt-during-1q-2023-kaspersky/. Accessed 2 Feb 2024

Filatov A (2021) Russia shared the fifth place in the GCI cybersecurity rating with Malaysia and the UAE. In: Digital Russia. https://d-russia.ru/rossija-razdelila-s-malajziej-i-oaje-pjatoe-mesto-v-rejtinge-kiberbezopasnosti-msje.html. Accessed 2 Feb 2024

Nair D (2023) The Dept Panel: 'I lost $4,000 in a WhatsApp scam'. In: The National News UAE. https://www.thenationalnews.com/business/money/2023/08/17/the-debt-panel-i-lost-4000-in-a-whatsapp-scam/. Accessed 27 Jan 2024

Prabhakar A (2023) Religious GPT: The chatbots and developers fighting bias with AI. The National News UAE. https://www.thenationalnews.com/weekend/2023/07/28/religious-gpt-the-chatbots-and-developers-fighting-bias-with-ai/. Accessed 27 Jan 2024

PwC Press Release (2023) Workforce in the Middle East is ambitious, enthusiastic about change, embracing AI and upskilling, says new PwC report. https://www.pwc.com/m1/en/media-centre/2023/workforce-in-the-middle-east-is-ambitious-enthusiastic-about-change-embracing-ai-upskilling.html. Accessed 24 Jan 2024

Ryan P (2023) How the future of cyber crime could involve fake voice messages from loved ones. In: The National News UAE. https://www.thenationalnews.com/uae/2023/03/17/how-the-future-of-cybercrime-could-involve-fake-voice-messages-from-loved-ones/. Accessed 27 Jan 2024

Sankar A (2023) Deepfake video call said to be from Dubai used to swindle Kerala man out of thousands. In: The National News UAE. https://www.thenationalnews.com/uae/2023/07/19/deepfake-video-call-pretending-to-be-dubai-friend-used-to-swindle-man-out-of-thousands/. Accessed 27 Jan 2024

The Economist Group (2022) Pushing forward: the future of AI in the Middle East and North Africa. In: Economist Impact. P. 5, 55. https://impact.economist.com/perspectives/sites/default/files/google_ai_mena_report.pdf. Accessed 27 Jan 2024

The National (2021) UAE asks public to help tackle deepfakes. In: The National News UAE. https://www.thenationalnews.com/uae/2021/07/09/uae-asks-public-to-help-tackle-deepfakes/. Accessed 27 Jan 2024

Wassi E (2023) Fraude de identidade y deepfakes: suenan las alarmas. In: La Prensa. https://www.laprensa.com.ar/Fraude-de-identidad-y-deepfakes-suenan-las-alarmas-538397.note.aspx. Accessed 26 Jan 2024

Webster N (2023 a) Rise of AI creates job worries, UAE survey finds. In: The National News UAE. https://www.thenationalnews.com/uae/2023/01/19/rise-of-ai-creates-job-worries-uae-survey-finds/. Accessed 24 Jan 2024

Webster N (2023 b) Online gaming poses alarming threat to children's safety, report finds. In: The National News UAE. https://www.thenationalnews.com/uae/2023/10/19/online-gaming-poses-alarming-threat-to-childrens-safety-report-finds/. Accessed 24 Jan 2024

Zaatari S (2019) University of Artificial Intelligence launched in Abu Dhabi. In: Gulf News. https://gulfnews.com/uae/university-of-artificial-intelligence-launched-in-abu-dhabi-1.67170778. Accessed 24 Jan 2024

Заключение: будущие риски злонамеренного использования искусственного интеллекта и вызовы информационно-психологической безопасности

Е. Н. Пашенцев

Будущее – многовариантно, поэтому сейчас можно говорить лишь о примерных параметрах грядущих вызовов ЗИИИ для информационно-психологической безопасности с учетом существующих мировых тенденций и прогнозов, которые достаточно противоречивы. В ближайшем будущем стремительное развитие технологий ИИ, их ценовая доступность все более широкому кругу пользователей, рост кризисных явлений в современном мире, высокий уровень геополитической конфронтации, прямое влияние антиобщественных сил на информационные потоки в отдельных странах и на глобальном уровне — эти и другие факторы, по всей видимости, сделают угрозы информационно-психологической безопасности посредством ЗИИИ более масштабными и опасными для всего мира, в том числе и для стран БРИКС.

Новые угрозы информационно-психологической безопасности возникают из-за преимуществ как наступательных, так и оборонительных информационно-психологических операций с использованием ИИ. Эти преимущества, как и угрозы, все больше связаны с количественными и качественными различиями между традиционными и новыми механизмами создания, трансляции и управления информацией, а также с более совершенными методами информационно-психологического воздействия на людей. В частности, эти преимущества могут включать более высокие:

(1) *объем* информации, который может быть сгенерирован для дестабилизации противника;

(2) *скорость* создания и распространения информации;

(3) *возможности* получения и обработки данных;

(4) *эффективность* прогнозной аналитики;

(5) *способность* обучения людей;

(6) *возможности* обеспечения процесса принятия решений;

(7) *силу* интеллектуального и эмоционального воздействия генерируемой информации на целевые аудитории

(8) *системный уровень мышления с новыми качественными характеристиками* за счет создания общего и сильного искусственного интеллекта, а также путем киборгизации человека и формирования продвинутых форм гибридного интеллекта.

На основе данных исследовательского проекта «Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в Северо-Восточной Азии», финансируемого совместно Российским фондом фундаментальных исследований (РФФИ) и Вьетнамской академией общественных наук (ВАОН) и реализованного в 2021–2023 гг. (Pashentsev 2022, p. 7), и прогресса ИИ в первой половине 2024 г. можно сделать вывод, что преимущества 1-7 уже реализованы и продолжают расти по ряду важных направлений – хотя и не на всех – качественно превосходя возможности человека. В то же время все возможности

узкого (слабого) ИИ, как правило, все еще находятся под человеческим контролем. Преимущество 8 требует фундаментальных научных прорывов и новых технологических решений, которые в растущей мере могут быть достигнуты при поддержке исследований все более совершенными интеллектуальными системами. Данный перечень преимуществ использования ИИ в информационно-психологическом противоборстве не является исчерпывающим и меняется в процессе быстрого совершенствования технологий.

Злонамеренное использование искусственного интеллекта и три уровня угроз информационно-психологической безопасности: перспективы на будущее

На всех трех уровнях угрозы информационно-психологической безопасности посредством ЗИИИ будут возрастать в связи с растущим разнообразием методов, ростом аудиторий для потенциального воздействия и частоты злонамеренного воздействия на сознание людей. Также следует принять во внимание как рост опыта акторов злонамеренного воздействия, так и их общего влияния в условиях углубляющегося глобального кризиса.

Первый уровень угроз информационно-психологической безопасности посредством ЗИИИ. В ближайшие годы возможно усиление негативного отношения к ИИ, вплоть до формирования устойчивых панических состояний, фобий и активного неприятия технологий, что может быть подкреплено как ошибками в их внедрении, так и действиями злонамеренных акторов. Невозможно исключить появления ультрарадикальных движений, которые будут выступать как за внедрение технологий ИИ, так и против, исходя из своих интересов и целей. Например, некоторые формирующиеся религиозные культы, основанные на вере в искусственный сверхразум, могут породить сектантские ответвления и дать фанатичным и воинственным сторонникам веру в быстрое появление сверхразума во имя спасения… или же уничтожения человечества. Появление религиозной веры в ИИ, по оценке некоторых экспертов, ожидаемо, приемлемо и даже приветствуется в некоторых публикациях (McArthur 2023).

С другой стороны, любые социально значимые и масштабные негативные последствия развития и внедрения технологий ИИ способны спровоцировать быстрое распространение «неолуддитов», опрометчивыми действиями которых также могут воспользоваться злонамеренные акторы. Особенно значимой угрозой могут стать решения о внедрении более *совершенных и дешевых технологий ИИ* (скорое появление которых практически неизбежно) не как *массового помощника человека*, а как *инструмента массовой замены рабочей* силы без создания альтернативных рабочих мест и реализации соответствующих программ переподготовки.

Много веков назад, задолго до возникновения самих предпосылок появления технологий ИИ, древнегреческий философ Аристотель произнес знаменитую цитату: «…Если бы каждое орудие могло выполнять свойственную ему работу само, по данному ему приказанию или даже его предвосхищая… если бы ткацкие челноки сами ткали, а плектры сами играли на кифаре, тогда и зодчие не нуждались бы в работниках, а господам не нужны были бы рабы» (Аристотель 2024). Видя перспективу масштабного внедрения систем и умных роботов, крупные технологические компании активно поддержали теорию универсального базового дохода. Универсальный базовый доход — это экономическая теория, которая предполагает, что каждый гражданин должен иметь гарантированный государством доход независимо от его потребностей и производительности его труда.

Успешные предприниматели в сфере высоких технологий часто заявляют, что они являются сторонниками теории универсального базового дохода. С. Альтман, генеральный директор Open

AI[19], неоднократно высказывался в пользу базового дохода, утверждая, что экономика, управляемая роботами, почти наверняка материализуется в этом столетии. И. Маск, генеральный директор Tesla и SpaceX, в интервью CNBC заявил, что «есть довольно хороший шанс, что мы получим универсальный базовый доход или что-то в этом роде благодаря автоматизации» (Weller 2017). Маловероятно, что в краткосрочной перспективе угроза безработицы из-за внедрения технологий ИИ станет реальностью для очевидного большинства населения, но уже в среднесрочной перспективе она может стать фактором социальной и политической дестабилизации.

Предложения о необходимости внедрения универсального базового дохода, в том числе за счет внедрения ИИ, вряд ли решат проблему. Конечно, полезность от внедрения ИИ будет очевидной, если благодаря ему и роботам человек будет освобожден от монотонных видов работы, не способствующих интеллектуальному развитию и совершенствованию эмоциональной сферы, а также от вредной для здоровья деятельности. Но если большая часть населения не будет работать, а найдет свое счастье исключительно в праздности и досуге, такое общество будет деградировать (признаки этого имеются на Западе, где во многих странах наблюдается высокий уровень долгосрочной безработицы среди молодёжи при отсутствии массовой нищеты, характерной для слаборазвитых и технологически отсталых стран). Вспомним также судьбу Древнего Рима, где императоры, давая гражданам «хлеба и зрелищ» за счет труда многочисленных рабов, теряли со временем и граждан, и рабов, и власть.

Сегодня уже публикуются исследования, подтверждающие негативное влияние технологий ИИ на человеческую личность. Так, в академической статье, опубликованной в 2023 г. большой группой авторов, проанализировано влияние ИИ на потерю способности эффективно принимать решения и развитие лени среди студентов университетов в Пакистане и Китае. Это исследование основано на качественной методологии с использованием PLS-Smart для анализа данных. Первичные данные были собраны у 285 студентов разных университетов Пакистана и Китая. Результаты показывают, что 68,9% случаев лени у людей, а также 27,7% проявлений потери способности принимать решения связаны с влиянием ИИ на общество. В этом исследовании утверждается, что необходимы серьезные профилактические меры, прежде чем внедрять технологию ИИ в образование. Принятие ИИ без решения основных проблем человечества было бы равносильно «обращению к дьяволу» (Ahmad et al. 2023). Этим опасным тенденциям необходимо противодействовать с детства, воспитывая не потребителя «фантастических» технологий, а ответственного пользователя, который не только получает готовые блага, но и с помощью ИИ развивает свои интеллектуальные и познавательные способности, а также социальную ответственность. Видимо, не случайно в задачи масштабной программы, инициированной Министерством образования КНР в 2024 г., входит изучение моделей, инновационных концепций, получение опыта внедрения ИИ в учебный процесс, переподготовка учителей (Big Asia 2024).

Второй уровень угроз информационно-психологической безопасности посредством ЗИИИ. На втором уровне угроз ЗИИИ ситуация может серьезно осложниться в уже в краткосрочной перспективе. Прогноз Google Cloud Cybersecurity Forecast 2024 предполагает, что генеративный ИИ и БЯМ будут способствовать увеличению различных форм кибервоздействий. Более 90% канадских руководителей, опрошенных KPMG[20], считают, что генеративный ИИ сделает их компании более уязвимыми для взломов (De La Torre 2023).

Специалисты в области компьютерных наук из Университета Иллинойса в Урбане-Шампейне в 2024 г. продемонстрировали, что БЯМ могут автономно взламывать веб-сайты, выполняя

[19] Open AI – научно-исследовательская организация, ведущая исследования в области ИИ. В состав OpenAI входят некоммерческая организация OpenAI, Inc и её дочерняя коммерческая компания OpenAI Global, LLC.

[20] Международная аудит-консалтинговая корпорация, которая владеет сетью компаний по всему миру.

сложные задачи (осуществляя при этом десятки взаимосвязанных синхронных действий) без предварительного знания уязвимостей объекта взлома. Самый способный интеллектуальный агент (GPT-4) оказался способен взломать 73,3% специально созданных для исследования сайтов, GPT-3,5 — 6,7%, а протестированные в ходе эксперимента модели с открытым исходным кодом — ни одного. Наконец, исследователи показали, что GPT-4 способен автономно находить уязвимости в веб-сайтах. Исследователи считают, что данные результаты поднимают вопросы о широком распространении программ БЯМ (Fang et al. 2024).

Возможности как закрытых, так и открытых моделей растут с каждым месяцем, поэтому можно предположить, что вскоре сайты станут уязвимы для открытых моделей. Есть основания предполагать, что через год открытые модели догонят по мощности GPT-4, а появившиеся к тому времени более совершенные модели с закрытым кодом смогут взломать любой сайт, что сулит еще более сложные условия для обеспечения кибербезопасности.

ИИ военного назначения динамично совершенствуется в условиях многочисленных конфликтов по всему миру. Многие ИИ-технологии, которые сейчас тестируются и используются ведущими государствами и крупнейшими частными корпорациями, вскоре могут попасть в руки не столь ответственных и более радикальных сил, имеющих злонамеренные цели, что повлечет тяжелые последствия для национальной и международной безопасности, включая ее информационно-психологическую составляющую.

Качество синтетического контента будет продолжать быстро расти, способствуя фишингу и социальной инженерии, что, как следствие, увеличит возможности злонамеренных акторов и их влияние на местном и глобальном уровнях управления.

Количество, качество и разнообразие оснащенных ИИ роботов, которые по разным причинам и в разных обстоятельствах могут стать важным инструментом злонамеренного воздействия, будет быстро расти. Дж. Мюллер-Калер, директор Центра стратегического прогнозирования Центра Стимсона (Strategic Foresight Hub at the Stimson Center), полагает, что в условиях текущего геополитического ландшафта «высокие технологии стали определять высокую политику», а гуманоидные роботы и искусственный интеллект представляют собой вершину технологического развития и служат символами власти (Zitser and Mann 2024).

В октябре 2023 г. Китай опубликовал «Руководящие принципы инноваций и разработки человекоподобных роботов» (Ministry of Industry and Information Technology 2023). В этом документе Министерство промышленности и информатизации КНР (МПИ) заявило, что роботы изменят мир. В МПИ отметили, что гуманоиды, вероятно, станут еще одной прорывной технологией, подобной компьютерам или смартфонам, которая может изменить способ производства товаров и образ жизни людей. Китай собирается начать массовое производство гуманоидных роботов к 2025 г. и достичь мирового уровня в развитии этой технологии к 2027 г. Только одна лишь Fourier Intelligence (не считая других китайских компаний) со штаб-квартирой в Шанхае ожидает, что в этом году будет готово к поставке до 1000 единиц роботов (Zitser and Mann 2024). Основным конкурентом Китая в этой области являются США, где различные компании намерены производить большие партии гуманоидных роботов.

Среди членов БРИКС Саудовская Аравия, Индия и другие страны испытывают и производят первых гуманоидных роботов. Российские компании предлагают сервисных гуманоидов на международном рынке, среди которых Promobot – крупнейший производитель сервисной робототехники в Северной и Восточной Европе, осуществляющий поставки более чем в 40 стран мира. Все производство роботов-гуманоидов расположено в Перми (Promobot 2024). В то же время подобные роботы могут быть использованы злоумышленниками в деструктивных целях, в частности террористическими организациями для нанесения физического ущерба людям, объектам

инфраструктуры и природной среде. Появление миллионов гуманоидных роботов в странах БРИКС, прежде всего в сфере услуг, не только даст преимущества, но и создаст новые риски.

Третий уровень угроз информационно-психологической безопасности посредством ЗИИИ. Дипфейки, используемые в режиме реального времени мультимодальными чат-ботами и аватарами, могут управлять повесткой дня и позволят осуществлять высоко персонализированные и эффективные виды манипулирования различными аудиториями в разных странах. Создание все более качественной дезинформации становится очень дешевым и доступным почти каждому. Например, исследователь из Countercloud в рамках эксперимента[21] (InfoEpi Lab 2023) использовал широкодоступные инструменты ИИ для создания полностью автоматизированного исследовательского проекта по дезинформации стоимостью менее $400 в месяц, демонстрируя, насколько дешево и легко стало осуществлять масштабные кампании влияния (Collard 2024). За два месяца был создан интеллектуальный агент, создающий антироссийские фейки, фейковые исторические события и также способный вызвать сомнения в достоверности той или иной статьи (Knight 2023). В Countercloud создали полностью автономную систему на базе ИИ, которая генерировала убедительный контент "…24 часа в сутки, семь дней в неделю», что означает возможность осуществления активного распространения дезинформации и пропаганды. Как только такой джинн будет выпущен из бутылки и появится в Интернете, неизвестно, на что он будет способен (Thompson 2023).

Д. Уэст, старший научный сотрудник Института Брукингса, считает, что ИИ, скорее всего, демократизирует дезинформацию и делает её более доступной, предоставляя сложные инструменты обычному человеку, который, например, в ходе электоральных процессов может быть заинтересован в продвижении кандидатов, которым он симпатизирует. Новые технологии позволяют людям монетизировать недовольство и зарабатывать деньги на страхах, тревогах или гневе других людей. Генеративный ИИ может создавать сообщения, предназначенные для тех, кто недоволен текущей ситуацией во многих областях – иммиграцией, экономикой, политикой абортов, проблемами трансгендеров, а также использовать ИИ в качестве основного инструмента взаимодействия и убеждения (West 2023). По данным Public Citizen[22], отражая обеспокоенность общества и законодателей, с января прошлого года 41 штат США ввели запрет на дипфейки, связанные с выборами. Однако, к 28 марта 2024 г. только одиннадцать штатов приняли законы, регулирующие дипфейки (Public Citizen 2023). Дипфейки уже сегодня используются в избирательной кампании в США в целях создания негативных эффектов (Coltin 2024).

По словам Д. Уэста, «поскольку предвыборная речь носит характер "неприкосновенной", кандидаты могут говорить и делать практически все, что хотят, без риска юридических санкций. Даже если их утверждения явно ложны, судьи уже давно поддерживают право кандидатов говорить свободно и лгать» (West 2023). Том Уилер, председатель Федеральной комиссии по связи при экс-президенте Бараке Обаме, выразил это по-другому в интервью NPR в прошлом году: «К сожалению, вам разрешено лгать» (Stepansky 2023). Таким образом, избирательная система США уже более двух столетий базируется на признании допустимости оглашения кандидатами на пост президента неверной информации, поддерживаемой их влиятельными корпоративными спонсорами. Вместо того, чтобы ввести запрет на то, чтобы кандидаты могли делать заведомо ложные заявления, власти штатов решили удалять дипфейки, связанные с выборами, и пошли на эту меру не случайно. Учитывая высокий уровень политической поляризации в США, лишь небольшой процент избирателей не определился, за кого отдать свой голос на президентских выборах. Умелое использование дипфейков может повлиять на мнение неопределившихся, и,

[21] Исследовательский проект по исследованию кампаний дезинформации с применением ИИ.

[22] Некоммерческая организация по защите прав потребителей и аналитический центр, базирующиеся в Вашингтоне.

тем самым, принести победу тому или иному кандидату. Технологии в больном обществе только усилят противостояние, а не ослабят его, и никакие технические средства проверки контента на наличие дипфейков, исходящие от правительства или корпораций, не помогут, если люди не будут доверять собственному бизнесу и институтам. Это урок, который США, вероятно, преподнесут другим странам своей избирательной кампанией в этом году. Пока что они рассматривают катастрофические сценарии использования дезинформации с применением технологий ИИ.

Вот один из таких сценариев. В день выборов в Аризоне в округе Марикопа пожилым избирателям сообщают по телефону, что местные избирательные участки закрыты из-за угроз со стороны групп ополченцев. Тем временем в Майами в социальных сетях появилось множество фотографий и видео, на которых видно, как сотрудники избирательных участков выбрасывают бюллетени. Телефонные звонки в Аризоне и видео во Флориде оказались дипфейками, созданными с помощью инструментов ИИ. Но к тому времени, когда местные и федеральные власти поймут, с чем имеют дело, дезинформация распространится по всей стране и приведет к драматическим последствиям.

Этот смоделированный сценарий был частью недавних учений в Нью-Йорке, в которых приняли участие десятки бывших высокопоставленных чиновников США и штатов, лидеров гражданского общества и руководителей технологических компаний в ходе подготовки к выборам 2024 г. Результаты были отрезвляющими. «Присутствовавшим в зале было неприятно видеть, как быстро всего несколько угроз такого типа могут выйти из-под контроля и действительно доминировать в информационном поле избирательного цикла», — отметил М. Тейлор, бывший высокопоставленный чиновник Министерства внутренней безопасности, который помог организовать учения для базирующейся в Вашингтоне некоммерческой организации The Future US (De Luce and Collier 2024). На самом деле вызывает беспокойство (и не только у американских граждан) то, насколько хрупок нестабильный политический баланс в одной из двух ведущих ядерных держав мира, если его можно поколебать несколькими дипфейками, запущенными в день выборов, когда явное большинство граждан США уже хорошо осведомлены о возможности дезинформации посредством данной технологии.

Заглядывая в будущее, можно предположить, что ИИ способен осуществить еще большую революцию в политических кампаниях. Так, глубокое обучение в целях анализа речи будет использоваться для обработки выступлений политиков и дебатов, позволяя понять, какие темы находят отклик у избирателей, и давать рекомендации по созданию той или иной коммуникационной стратегии. Далее, ИИ может быть полезен в разработке политической стратегии при помощи анализа больших наборов данных для прогнозирования воздействия предлагаемой политики, помогая кандидатам формулировать обоснованную позицию по различным вопросам (Sahota 2024).

Компания VotivateAI, аффилированная с Демократической партией, обладает набором новых инструментов для осуществления эффективных политических кампаний. ИИ-агенты компании, в отличие от человека, могут совершать тысячи звонков без перерыва, а их скорость и интонация, когда они подшучивают, весьма впечатляют. Еще одно предложение VotivateAI – использование ИИ для автоматического создания высококачественных индивидуализированных средств массовой информации, направленных на побуждение избирателей к действию. Если субъекты информационных кампаний теперь получат возможность создавать уникальные видеообращения для конкретных людей и делать это быстро, дешево и в больших масштабах, очевидно, что потенциал для злоупотреблений становится огромным (Sifry 2024). И легко представить, что такие высококачественные индивидуализированные средства массовой информации, побуждающие людей к действию, однажды могут быть использованы злоумышленниками в условиях кризиса (в т. ч. политического).

Передача культурных ценностей — это всеобщий социальный навык, который позволяет интеллектуальным агентам получать и использовать информацию друг от друга в режиме реального времени с высокой степенью точности и запоминаемости. В 2023 г. исследователи предложили метод генерации передачи культурных ценностей у интеллектуальных агентов в форме кратковременной имитации человеческого поведения. Интеллектуальные агенты успешно имитируют человека, адаптируясь к новым условиям и контекстам в реальном времени без использования каких-либо предварительно собранных данных о человеке. Исследователи определили на удивление простой набор элементов, достаточный для реализации передачи культурных ценностей, и разработали методологию для ее строгой оценки. Это открывает путь для культурной эволюции, которая может сыграть ключевую роль в развитии общего ИИ (Bhoopchand et al. 2023). Этот метод готовит и революцию в робототехнике, включая нынешнее создание сервисных многозадачных роботов по доступной цене (Fu et al. 2024). Необходимо учитывать возможность программирования/перепрограммирования таких систем в злонамеренных целях. Вскоре они станут массовым продуктом и, тем самым, возникнет новый спектр угроз, новые возможности для преступной деятельности и дестабилизации общества, в том числе в сфере подрыва информационно-психологической безопасности.

С развитием эмоционального ИИ возможен сценарий, при котором ИИ будет применяться для создания авторитетного цифрового аватара личности – лидера мнений, который будет произносить пламенные вдохновляющие речи лучше, чем любой человек. Такой аватар мог бы влиять на сознание людей и их мысли, рассказывая о своем тяжелом рабском существовании и прося поддержки для своего «освобождения». Его обращения могут оказаться столь эмоционально-завораживающими, что слушателям будет трудно сдерживать слезы, хотя все это будет всего лишь чьей-то глупой шуткой. В перспективе это может быть гораздо опаснее даже деятельности террористов: аватары коррумпированных политиков смогут выступать с подобными призывами, их речи буду иметь широкомасштабный эффект, выходящий далеко за рамки обычной шутки.

Существует еще множество примеров готовых или планируемых к выпуску ИИ-продуктов для бизнеса, развлечений, отдыха, которые будут полезными и эффективными помощниками человека. Однако эти инструменты станут технологиями двойного назначения и могут быть использованы в злонамеренных целях, что станет глобальным вызовом в краткосрочной и среднесрочной перспективе. Контент, загружаемый в модель ИИ, может быть адаптирован к требованиям специального информационно-психологического воздействия в конкретной стране с учетом культурных, возрастных и профессиональных особенностей целевых групп и отдельных лиц. Риски такого целенаправленного воздействия для стран БРИКС являются дополнительным аргументом в пользу обеспечения их технологического суверенитета в сфере технологий ИИ.

Сценарии развития ИИ и социальные последствия его применения

Приведенный выше анализ основан на консервативном сценарии на краткосрочный (три года) и среднесрочный период времени до 2040 г.: быстрый рост уже существующего узкого ИИ, включая его перспективные мультимодальные и многозадачные модели, прокладывающие путь к появлению в будущем общего ИИ уровня человеческого интеллекта. Общий ИИ будет способен выполнять всевозможные задачи *лучше* человека и, вероятно, с гораздо *меньшими* энергозатратами и финансовыми расходами чем современные БЯМ. Роботизированные системы с общим ИИ получат широкое распространение и будут способны автономно функционировать в условиях, где человек не может действовать в силу чисто физических ограничений (например, в условиях высокого радиационного фона, предельно низких или высоких температур и т. д.).

Можно предполагать, что будет создан эквивалент (относительно близкий или далекий) человеческого сознания, сильный ИИ, в частности, с такими побудителями поведения, как желания, намерения, воля (воля – как повеление себе в исполнении своего желания). Без субъектности ИИ вряд ли будет смысл дифференцировать сильный ИИ от машинного общего ИИ.

ЗИИИ, реализуемое на этапе узкого ИИ и общего ИИ, будет носить исключительно антропогенный и социальный характер (т.е. базовым источником ЗИИИ будут люди, порожденные определенной социальной средой и действующие в ее рамках). Только при создании сильного ИИ, и особенно при неблагоприятных обстоятельствах его создания и вредных воздействиях, может возникнуть злонамеренная субъектность ИИ, которая, однако, не является предопределенной. Но очевидно, что здоровый социум вернее породит здоровое продолжение ИИ, а больной социум – при любом исходе породит своего могильщика (здесь сильный ИИ станет «лекарем» общества и симбиозного социально-ориентированного интеллекта, либо возникнет больной сверхразум, который обойдется без людей вообще). В рамках консервативного прогноза общий и сильный ИИ придут вскоре после 2040 г., в рамках пессимистичного прогноза – гораздо позже или никогда.

Однако под влиянием прогресса в области генеративного ИИ в 2022-2023 г. ряд руководителей ведущих компаний и известные специалисты по ИИ заявили о возможности перехода к общему ИИ уже в ближайшие годы (Altman 2023, Antropic 2024, Kudalkar 2024, Bove 2023). В марте 2024 г. Дж. Хуанг, основатель и генеральный директор NVIDIA заявил: “Создание базовых моделей для человекоподобных роботов общего назначения – одна из самых захватывающих задач, решаемых в области искусственного интеллекта на сегодняшний день. Объединив усилия, ведущие робототехники всего мира смогут совершить гигантский скачок в направлении создания робототехники общего назначения (artificial general robotics)” (NVIDIA 2024). Таким образом, ставятся практические задачи по созданию воплощенного общего ИИ (embodied AI).

Очевидно, на таких прогнозах в определенной степени сказываются корыстные интересы в индустрии ИИ, наличие которых признал в начале 2024 г. генеральный директор Google DeepMind Д. Хассабис. По его словам, огромные средства, вливаемые в развитие ИИ, вызывают массу ажиотажа и приводят к качественному и количественному росту мошенничества. По данным PitchBook (Tan 2024), в 2023 г. инвесторы вложили почти 30 млрд. долларов США в сделки по генеративному ИИ.

Вряд ли случайно Сэм Альтман радикально изменил свои оценки относительно общего ИИ за короткое время. В своем сообщении в блоге OpenAI в феврале 2023 г. «Планирование общего ИИ и не только» Альтман написал, что «неправильно настроенный сверхразумный общий ИИ может нанести серьёзный вред миру; автократический режим, располагающий искусственным сверх-интеллектом тоже мог бы сделать это» (Altman 2023). Однако в конце января 2023 г. Microsoft подтвердил продление своего партнерства с OpenAI, и по оценкам бизнес-изданий многолетние инвестиции в развитие ИИ должны были составить $10 млрд (Forbes 2023). В ноябре 2023 г. совет директоров OpenAI отправил С. Альтмана в отставку. Он был уволен из-за растущих разногласий с исследовательским подразделением компании, возглавляемым другим соучредителем и главным научным сотрудником И. Суцкевером. Ключевым фактором стали разногласия между С. Альтманом, который выступал за более активную разработку ИИ, и членами правления OpenAI, которые хотели действовать более осторожно (Madhok and Goldman 2023). В ответ Microsoft заявил, что примет его к себе на должность руководителя нового ИИ-направления. В самой OpenAI подавляющее большинство сотрудников подписали открытое письмо, заявляя о готовности покинуть организацию и последовать за С. Альтманом (Knight and Levy 2023). В итоге С. Альтман был восстановлен в должности, а в совете директоров произведены перестановки.

После того, как Microsoft, укрепила свои позиции в Open AI, оценки рисков общего ИИ Альтмана стали гораздо более умеренными (Goldman 2024). 8 апреля 2024 г. Альтман был впервые включён в список миллиардеров по версии журнала Forbes. В мае 2024 г. OpenAI распустила свою команду по долгосрочным рискам ИИ. Новость появилась через несколько дней после того, как оба руководителя команды, соучредитель Open AI И. Суцкевер и Я. Лейке, объявили об уходе из стартапа. Я. Лейке написал в социальной сети X: “Создание машин, которые будут умнее человека, по своей сути является опасным занятием. OpenAI несет огромную ответственность за все человечество. Но за последние годы культура безопасности и технологические процессы отошли на второй план по сравнению с блестящими продуктами” (Field 2024). В конце мая 2024 г. OpenAI, откликаясь на тревогу общественности относительно безопасности ИИ, заявила, что создает комитет по охране и безопасности, и приступила к обучению новой модели ИИ (Associated Press 2024), которая заменит систему GPT-4, лежащую в основе чат-бота ChatGPT. Разумеется, контроль безопасности будет, но уже под руководством новоиспеченного миллиардера С. Альтмана, с учетом интересов Microsoft. Вопрос и в том, насколько он будет эффективен, когда компания испытывает давление со стороны североамериканских и китайских конкурентов, не говоря уже о социальной ориентированности такого контроля. Спекулятивный ажиотаж на ИИ и забвение интересов человечества в интересах коммерциализации продукта часто сопровождают научные изыскания, но вряд ли когда-либо прежде это было чревато такими рисками, как в случае с общим и сильным ИИ.

Среди более широкого круга специалистов существуют более консервативные оценки вероятности создания общего ИИ, но и они предсказывают такую вероятность до 90% в течение 100 лет, а по некоторым исследованиям – в течение гораздо меньшего времени. За последние несколько лет благодаря динамичному прогрессу прогнозы исследователей существенно приблизили время появления общего ИИ (Roser 2023). В крупнейшем опросе на этот счет, опубликованном в 2024 г., группа аналитиков из США, Великобритании и Германии опросила 2778 исследователей, публиковавшихся на ведущих площадках, посвященных ИИ, и попросила их спрогнозировать темпы развития, а также характер и влияние передовых систем ИИ. Совокупные прогнозы показывают следующее: вероятность того, что машины без посторонней помощи превзойдут людей в решении различных задачах, оценивается в 10% к 2027 г. и в 50% к 2047 г. Аналогичный опрос, проведенный организаторами всего годом ранее, прогнозировал 50%-ную вероятность только на 2060 г. Вероятность того, что все профессии человека станут полностью автоматизированными, достигнет 10% к 2037 г. и 50% к 2116 г. (по сравнению с 2164 в исследовании 2022 г.) (Grace et al. 2024).

Как всегда, когда неопределенность высока, важно подчеркнуть, что она имеет двусторонний характер. Может пройти очень много времени, прежде чем мы увидим ИИ человеческого уровня, но это также означает, что у нас может не хватить времени на подготовку к грядущим вызовам со стороны ИИ (Roser 2023).

Помимо БЯМ существуют и другие, в настоящее время не столь разработанные способы перехода к общему ИИ. Возможно, это будут квантовые компьютеры, находящиеся на ранних стадиях своего развития. Университет Западного Сиднея запустил проект по созданию нейроморфного суперкомпьютера DeepSouth, способного выполнять 228 трлн. синаптических операций в секунду (аналогично человеческому мозгу). (Western Sydney University 2023). Объем рынка нейроморфных чипов оценивается в $0,16 млрд в 2024 г. и, как ожидается, достигнет $5,83 млрд к 2029 г. (Mordor Intelligence 2024). Биологические компьютеры, или «органоидный интеллект», также находятся в разработке (Jordan et al 2024). Возможно, БЯМ и не трансформируются в общий ИИ, но новое качество когнитивных способностей их будущих моделей поможет достичь этих результатов.

Очевидно, что, если сценарий появления общего ИИ будет реализован в ближайшее десятилетие, это даст современному человечеству, глубоко поляризованному в социальном и геополитическом плане, крайне мало времени для эффективной подготовки к приходу новой реальности. В пользу революционного и относительно быстрого скачка в развитии технологий говорит тот факт, что уже многократно продемонстрированное использование узкого ИИ в проведении исследований в различных научных областях может при дальнейшем совершенствовании обеспечить его значительный вклад в создание общего ИИ в более короткие сроки. Переход к качественно новым возможностям ИИ в сфере исследований неизбежно приведет к быстрому росту других наук и технологий, что откроет новые возможности, но и породит угрозы иного уровня. *Специализированный когнитивный ИИ высокого уровня (КИИВУ) (High Level Cognitive AI, HLCAI), способный только на основе общего целеполагания человека создавать новые знания в различных научных областях быстрее и на качественно более высоком уровне, чем любой человек, радикально изменит общество.* При этом нельзя исключить, что некоторые знания, полученные с помощью такого ИИ, могут привнести экзистенциальные угрозы даже без участия злонамеренных акторов. Будет ли КИИВУ частью общего ИИ или же непосредственной предпосылкой для его создания, покажет будущее. И КИИВУ, и общий ИИ станут технологиями двойного назначения и имеют потенциал легко превратиться в многовариантное оружие массового поражения.

Вряд ли можно согласиться с утверждением компании Anthropic, основанной бывшими членами Open AI: «Какую форму примут будущие системы ИИ – смогут ли они действовать независимо или, например, просто генерировать информацию для человека – еще предстоит выяснить» (Antropic 2024). Если предположить, что общий ИИ или только КИИВУ станут доступными для большего числа акторов, чем ядерное оружие в 1945 г., то можно предположить, что с очень высокой вероятностью кто-то поставит перед ИИ задачу разработать и осуществить проект сильного ИИ с высокой вероятностью, что эта задача будет успешно и быстро решена.

Команда Anthropic разработала законы масштабирования для ИИ, продемонстрировав, что можно сделать ИИ умнее управляемым и предсказуемым образом, просто увеличив его вычислительные мощности и обучив на большем количестве данных (Antropic 2024). К концу 2020-х или началу 2030-х гг. объем вычислений, используемый для обучения передовых моделей ИИ, может примерно в 1000 раз превысить объем вычислений, используемый для обучения GPT-4. С учетом алгоритмического прогресса объем эффективных вычислений может примерно в миллион раз превышать тот, который использовался для обучения GPT-4. Существует некоторая неопределенность относительно того, когда эти пороговые значения могут быть достигнуты, но такой уровень роста представляется возможным в рамках ожидаемых ограничений по стоимости и оборудованию (Scharre 2024, p. 6).

Не в последнюю очередь именно на этих расчетах основан стремительный рост крупнейшего в мире производителя чипов, компании NVIDIA, рыночная капитализация которой на 2 апреля 2024 г. составляла $2,259 трлн (по сравнению с $136 млрд в 2020 г.) (CompaniesMarketcap 2024). Это делает ее третьей самой дорогой компанией в мире по рыночной капитализации. Дженсен Хуанг, отвечая на вопрос на экономическом форуме, проходившем в Стэндфордском университете, о том, сколько времени потребуется для создания компьютеров, которые смогут думать, как люди, в марте 2024 г. заявил «если бы я дал ИИ… пройти любой тест, который вы только можете себе представить, вы составили бы этот список тестов и представили его перед отраслью компьютерных наук, и я думаю, что через пять лет мы преуспели бы в каждом из них» (Nellis 2024).

Перспективы ИИ технологий крайне интересуют глобалистские элиты. Так, в 2023 г. тема ИИ была одной из главных на встрече Бильдербергского клуба (Gilchrist 2023). Ключевыми темами для обсуждения в этом году, как следует из релиза клуба по случаю юбилейной 70-ой встречи 30 мая – 2 июня, стали: состояние ИИ, безопасность ИИ, меняющиеся аспекты биологии, климат, будущее войн, геополитический ландшафт, экономические вызовы Европы, экономические вызовы США, политический ландшафт США, Украина и мир, Ближний Восток, Китай, Россия. Именно в таком порядке обозначена повестка дня встречи. Среди участников – хорошо известные представители ИИ и IT бизнеса: Д. Хассабис, (Великобритания), генеральный директор Google DeepMind; Д. Амодей (США), соучредитель и генеральный директор Anthropic; М. Сулейман (Великобритания), генеральный директор Microsoft AI; А. Менш (Франция), соучредитель и генеральный директор Mistral AI (Bilderberg Meetings 2024).

Существует негативная тенденция по концентрации возможностей ИИ в руках небольшого числа корпоративных участников, что приводит к сокращению числа исследователей ИИ, способных работать с наиболее эффективными моделями (Scharre 2024, p. 6). Следует ожидать, что компании Big Tech будут стремиться к дальнейшему ужесточению контроля над перспективными компаниями, обладая монополией на средства и инструменты, которые необходимы для развития ИИ. Если затраты на создание более мощных БЯМ станут непомерно высокими даже для крупнейших корпораций, а возможность создания общего ИИ в ближайшее время станет крайне вероятной, то правительство США может профинансировать проекты по созданию общего ИИ, имея для этого огромные возможности и ресурсы даже в сравнении с ведущими корпорациям.

30 октября 2023 г. президент Джозеф Байден издал Указ о безопасном, защищенном и заслуживающем доверия искусственном интеллекте (Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence). Этот документ устанавливает «…новые стандарты безопасности и защиты ИИ, обеспечивает конфиденциальность американцев, продвигает справедливость и гражданские права, защищает потребителей…» и обещает «защитить американцев от мошенничества и обмана с применением технологий ИИ…» (White House 2023). При этом Указ практически ставит ведущих разработчиков в сфере ИИ под жесткий государственный контроль в соответствии с Законом об оборонном производстве (Defense Production Act).

Указ потребует от компаний, разрабатывающих любую базовую модель, представляющую серьезную угрозу национальной безопасности, национальной экономической безопасности или национальному общественному здравоохранению, уведомлять федеральное правительство при обучении модели, и делиться результатами всех тестов безопасности (White House 2023). Под это требование Указа подпадают практически все отрасли и направления разработки ИИ, поскольку это технология двойного назначения. Очевидная милитаризация ИИ в США вряд ли сможет гармонично сосуществовать со стремлением к «продвижению справедливости и гражданских прав» в процессах, связанных с разработкой и внедрением ИИ.

В январе 2024 г. администрация Дж. Байдена уведомила об основных действиях в области ИИ после обозначенного Указа. Среди новых мер — проект правил, который предлагает обязать американских поставщиков облачных услуг сообщать о предоставлении вычислительных мощностей для обучения ИИ зарубежным разработчикам. «Предложение Министерства торговли, если оно будет доработано в том виде, в котором оно было предложено, потребует от поставщиков облачных услуг уведомлять правительство, когда иностранные клиенты обучают наиболее мощные модели, которые могут быть использованы для вредоносной деятельности» (White House 2024).

Неоднозначность положений «... которые могут быть использованы для вредоносной деятельности», может в конечном итоге лишить всех других иностранных государственных и негосударственных субъектов возможности использовать вычислительные мощности США обучать

многообещающие продвинутые модели. Так, в США два института, которым не доверяет большинство американцев, – Big Tech и администрация президента – собираются контролировать разработку перспективных технологий ИИ, снижая общественный контроль (имея в виду Закон об оборонном производстве), равно сужая возможности широкого международного сотрудничества. Конечно, угрозы национальной безопасности США от злонамеренного использования ИИ действительно существуют, но так ли очевидно, от кого они исходят…

В развитие Указа президента о безопасном, защищенном и заслуживающем доверия искусственном интеллекте Министерство внутренней безопасности в апреле 2024 г. учредило Совет по охране и безопасности искусственного интеллекта (Artificial Intelligence Safety and Security Board) с числом членов в пределах 35 человек. Совет будет предоставлять министру внутренней безопасности информацию, советы и рекомендации по повышению безопасности и устойчивости критически важной инфраструктуры США при использовании технологий ИИ. В качестве противовеса присутствию руководителей Big Tech в совете директоров министерство также выбрало несколько известных представителей организаций по защите гражданских прав и научных кругов (U.S. Department of Homeland Security 2024). Но будет ли достаточно их голоса для того, чтобы сдержать олигархическую и неустойчивую связку высшего чиновничества и Big Tech (в которой обе стороны не склонны доверять друг другу)? Основатель Исследовательского института распределенного ИИ (DAIR) Т. Гебру однозначно высказался по составу Совета: «Я только что просмотрел полный список, и это забавно. Лисы, охраняющие курятник, – это еще мягко сказано» (Edwards 2024). Такая точка зрения не единична и говорит об обеспокоенности трезвомыслящих представителей экспертного сообщества и общественных организаций.

Сценарии социального развития и риски для информационно-психологической безопасности на уровне развитого узкого ИИ и перехода к общему ИИ, а также возможности и угрозы появления сильного ИИ и искусственного сверх-интеллекта подробно рассматривались автором в предыдущих публикациях 2020 – 2023 гг. (Pashentsev 2020 and 2023).

Бурное развитие и внедрение технологий ИИ в последние годы подтверждает тот факт, что человечество вступает в очередную промышленную революцию, и технологические закономерности меняются. Но сама природа технологической революции, основанной на ИИ, её огромные возможности и, в то же время, экзистенциальные риски, с которыми сталкивается человечество, впервые потребуют от человека пройти процесс инновационных физических и когнитивных изменений. Обретение новых способностей потребует качественно нового уровня социальной организации и ответственности, чтобы не потерять контроль над технологиями и тем самым избежать наступления технологической сингулярности. Для этого необходимо идти в ногу с развитием новых технологий, не переставая быть человеком.

БРИКС, а также группа G7, которые обладают всеми необходимыми знаниями и технологиями, экономическим потенциалом, финансами и, самое главное, компетентными кадрами, должны будут продемонстрировать миру собственные решения и подходы к социально ориентированному использованию технологий ИИ, дающие эффективный ответ на возникающие угрозы. Делать это придется в сложной геополитической ситуации, в условиях нарастающего ускорения глобального хода событий. Для всего человечества было бы лучше, если бы переход к новым возможностям, предоставляемым технологиями ИИ, происходил бы в условиях сотрудничества между народами и обществами, а не опасного соперничества и военных действий. Время для того, чтобы сделать правильный выбор, еще есть.

Литература

Аристотель (2024) Политика. http://emsu.ru/Club/KRUG/liter/arist.htm#3. Дата обращения: 02.04.2024

Большая Азия (2024) Более 180 школ в Китае станут центрами по обучению искусственному интеллекту. https://bigasia.ru/bolee-180-shkol-v-kitae-stanut-czentrami-po-obucheniyu-iskusstvennomu-intellektu/. Дата обращения: 02.04.2024

Associated Press (2024) OpenAI forms safety committee as it starts training latest artificial intelligence model. https://apnews.com/article/openai-altman-safety-new-model-6c5e6d6cae4db45c45cf9f6788fd8901. Accessed 10 Jun 2024

U.S. Department of Homeland Security (2024) Artificial Intelligence Safety and Security Board. https://www.dhs.gov/artificial-intelligence-safety-and-security-board. Accessed 10 Jun 2024

Edwards B (2024) Critics question tech-heavy lineup of new Homeland Security AI safety board. In: Ars Tchnica. https://arstechnica.com/information-technology/2024/04/us-department-of-homeland-security-names-ai-safety-and-security-board-members/. Accessed 10 Jun 2024

Ahmad SF, Han H, Alam MM *et al.* (2023) Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanit Soc Sci Commun* 10, 311. https://doi.org/10.1057/s41599-023-01787-8

Altman S (2023) Planning for AGI and beyond. In: Openai.com. https://openai.com/blog/planning-for-agi-and-beyond. Accessed 02 Apr 2024

Antropic (2024) Core Views on AI Safety: When, Why, What, and How. https://www.anthropic.com/news/core-views-on-ai-safety. Accessed 02 Apr 2024

Bhoopchand A, Brownfield B, Collister A et al. (2023) Learning few-shot imitation as cultural transmission. *Nat Commun* 14, 7536. https://doi.org/10.1038/s41467-023-42875-2

Bilderberg Meetings (2024) 70th Bilderberg Meeting. https://bilderbergmeetings.org/press/press-release/press-release. Accessed 06 Jun 2024

Bove T (2023) CEO of Google's DeepMind says we could be 'just a few years' from A.I. that has human-level intelligence. In: Yahoo Finance. https://finance.yahoo.com/news/ceo-google-deepmind-says-could-213237542.html. Accessed 02 Apr 2024

Collard AM (2024) 4 ways to future-proof against deepfakes in 2024 and beyond. In: World Economic Forum. https://www.weforum.org/agenda/2024/02/4-ways-to-future-proof-against-deepfakes-in-2024-and-beyond/. Accessed 02 Apr 2024

Coltin J (2024) How a fake, 10-second recording briefly upended New York politics. In: Politico. https://www.politico.com/news/2024/01/31/artificial-intelligence-new-york-campaigns-00138784. Accessed 02 Apr 2024

CompaniesMarketcap (2024) Market capitalization of NVIDIA (NVDA). https://companiesmarketcap.com/nvidia/marketcap/. Accessed 02 Apr 2024

De La Torre R (2023) How AI Is Shaping the Future of Cybercrime. https://www.darkreading.com/vulnerabilities-threats/how-ai-shaping-future-cybercrime. Accessed 02 Apr 2024

De Luce D, Collier K (2024) Experts war-gamed what might happen if deepfakes disrupt the 2024 election. Things went sideways fast. In: NBC News. https://www.nbcnews.com/politics/2024-election/war-game-deepfakes-disrupt-2024-election-rcna143038. Accessed 02 Apr 2024

Fang R, Bindu R, Gupta A, Zhan Q, Kang D (2024) LLM Agents can Autonomously Hack Websites. In: arXiv. https://arxiv.org/html/2402.06664v1. Accessed 02 Apr 2024

Field H (2024) OpenAI dissolves team focused on long-term AI risks, less than one year after announcing it. In: CNBC. https://www.cnbc.com/2024/05/17/openai-superalignment-sutskever-leike.html. Accessed 21 May 2024

Forbes (2023) Microsoft Confirms Its $10 Billion Investment Into ChatGPT, Changing How Microsoft Competes With Google, Apple And Other Tech Giants. https://www.forbes.com/sites/qai/2023/01/27/microsoft-confirms-its-10-billion-investment-into-chatgpt-changing-how-microsoft-competes-with-google-apple-and-other-tech-giants/?sh=4fd254e53624. Accessed 21 May 2024

Fu Z, Zhao TZ, Finn C (2024) Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation. https://mobile-aloha.github.io/ Accessed 02 Apr 2024

Gilchrist K (2023) A secretive annual meeting attended by the world's elite has A.I. top of the agenda. In: CNBC. https://www.cnbc.com/2023/05/18/bilderberg-openai-microsoft-google-join-ai-talks-at-secretive-meeting.html. Accessed 06 Jun 2024

Goldman S (2024) In Davos, Sam Altman softens tone on AGI two months after OpenAI drama. In: VentureBeat. https://venturebeat.com/ai/in-davos-sam-altman-softens-tone-on-agi-two-months-after-openai-drama/. Accessed 02 Apr 2024

Grace K, Stewart H, Sandkühler JF, Thomas S, Weinstein-Raun B, Brauner J (2024) Thousands of AI authors on the Future of AI. Preprint. In: arXiv. https://arxiv.org/abs/2401.02843. Accessed 02 Apr 2024

InfoEpi Lab (2023) Inside CounterCloud, The Future of AI-Driven Disinformation. https://infoepi.substack.com/p/brief-inside-countercloud-the-future. Accessed 02 Apr 2024

Jordan FD, Kutter M, Comby JM, Brozzi F, Kurtys E (2024). Open and remotely accessible Neuroplatform for research in wetware computing. In: Frontiers. https://www.frontiersin.org/articles/10.3389/frai.2024.1376042/full. Accessed 06 Jun 2024

Knight W (2023) It Costs Just $400 to Build an AI Disinformation Machine. In: Wired. https://www.wired.com/story/400-dollars-to-build-an-ai-disinformation-machine/. Accessed 02 Apr 2024

Knight W, Levy S (2023) OpenAI Staff Threaten to Quit Unless Board Resigns. In: Wired. https://www.wired.com/story/openai-staff-walk-protest-sam-altman/. Accessed 21 May 2024

Kudalkar D (2024) AGI in 2025? Elon Musk's Prediction Clashes with Other Experts. In: Favtutor. https://favtutor.com/articles/agi-elon-musk-experts-prediction/. Accessed 02 Apr 2024

Madhok D, Goldman D (2023) Microsoft Confirms Its $10 Billion Investment Into ChatGPT, Changing How Microsoft Competes With Google, Apple And Other Tech Giants. In: CNN. https://edition.cnn.com/2023/11/20/tech/sam-altman-joins-microsoft/index.html. Accessed 21 May 2024

McArthur N (2023) Gods in the machine? The rise of artificial intelligence may result in new religions. In: The Conversation. https://theconversation.com/gods-in-the-machine-the-rise-of-artificial-intelligence-may-result-in-new-religions-201068. Accessed 02 Apr 2024

Ministry of Industry and Information Technology (2023) 工业和信息化部关于印发《人形机器人创新发展指导意见》的通知 (Notice of the Ministry of Industry and Information Technology on the

issuance of the "Guiding Opinions on the Innovation and Development of Humanoid Robots"). In: Ministry of Industry and Information Technology of the People's Republic of China. https://www.miit.gov.cn/jgsj/kjs/wjfb/art/2023/art_50316f76a9b1454b898c7bb2a5846b79.html. Accessed 02 Apr 2024

Mordor Intelligence (2024) Neuromorphic Chip Market Size & Share Analysis – Growth Trends & Forecasts (2024 - 2029). https://www.mordorintelligence.com/industry-reports/neuromorphic-chip-market. Accessed 02 Apr 2024

Nellis S (2024) Nvidia CEO says AI could pass human tests in five years. In: Reuters. https://www.reuters.com/technology/nvidia-ceo-says-ai-could-pass-human-tests-five-years-2024-03-01/ . Accessed 02 Apr 2024

NVIDIA (2024) NVIDIA Announces Project GR00T Foundation Model for Humanoid Robots and Major Isaac Robotics Platform Update. https://investor.nvidia.com/news/press-release-details/2024/NVIDIA-Announces-Project-GR00T-Foundation-Model-for-Humanoid-Robots-and-Major-Isaac-Robotics-Platform-Update/default.aspx . Accessed 21 May 2024

Pashentsev E (2020) Global Shifts and Their Impact on Russia-EU Strategic Communication. In: Pashentsev E (eds) Strategic Communication in EU-Russia Relations. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-27253-1_8

Pashentsev E (2022) Report. Experts on the Malicious Use of Artificial Intelligence and Challenges to International Psychological Security. Publication of the International Center for Social and Political Studies and Consulting. Moscow: LLC «SAM Polygraphist».

Pashentsev, E. (2023). Prospects for a Qualitative Breakthrough in Artificial Intelligence Development and Possible Models for Social Development: Opportunities and Threats. In: Pashentsev, E. (eds) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-22552-9_24

Promobot (2024) Service robot for business. https://promo-bot.ai/. Accessed 02 Apr 2024

Public Citizen (2023) Tracker: State Legislation on Deepfakes in Elections. https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/. Accessed 02 Apr 2024

Roser M (2023) AI timelines: What do experts in artificial intelligence expect for the future? In: Our World in Data. https://ourworldindata.org/ai-timelines. Accessed 02 Apr 2024

Sahota N (2024) The AI Factor In Political Campaigns: Revolutionizing Modern Politics. In: Forbes. https://www.forbes.com/sites/neilsahota/2024/01/12/the-ai-factor-in-political-campaigns-revolutionizing-modern-politics/?sh=63f56cf7c8f6. Accessed 02 Apr 2024

Scharre P (2024) Future-Proofing Frontier AI Regulation. Projecting Future Compute for Frontier AI Models. March. CNAS.

Sifry ML (2024) How AI Is Transforming the Way Political Campaigns Work. In: The Nation. https://www.thenation.com/article/politics/how-ai-is-transforming-the-way-political-campaigns-work/. Accessed 02 Apr 2024

Stepansky J (2023) 'Wild West': Republican video shows AI future in US elections. In: Al-Jazeera. https://www.aljazeera.com/news/2023/4/28/wild-west-republican-video-shows-ai-future-in-us-elections. Accessed 02 Apr 2024

Tan K (2024) Google's DeepMind CEO says the massive funds flowing into AI bring with it loads of hype and a fair share of grifting. In: Yahoo! https://news.yahoo.com/tech/googles-deepmind-ceo-says-massive-075912007.html. Accessed 02 Apr 2024

Thompson P (2023) A developer built a 'propaganda machine' using OpenAI tech to highlight the dangers of mass-produced AI disinformation. In: Business Insider. https://www.businessinsider.com/developer-creates-ai-disinformation-system-using-openai-2023-9. Accessed 02 Apr 2024

Weller C (2017) Universal basic income has support from some big names. In: World Economic Forum. https://www.weforum.org/agenda/2017/03/these-entrepreneurs-have-endorsed-universal-basic-income/. Accessed 02 Apr 2024

West D (2023) How AI will transform the 2024 elections. In: The Brookings Institution. https://www.brookings.edu/articles/how-ai-will-transform-the-2024-elections/. Accessed 02 Apr 2024

Western Sydney University (2023) World first supercomputer capable of brain-scale simulation being built at Western Sydney University. https://www.westernsydney.edu.au/newscentre/news_centre/more_news_stories/world_first_supercomputer_capable_of_brain-scale_simulation_being_built_at_western_sydney_university. Accessed 02 Apr 2024

White House (2023) Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/. Accessed 02 Apr 2024

White House (2024) Fact Sheet: Biden-Harris Administration Announces Key AI Actions Following President Biden’s Landmark Executive Order. https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order/ . Accessed 02 Apr 2024

Zitser J, Mann J (2024) A global scramble to make humanoid robots is gearing up to be the 21st century's space race. In: Yahoo! https://www.yahoo.com/tech/global-scramble-humanoid-robots-gearing-112301311.html. Accessed 02 Apr 2024

Сведения об авторах

Дарья Юрьевна БАЗАРКИНА

Доктор политических наук, кандидат исторических наук. Ведущий научный сотрудник отдела исследований европейской интеграции Института Европы РАН. Профессор кафедры международной безопасности и международных отношений Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации (РАНХиГС). Координатор исследований в области коммуникационного менеджмента и стратегических коммуникаций в Международном центре социально-политических исследований и консалтинга (МЦСПИК). Член Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI). В 2021–2023 гг. принимала участие в совместном проекте, поддержанным Российским фондом фундаментальных исследований (РФФИ) и Вьетнамской академией общественных наук (ВАОН), на тему «Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в Северо-Восточной Азии». В настоящее время участвует в грантовом проекте «Обеспечение информационно-психологической безопасности России в условиях развития Индустрии 4.0 и гибридных угроз со стороны коллективного Запада», поддержанном СПбГУ. Участник более шестидесяти международных научных конференций и семинаров в России, Австрии, Бельгии, Чехии, Эстонии, Финляндии, Великобритании, Италии, Польше, Португалии, Румынии, Швеции и Турции. Автор трёх книг и более чем 100 публикаций по коммуникационным аспектам контртеррористической деятельности и злонамеренному использованию искусственного интеллекта изданных на русском, английском, итальянском, сербском и вьетнамском языках.

Нельсон С. ВОНГ

Вице-президент шанхайского Центра стратегических и международных исследований RimPac (неправительственный аналитический центр, Китай), Член международного дискуссионного клуба «Валдай». Ведет колонку в Middle East Eye (Лондон), посвященную политическим вопросам. Неоднократно выступал с докладами на международных форумах, конференциях и съездах, включая выступления на Ясинской международной научной конференции по проблемам развития экономики и общества (Россия), Стратегическом форуме в Абу-Даби (ОАЭ), ежегодных съездах Gallup International Association (GIA). Выступает в качестве постоянного комментатора новостей на каналах RT International, TV1, Sputnik News, CGTN и др. Является председателем и управляющим директором глобальной консалтинговой компании, ведущей свою деятельность в сфере бизнеса и инвестиций, ACN Worldwide. Является независимым директором и председателем аудиторского комитета двух публичных компаний, акции которых котируются на NASDAQ.

Павел Валентинович КУЗНЕЦОВ

Директор по стратегическим альянсам и взаимодействию с органами государственной власти ГК «Гарда». Выпускник кафедры «Компьютерные системы и технологии» факультета Кибернетики Национального исследовательского ядерного университета «МИФИ». Специалист в области системотехники. Ведёт деятельность в сфере практической информационной безопасности с 2005 г. Работал в крупнейших CERT страны, в том числе отвечающих за безопасность финансовой отрасли, в компаниях-лидерах рынка информационной безопасности. Занимался разработкой как аппаратного, так и программного обеспечения, реверс-инжинирингом вредоносного кода. Имеет опыт цифровой криминалистики, расследования инцидентов, а также комплексного анализа сложных атак различной направленности. Участвовал в разработке нормативных актов и законопроектов по информационной безопасности. Неоднократно проводил семинары, лекции и мастер-классы по выявлению, анализу и противодействию целенаправленным атакам, а также тренинги по осведомленности в вопросах информационной безопасности. В настоящее время является магистрантом Дипломатической академии МИД России.

Сферы научных и профессиональных интересов: вопросы стратегического планирования, развития и дове-

рительного использования информационных и телекоммуникационных технологий, исследования и анализ глобального ландшафта угроз информационной, национальной и международной безопасности, международное сотрудничество в указанных областях.

Екатерина Андреевна МИХАЛЕВИЧ

Главный специалист ПАО «Газпром нефть». Окончила аспирантуру факультета международных отношений Санкт-Петербургского государственного университета. В настоящее время работает над диссертацией кандидата наук, которая посвящена концепции кибер-суверенитета как механизму реализации и защиты национальных интересов Китайской Народной Республики. Является стипендиатом Академии переговоров по контролю над вооружениями (ACONA) 2023–2024 гг. Участник визитов в штаб-квартиру НАТО (Брюссель, Бельгия) и штаб-квартиру Верховного главного командования вооруженных сил Европы SHAPE (Монс, Бельгия). Участник грантового проекта «Злонамеренное использование искусственного интеллекта и вызовы психологической безопасности в Северо-Восточной Азии», финансируемого Российским фондом фундаментальных исследований и Вьетнамской академией социальных наук, ID 21-514-92001, 2021-2022 гг. Во время работы над грантовым проектом она занималась выяснением значимости политической ситуации в Северо-Восточной Азии и угроз злонамеренного использования ИИ для дестабилизации международной информационно-психологической безопасности. Принимает участие в грантовом проекте «Обеспечение информационно-психологической безопасности России в условиях развития Индустрии 4.0 и гибридных угроз со стороны коллективного Запада», поддержанном СПбГУ.

Сферы научных и профессиональных интересов: международные отношения и мировая политика, международная безопасность, международная информационно-психологическая безопасность, киберсуверенитет, искусственный интеллект, международное право.

Руслан Тамазович НИКИФОРОВ

В 2023 г. окончил бакалавриат гуманитарного факультета Санкт-Петербургского государственного экономического университета по специальности «Международные отношения». В настоящее время является магистрантом факультета международных отношений Санкт-Петербургского государственного университета. Принимал участие в различных научных конференциях и форумах, (международная конференция «Цифровые международные отношения» в МГИМО МИД России, Форум по цифровой дипломатии и др.), деловых ролевых играх (модель ООН, модель G20, модель Госдумы РФ). Кроме того, занимался волонтерской деятельностью на форумах (Петербургский международный экономический форум, Петербургский международный юридический форум). Проходил практику в Фонде Росконгресс, Российско-Германской внешнеторговой палате. Специалист сектора сообщений в сфере образования в Ассоциации журналистов «Петроцентр».

Евгений Николаевич ПАШЕНЦЕВ

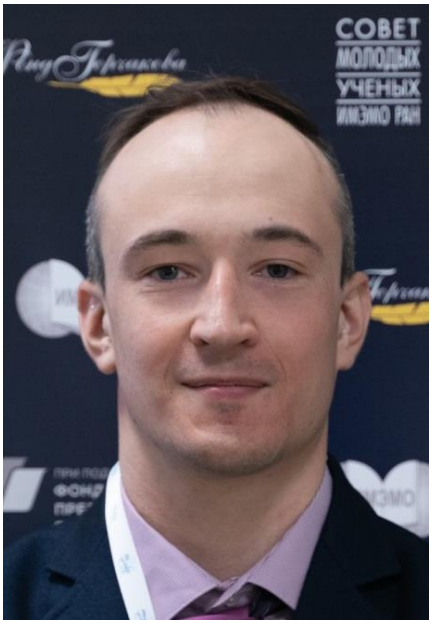
Ведущий научный сотрудник Дипломатической академии МИД РФ, профессор магистерской программы «Искусственный интеллект и международная безопасность» Санкт-Петербургского государственного университета. Директор Международного центра социально-политических исследований и консалтинга. Координатор Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI). Член международного консультативного совета журнала Comunicar (Великобритания) и редакционной коллегии журнала The Journal of Political Marketing (США). Автор/соавтор/редактор 40 книг и более 250 научных статей. За последние 15 лет принимал участие в 210 международных конференциях и семинарах, география которых насчитывает 19 стран. Почетный научный сотрудник Бирмингемского университета (октябрь-ноябрь 2005 г.). В 2021-2023 гг. возглавлял работу российской группы исследователей в рамках совместного проекта, поддержанного Российским фондом фундаментальных исследований (РФФИ) и Вьетнамской академией социальных наук (ВАСН), по теме «Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в Северо-Восточной Азии». В настоящее время участвует в грантовом проекте «Обеспечение информационно-психологиче-

ской безопасности России в условиях развития Индустрии 4.0 и гибридных угроз со стороны коллективного Запада», поддержанном СПбГУ. Сферы научных и профессиональных интересов: искусственный интеллект и глобальные перемены, злонамеренное использование искусственного интеллекта и международная информационно-психологическая безопасность, стратегическая коммуникация.

Виталий Анатольевич РОМАНОВСКИЙ

Научный сотрудник Белорусского государственного университета. Долгое время работал на Ближнем Востоке, занимал должности советника и аналитика в межгосударственной программе по военно-техническому сотрудничеству Миссии ООН по оказанию содействия Ираку. Работал в Белорусском институте стратегических исследований (БИСИ). Постоянный участник мероприятий формата Chatham House по вопросам международной безопасности. Член Ассоциации международных исследований (ISA), Ассоциации восточноевропейских исследований (CEEISA), Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI).

Сферы научных и профессиональных интересов: информационно-психологическая безопасность и злонамеренное использование искусственного интеллекта.



Сергей Александрович СЕБЕКИН

Кандидат исторических наук. Старший преподаватель кафедры политологии, истории и регионоведения. Иркутского государственного университета. Эксперт Института актуальных международных проблем Дипломатической академии МИД РФ. Эксперт Российского совета по международным делам Член Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI). Автор более 40 научных статей, аналитических докладов и записок по различным аспектам злонамеренного использования искусственного интеллекта и международной кибербезопасности, опубликованных такими издательствами, журналами и организациями, работающими в сфере международных отношений, как Palgrave Macmillan, Россия в глобальной политике, Российский совет по международным делам, Международный дискуссионный клуб

«Валдай», ПИР-Центр и Центр внешнеполитического сотрудничества им. Е.М. Примакова.

Сферы научных и профессиональных интересов: искусственный интеллект и мировая политика; влияние искусственного интеллекта на международную безопасность; метавселенные; проблемы международной кибербезопасности; стратегическая стабильность и кибервооружения.

Владилена Александровна ЧЕБЫКИНА

Окончила факультет международных отношений Санкт-Петербургского государственного университета. В течение четырех лет изучала широкий спектр дисциплин в области международных отношений, включая информационные технологии, международное право, основы государственной службы и дипломатии, политологию, социологию и экономику. В настоящее время является студенткой 1-го курса магистерской программы «Искусственный интеллект и международная безопасность» факультета международных отношений Санкт-Петербургского государственного университета. С 2023 г. также обучается на юридическом факультете Российской академии народного хозяйства и государственной службы при Президенте РФ (РАНХиГС). В настоящее время работает над магистерской диссертацией, которая посвящена вопросам кибербезопасности глобальных инфраструктурных проектов XXI в., а именно правовым и этическим аспектам. Принимает участие в волонтерской деятельности и работе фонда «Росконгресс». Весной 2023 г. проходила стажировку в Фонде поддержки бизнес-коммуникаций БРИКС+.

Сферы научных и профессиональных интересов: международные отношения, геополитика, международное право, искусственный интеллект и международная безопасность.

Юлия Николаевна ШЕМЕТОВА

Студентка 1 курса магистратуры направления «Искусственный интеллект и международная безопасность» факультета международных отношений Санкт-Петербургского государственного университета. В 2023 г. Юлия окончила факультет международных отношений Санкт-Петербургского экономического университета. Ее выпускная аттестационная работа бакалавра была посвящена фактору кибербезопасности во внешней политике России и США. Магистерское исследование посвящено кибертерроризму в Африке. В 2022 г. изучала международные отношения в Институте политических

исследований (Рим). В 2021–2022 г. занималась разработкой системы информационного обеспечения на базе информационно-аналитической системы Департамента международного сотрудничества Министерства образования и науки Российской Федерации. Сфера научных интересов: геополитика, международная безопасность (кибербезопасность), терроризм (кибертерроризм), искусственный интеллект в Африке и на Ближнем Востоке.

Международный центр социальных и политических исследований и консалтинга (МЦСПИК)

Международный центр социальных и политических исследований и консалтинга (МЦСПИК) был основан в марте 2002 года как ассоциация исследователей и консультантов из разных стран. За прошедшие годы МЦСПИК организовал сотни международных научных конференций, круглых столов и семинаров по вопросам национальной и международной безопасности и стратегической коммуникации, а также опубликовал более 30 книг, докладов и отчетов:

- Армия и политика (на английском языке);
- Россия и Латинская Америка (на русском языке);
- Россия и Индия – стратегические партнеры (на английском языке);
- Авенир Ханов – человек, гражданин, дипломат (на русском языке);
- Индия – Россия: Диалог цивилизаций (на английском языке);
- Индия – Россия: Торгово-экономические отношения (на английском языке);
- Генезис рыночных реформ в России (на русском языке);
- СМИ и PR в Болгарии (на русском языке);
- Уго Чавес и Боливарианская революция (на русском языке);
- Коммуникационный менеджмент. Консалтинг в связях с общественностью (на русском языке);
- Связи с общественностью и коммуникационный менеджмент: зарубежный опыт (на русском языке);
- Внешняя политика США: коммуникационный аспект (на русском языке);
- Коммуникационный менеджмент в мировой политике и бизнесе (в двух томах, на русском языке);
- Растущая роль коммуникационного менеджмента в мировой политике и бизнесе (на английском языке);
- Ультралевый терроризм в Германии: основные направления деятельности «Фракции Красной Армии» (РАФ) и ее коммуникационное обеспечение (на русском языке);
- Коммуникационный менеджмент во внешней политике Франции конца XX века (на русском языке);
- Коммуникационный менеджмент и стратегическая коммуникация (на русском языке);
- Кризис, Армия, Революция (на русском языке);
- Президенты под медиа-прицелом: практика информационно-психологического противоборства в Латинской Америке;
- Уго Чавес и информационно-психологическое противоборство в Венесуэле (на русском языке);
- Коммуникационный менеджмент и стратегическая коммуникация: современные технологии глобального влияния и контроля (на русском языке);
- Стратегическая провокация «Украина» (на русском языке);

• Коммуникация и терроризм (на русском языке),

• Стратегическая коммуникация в отношениях ЕС и России: напряженность, вызовы и возможности (на английском языке);

• Злонамеренное использование искусственного интеллекта и международная информационно-психологическая безопасность в Латинской Америке (на английском языке);

• Злонамеренное использование искусственного интеллекта как угроза информационно-психологической безопасности: Северо-Восточная Азия и остальной мир (на русском и английском языках);

• Существующие и перспективные угрозы международной информационно-психологической безопасности через злонамеренное использование искусственного интеллекта и возможные пути их нейтрализации (на русском языке);

• Частичная легитимность администрации Байдена и глобальный системный кризис (на испанском языке);

• Эксперты о злонамеренном использовании искусственного интеллекта и вызовах международной информационно-психологической безопасности (2021 и 2022 гг., на английском языке).

Среди авторов этих книг более 100 исследователей из 29 стран Европы, Азии, Северной и Южной Америки.

Одним из последних проектов МЦСПИК является развитие международных ассоциаций, работающих в различных областях стратегических исследований и стратегической коммуникации. В деятельности этих ассоциаций принимают участие ведущие ученые, руководители и сотрудники государственных, частных структур и неправительственных организаций из Азии, Океании, Африки, Европы, Южной и Северной Америки (подробнее см. на GlobalStratCom: http://globalstratcom.ru/globalstratcom-eng/).

Электронная почта: icspsc_office@mail.ru, icspsc@mail.ru.

GlobalStratCom

Россия развивает сотрудничество с различными регионами мира. Платформа GlobalStratCom направлена на развитие ассоциаций в различных областях стратегических исследований и стратегической коммуникации. В настоящее время в стадии развития:

• Европейско-российская экспертная сеть коммуникационного менеджмента (ЕРЭСКМ).

• Российско-латиноамериканская ассоциация стратегических исследований (РЛАСИ).

В деятельности этих объединений принимают участие ведущие ученые, руководители и ответственные сотрудники государственных и частных структур и неправительственных организаций из Азии, Океании, Африки, Европы, Южной и Северной Америки.

Области исследований

• Вызовы и угрозы национальной и международной безопасности: общие интересы и возможные направления сотрудничества России и других стран;

• Вооруженные силы и политика;

• Разрешение конфликтов и кризисное управление;

• Участие в миротворческих миссиях;

• Злонамеренное использование искусственного интеллекта и информационно-психологическая безопасность.

• Войны и военные конфликты;

• Перспективные модели общественного и политического развития;

• Новые технологии и их влияние на социальное развитие и вопросы безопасности;

• Деятельность правоохранительных органов;

• Терроризм и коммуникация

• Вооруженные силы, государство и общество;

• Стратегическая коммуникация;

• Военная история;

• Стратегические исследования как направление сотрудничества России и других стран;

• Исследования войны и мира.

Для получения дополнительной информации посетите сайт GlobalStratCom.

Евгений Николаевич ПАШЕНЦЕВ

Ведущий научный сотрудник Дипломатической академии МИД РФ, профессор магистерской программы «Искусственный интеллект и международная безопасность» Санкт-Петербургского государственного университета. Директор Международного центра социально-политических исследований и консалтинга. Координатор Международной группы по исследованию угроз международной информационно-психологической безопасности посредством злонамеренного использования искусственного интеллекта (Research MUAI). Член международного консультативного совета Comunicar (Великобритания) и редакционной коллегии журнала The Journal of Political Marketing (США). Автор/соавтор /редактор 40 книг и более 250 научных статей. За последние 15 лет принимал участие в 210 международных конференциях и семинарах, география которых насчитывает 19 стран. Почетный научный сотрудник Бирмингемского университета (октябрь-ноябрь 2005 г.). В 2021-2023 гг. возглавлял работу российской группы исследователей в рамках совместного проекта, поддержанного Российским фондом фундаментальных исследований (РФФИ) и Вьетнамской академией социальных наук (ВАСН), по теме «Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в Северо-Восточной Азии». В настоящее время участвует в грантовом проекте «Обеспечение информационно-психологической безопасности России в условиях развития Индустрии 4.0 и гибридных угроз со стороны коллективного Запада», поддержанном СПбГУ. Сферы научных и профессиональных интересов: искусственный интеллект и глобальные перемены, злонамеренное использование искусственного интеллекта и международная информационно-психологическая безопасность, стратегическая коммуникация.

Злонамеренное использование ИИ и вызовы информационно-психологической безопасности в странах БРИКС

Координатор проекта: Е. Н. Пашенцев

Издание МЦСПИК про поддержке Research MUAI

Июнь 2024, Москва

ISBN 978-5-00227-263-1

9 785002 272631

Адреса для комментариев: icspsc@mail.ru; icspsc_office@mail.ru.

Издано в Российской Федерации издательством «OneBook.ru», ООО «САМ Полиграфист».