DOI 10.35775/PSI.2024.106.6.021
УДК 32.327

И.В. СУРМА
кандидат экономических наук,
начальник отдела НАМИБ, доцент кафедры
международной и национальной безопасности
Дипломатической академии МИД РФ,
профессор Академии Военных Наук,
Россия, г. Москва

ВЫЗОВЫ И УГРОЗЫ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА КАК УНИВЕРСАЛЬНОГО ИНСТРУМЕНТА СОЦИАЛЬНО-ПОЛИТИЧЕСКОЙ И ЭКОНОМИЧЕСКОЙ ТРАНСФОРМАЦИИ СОВРЕМЕННОГО ОБЩЕСТВА

В статье показано, что сегодня цифровые технологии и технологии искусственного интеллекта (ИИ) выступают универсальным инструментом социально-политической и экономической трансформации современного общества. Автор отмечает, что этот процесс существенно ускорится с появлением генеративного искусственного интеллекта (GenAI) и на фоне явных успехов, генеративный искусственный интеллект ставит новые серьезные задачи в области кибербезопасности. Эта новая технология может быть использована для создания более сложных фишинговых сообщений и электронных писем, а также для создания возможностей для злоумышленников выдавать себя за людей или организации, что приведет к увеличению числа случаев хищения персональных данных или мошенничества. Распространение «глубоких подделок», создающих более реалистичные видео-, аудиозаписи или изображения, может нанести серьезный ущерб как государствам и организациям, так и отдельным лицам. Автор отмечает, что достаточно высок потенциал использования технологий GenAI с целью манипулирования и распространения дезинформации, а также применения их в формате государственного кибертерроризма, в том числе и в процессе организации и проведения цветных революция и операций по политической дестабилизации.

Ключевые слова: *искусственный интеллект, кибербезопасность, ООН, угрозы информационной безопасности, кибертерроризм, генеративный искусственный интеллект*

По оценкам мировой финансовой индустрии вклад технологии искусственного интеллекта (ИИ) в глобальную экономику может составить от 10 до 15 триллионов долларов США к 2030 году. Об этом заявил генеральный секретарь ООН Антониу Гутерриш на заседании Совета Безопасности всемирной организации

по проблематике искусственного интеллекта в штаб-квартире ООН в Нью-Йорке 18 июля 2023 года [8]. Совет из 15 членов, заслушал Джека Кларка, соучредителя известного стартап-проекта в области искусственного интеллекта Anthropic, и профессора Цзэн И, содиректора Китайско-британского исследовательского центра этики и управления искусственным интеллектом. В ходе брифинга Генеральный секретарь ООН подчеркнул, что скорость и охват новой технологии беспрецедентны, поражают и шокируют возможности генеративного искусственного интеллекта, а число пользователей ChatGPT только за два месяца после запуска достигло 100 миллионов человек. Заметим, что по недавним оценкам экономистов Goldman Sachs после того, как по крайней мере половина компаний во всем мире перейдет на технологии искусственного интеллекта, мировой ВВП может увеличиться в среднем на 7% за 10 лет. Эксперты прогнозируют, что ИИ способен повысить мировую производительность в среднем на 1,4% в год.

При этом, в своем выступлении генеральный секретарь ООН отметил, что даже создатели ИИ не имеют достаточного представления о том, к чему может привести этот технологический прорыв, но уже сейчас понятно, что искусственный интеллект будет оказывать влияние на все сферы жизни человечества, включая ООН.

Сегодня цифровые технологии и технологии искусственного интеллекта (ИИ) выступают универсальным инструментом социально-политической и экономической трансформации современного общества и этот процесс ускорится с недавним появлением генеративного ИИ (GenAI). Но на фоне явных успехов, GenAI ставит новые серьезные задачи в области кибербезопасности. Эта новая технология может быть использована для создания более сложных фишинговых сообщений и электронных писем, а также для создания возможностей для злоумышленников выдавать себя за людей или организации, что приведет к увеличению числа случаев хищения персональных данных или мошенничества. Распространение «глубоких подделок», создающих более реалистичные видео-, аудиозаписи или изображения, может нанести серьезный ущерб как государствам и организациям, так и отдельным лицам [7].

Модели GenAI могут быть уязвимы к отравлению данных и атакам на входные данные [4]. Атаки с отравлением данных пытаются воздействовать на модели ИИ на этапе обучения путем добавления специальных элементов в набор обучающих данных, чтобы снизить точность обучения или скрыть вредоносные действия, которые ожидают специальных входных данных [6]. Атаки на входные данные аналогичны, но они пытаются воздействовать на модели ИИ во время работы. GenAI может быть подвержен аналогичным атакам при помощи манипулирования данными. Инструменты, такие как SEO или контент, созданный GenAI, могут быть использованы для манипулирования средой данных GenAI в злонамеренных целях. Хотя в настоящее время этот риск может быть несущественным, поскольку текущие модели GenAI обучены и работают на данных, полученных из Интернета до 2021 года, ситуация может быстро измениться по мере того, как все больше людей будут узнавать о возможностях GenAI и стремительно

внедрять его. Кроме того, особенно уязвимыми могут оказаться приложения GenAI корпоративного уровня, поскольку они используют более узкие наборы данных, которые могут стать мишенью для специально разработанных средств кибервзлома. Еще более грозным видится сочетание использования технологий GenAI и ядерного оружия, так как обезоруживающий (превентивный) киберядерный удар под управлением искусственного интеллекта может привести человечество к апокалипсису. Кроме всего прочего, достаточно высок потенциал использования технологий GenAI по мониторингу и манипуляции, что может привести к усилению концентрации власти. Так, с одной стороны, правительства ряда стран могут использовать технологии искусственного интеллекта для нарушения гражданских прав и свобод, распространения дезинформации и подавления инакомыслия, а, с другой, те же правительства могут применять эти технологии в формате государственного кибертерроризма, в том числе и в процессе проведения цветных революций.

Еще один фактор, вызывающий озабоченность – это то, что современные модели GenAI все чаще подвергаются успешным атакам «взлома» [2]. Эти атаки основаны на разработке наборов тщательно продуманных подсказок (последовательностей слов, фраз или предложений), позволяющих обойти правила и фильтры GenAI или даже вставить вредоносные данные или инструкции (последнее иногда называют «атакой внедрения подсказок»). Такие атаки могут нарушить работу GenAI или привести к утечке конфиденциальных данных.

Учитывая, что технология GenAI – явление относительно новое, полный спектр и масштаб ее уязвимости к кибератакам еще не до конца изучен. Тем не менее уже первые признаки указывают на потенциально серьезные проблемы, которые требуют тщательного изучения, особенно когда лица, принимающие решения, рассматривают возможность широкомасштабного внедрения технологии в таких чувствительных и жестко регулируемых секторах, как энергетика или финансы, а также в случае корпоративных систем GenAI.

Кроме того, технологии ИИ потенциально могут привести к появлению новых источников и каналов передачи системных рисков. В частности, широкое использование ИИ может привести к повышению однородности оценок рисков (например, кредитных решений в финансовом секторе), а также к возникновению риска «вне выборки», что в сочетании с ростом взаимосвязанности может создать условия для нарастания системных рисков.

GenAI, вероятно, приведет к возникновению системных рисков, аналогичных рискам ИИ, но при этом вызовет и свои собственные. Эти проблемы могут усугубляться легкостью и практичностью формирования отчетов GenAI, а также отсутствием эффективного режима регулирования. Такая ситуация позволит усилить соблазн чрезмерно полагаться на GenAI, что, в свою очередь, может привести к увеличению риска заражения и формированию системных рисков [3].

Таким образом, по мнению экспертов технологии GenAI имеют большие перспективы для применения в финансово-экономическом секторе, однако к ним следует подходить с осторожностью. Технологии GenAI могут значительно

повысить эффективность, улучшить качество обслуживания клиентов, укрепить систему управления рисками и соблюдения нормативных требований. Однако присущие GenAI риски могут существенно повлиять на репутацию и надежность и, в конечном счете, подорвать доверие общества. Приложения GenAI корпоративного уровня потенциально могут помочь снизить некоторые риски, присущие государственным GAI.

Аналитики считают, что со временем нормативно-правовая база будет развиваться и усовершенствоваться, чтобы обеспечить нормальное и эффективное использование приложений GenAI, однако в настоящее время, исходя из выявленных рисков и угроз, необходимо принять промежуточные меры. Использование GenAI требует тщательного человеческого контроля, соизмеримого с рисками, которые могут возникнуть в результате применения технологии в деятельности различных организаций (например, использование ИИ для анализа или рекомендаций по сравнению с внедрением систем ИИ, способных принимать и исполнять решения). Органы государственного надзора должны укреплять свой институциональный потенциал и активизировать мониторинг и наблюдение за развитием технологий, уделяя наиболее пристальное внимание тому, как они применяются в финансово-экономическом секторе, так как эта сфера более других готова наиболее динамично внедрять технологии GenAI. Для этого необходимо улучшить взаимодействие с заинтересованными сторонами из государственного и частного секторов, а также сотрудничать с юрисдикциями на международном и региональном уровнях.

Следует отметить, что методы современных злоумышленников становятся все более и более изощренными по мере технологического развития и увеличения ландшафта угроз. Рассматривая особенности обеспечения информационной безопасности в условиях цифрового общества, нужно не забывать и о недавних случаях с аппаратными уязвимостями класса Spectre и Meltdown, которые наиболее ярко отражают актуальные проблемы кибербезопасности. Существует 27 уязвимостей (13 Spectre-подобных уязвимостей и 14 Meltdown-подобных уязвимостей), которые были обнаружены в архитектурах трех крупнейших мировых производителей процессоров AMD, ARM и Intel, использующихся в миллионах компьютерах по всему миру. Последствия этих проблем примерно одинаковые, то есть возможность получения несанкционированного доступа к данным, которые дают злоумышленникам дополнительные возможности по установлению контроля над практически каждым устройством в мире, включая ноутбуки, мобильные устройства и промышленные системы, повышая риски их эксплуатации, как минимум, на порядок. К сожалению, это серьезные аппаратные уязвимости программными решениями, которые так просто и полностью не исправить. Производители процессоров слишком долго игнорировали эти уязвимости, пытаясь оптимизировать процессоры, для того чтобы увеличить их быстродействие и повысить производительность. Но за все в этом мире приходится платить, что-то удается улучшить, но, а где-то при этом появляются новые и более серьезные проблемы, так как для разработки новых приемов киберпреступники

стали все активнее использовать возможности искусственного интеллекта и, конечно, они не упустят возможность использовать аппаратные уязвимости в архитектурах процессоров ведущих мировых производителей.

Отметим, что влияние ИИ на международную торговлю и мировой экономический рост будет оказываться несколькими способами. Одним из них является макроэкономическое воздействие ИИ и связанные с ним экономические эффекты. Например, если ИИ увеличит рост производительности, то это ускорит экономический рост и предоставит новые возможности для международной торговли. Сегодня темпы роста производительности во всем мире достаточно низкие, и существуют различные этому причины. Одной из причин низкого роста производительности, особенно важной для понимания потенциальной связи с ИИ, является то, что требуется время (технологическое) для внедрения и эффективного использования новых технологий, особенно сложных, оказывающих влияние на мировую экономику в целом, таких как ИИ. Это включает также время на создание достаточно большого запаса капитала для получения совокупного эффекта и дополнительных инвестиций, необходимых для полного использования преимуществ инвестиций в ИИ, включая доступ к квалифицированным специалистам и деловым практикам.

Выделим еще несколько немаловажных потенциальных угроз использования технологии ИИ. Косвенно внедрение ИИ может увеличить разрыв между странами, усилив нынешний цифровой разрыв. Различным странам могут потребоваться разные стратегии и ответные меры, поскольку показатели внедрения ИИ и технологический уровень развития отличаются.

Государства, лидирующие в разработке и внедрению ИИ (в основном это компании из наиболее развитых стран), могли бы увеличить свое доминирование над развивающимися странами. Страны-лидеры в области искусственного интеллекта могли бы получать дополнительные 20-25% чистых выгод по сравнению с сегодняшним днем, в то время как развивающиеся страны могли бы получать только от 5 до 15%. У многих развитых стран, возможно, не останется иного выбора, кроме как использовать ИИ для обеспечения более высокого роста производительности по мере замедления темпов роста их ВВП (во многих случаях, что частично отражает проблему, связанную со старением населения) [5]. Более того, в этих странах размеры заработной платы достаточно высоки, что означает существование больше стимулов для замены рабочей силы машинами, чем в развивающихся странах с низкой заработной платой. Напротив, развивающиеся страны, как правило, используют другие способы повышения производительности, включая внедрение передового опыта и реструктуризацию своих отраслей экономики. Следовательно, у них может быть меньше стимулов продвигать технологии ИИ (которые в любом случае способны принести им относительно меньшую выгоду, чем в странах с развитой экономикой). Некоторые развивающиеся страны могут оказаться исключениями из этого правила. Например, у Китая есть национальная стратегия, направленная на то, чтобы стать мировым лидером в цепочке поставок ИИ, и он вкладывает в это значительные средства.

Еще один вопрос связан с тем, как ИИ может повлиять на компании. Вполне возможно, что технологии искусственного интеллекта могут привести к разрыву в производительности между лидерами (компаниями, которые полностью внедрят инструменты искусственного интеллекта на своих предприятиях в течение следующих пяти-семи лет) и консерваторами (компаниями, которые вообще не внедряют технологии искусственного интеллекта или не полностью внедрят их на своих предприятиях к 2030 году). На одном конце спектра лидирующие компании, вероятно, получат непропорционально большую выгоду и к 2030 году они потенциально могут удвоить свой денежный поток (полученная экономическая выгода за вычетом сопутствующих инвестиций и затрат на переходный период). Это подразумевает дополнительный ежегодный рост чистого денежного потока примерно на 6 % в течение более длительного периода, чем в следующем десятилетии. Лидеры, как правило, имеют прочную стартовую базу в области информационных технологий, более склонны инвестировать в ИИ и положительно оценивают бизнес-обоснование ИИ.

С другой стороны, у консерваторов может наблюдаться примерно 20% снижение денежного потока по сравнению с сегодняшним уровнем при той же модели затрат и прибыли. Одним из важных факторов такого давления на прибыль является наличие сильной конкурентной динамики среди компаний, которая может переместить долю рынка от отстающих к лидерам и может вызвать серьезные дискуссии о неравномерном распределении преимуществ ИИ.

Следующий вопрос касается того, как ИИ может повлиять на рынок труда и работников. Спрос на рабочие места может сместиться с повторяющихся задач в сторону тех, которые ориентированы на социальные и когнитивные аспекты и требуют большего количества цифровых навыков. В профилях должностей, характеризующихся повторяющимися действиями или требующих низкого уровня цифровых навыков, доля общей занятости может сократиться в наибольшей степени примерно до 30% к 2030 году с примерно 40%. Наибольший прирост доли рынка труда может быть достигнут в неповторяющихся видах деятельности и тех, которые требуют высоких цифровых навыков, увеличившись примерно с 40% до более чем 50%. По оценкам экспертов, семь из десяти работников в США столкнутся с влиянием ИИ, но лишь небольшую часть из них полностью заменят технологии. Между тем, ИИ сможет дополнить работу почти двух третей людей. Таким образом, по мнению аналитиков, прямое влияние ИИ на спрос на рабочую силу в ближайшей перспективе может быть отрицательным, но его влияние на рост производительности труда все равно будет положительным.

Нельзя отрицать, что в использовании ИИ есть много преимуществ. Есть причина, по которой он становится таким популярным, и это потому, что технология во многих отношениях делает нашу жизнь лучше, но тем не менее, существуют угрозы и риски, присущие технологии ИИ. Видимо поэтому по результатам обсуждений на заседании Совета Безопасности ООН Антониу Гутерриш заявил в своем выступлении: «Упор на защиту будущих поколений в Уставе ООН дает нам право на то, чтобы привлечь все стороны к предотвращению долгосрочных

глобальных угроз на коллективной основе. ИИ представляет такую угрозу» [1], – и поддержал идею создания органа всемирной организации для регулирования сферы искусственного интеллекта, отметив, что ООН может стать идеальной площадкой для выработки глобальных стандартов и подходов в сфере искусственного интеллекта.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК:

1. Гутерриш поддержал создание органа ООН для контроля за искусственным интеллектом // https://www.belta.by/world/view/guterrish-podderzhal-sozdanie-organa-oon-dlja-kontrolja-za-iskusstvennym-intellektom-577561-2023/.
2. ADVERSA. 2023. «Universal LLM Jailbreak: CHATGPT, GPT-4, BARD, BING, ANTHROPIC, and Beyond» // https://adversa.ai/blog/universal-llm-jailbreak-chatgpt-gpt-4-bard-bing-anthropic-and-beyond/.
3. **Atreides Kyrtin.** 2023 Automated Bias and Indoctrination at Scale... Is All You Need. ResearchGate // http://dx.doi.org/10.13140/RG.2.2.16741.88803.
4. **Boukherouaa, El Bachir, Ghiath Shabsigh, Khaled AlAjmi, Jose Deodoro, Aquiles Farias, Ebru Iskender, Alin T. Mirestean and Rangachary Ravikumar.** 2021 «Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance». IMF Departmental Paper 2021/024, International Monetary Fund, Washington, DC.
5. **Nicoletti Leonardo and Dina Bass.** 2023. Humans Are Biased: Generative AI Is Even Worse // Bloomberg Technology + Equality // https://www.bloomberg.com/graphics/2023-generative-ai-bias/.
6. **Papenbrock Jochen and Alexandra Ebert.** 2022. «Best Practices: Explainable AI Powered by Synthetic Data». NVIDIA Technical Blog. May 20, 2023 // https://developer.nvidia.com/blog/best-practices-explainable-ai-powered-by-synthetic-data/.
7. **Ullah Ihsan, Andre Rios, Vaibhav Gala and Susan McKeever.** 2020 «Explaining Deep Learning Models for Structured Data Using Layer-Wise Relevance Propagation» // https://doi.org/10.48550/arXiv.2011.13429.
8. UN Security Council meets for first time on AI risks. By Michelle Nichols // https://www.reuters.com/technology/un-security-council-meets-first-time-ai-risks-2023-07-18/.

I.V. SURMA
Candidate of Economic Sciences,
Head of the Department of the National Association for International Information Security, Associate Professor of the Department of International and National Security of the Diplomatic Academy of the Ministry of Foreign Affairs of the Russian Federation;
professor of the Academy of Military Sciences,
Moscow, Russia

CHALLENGES AND THREATS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES AS A UNIVERSAL TOOL FOR SOCIO-POLITICAL AND ECONOMIC TRANSFORMATION OF MODERN SOCIETY

The article shows that today digital and artificial intelligence (AI) technologies act as a universal tool for socio-political and economic transformation of modern society. The author notes that this process will significantly accelerate with the emergence of generative artificial intelligence (GenAI) and against the background of clear successes, generative artificial intelligence poses new serious challenges in the field of cybersecurity. This new technology could be used to create more sophisticated phishing messages and emails, as well as create opportunities for attackers to impersonate people or organizations, leading to an increase in identity theft or fraud. The proliferation of «deep spoofs» that create more realistic video, audio or image recordings can cause serious harm to States and organizations as well as individuals. The author notes that the potential for using GenAI technologies to manipulate and disseminate disinformation, as well as their use in the format of state cyberterrorism, including in the process of organizing and conducting color revolutions and political destabilization operations, is quite high.

Key words: *artificial intelligence, cybersecurity, UN, threats to information security, cyberterrorism, generative artificial intelligence.*