Злонамеренное использование искусственного интеллекта как угроза информационно-психологической безопасности: Северо-Восточная Азия и остальной мир

Материалы научных семинаров молодых исследователей

Издание Международного центра социально-политических исследований и консалтинга

Москва

2021

Злонамеренное использование искусственного интеллекта как угроза информационно-психологической безопасности: Северо-Восточная Азия и остальной мир

Материалы научных семинаров молодых исследователей

Под редакцией доктора политических наук Д. Ю. Базаркиной

Издание Международного центра социально-политических исследований
и консалтинга

Москва

2021

УДК 004.8-049.5(082)
ББК 32.813я43
3-68

Научное издание

Злонамеренное использование искусственного интеллекта как угроза информационно-психологической безопасности: Северо-Восточная Азия и остальной мир. Материалы научных семинаров молодых исследователей в Санкт-Петербургском государственном университете и Дипломатической академии МИД России, 27 октября и 25 ноября 2021 г. / Под ред. доктора политических наук Д. Ю. Базаркиной. Издание Международного центра социально-политических исследований и консалтинга. – М.: ООО «Сам Полиграфист», 2021. – 52 с.

ISBN 978-5-00166-511-3
Оформление обложки: Shutterstock.

В сборнике представлены доклады молодых исследователей по проблеме злонамеренного использования искусственного интеллекта (ИИ) с целью подрыва информационно-психологической безопасности общества. Доклады, сделанные на научных семинарах в Санкт-Петербургском государственном университете и Дипломатической академии МИД России осенью 2021 года, содержат как анализ политически обусловленных угроз информационно-психологической безопасности отдельных стран и регионов (с особым вниманием к положению в Северо-Восточной Азии), так и оценки текущих и будущих угроз злонамеренного использования отдельных технологий ИИ (таких, как дипфейки).

Исследования, опубликованные в ряде докладов, выполнены при финансовой поддержке Российского фонда фундаментальных исследований (РФФИ), проект №21-514-92001 («Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в Северо-Восточной Азии») РФФИ и Вьетнамской академии общественных наук (ВАОН).

Malicious Use of Artificial Intelligence as a Threat to Psychological Security: Northeast Asia and the Rest of the World. Proceedings of the seminars of young researchers at Saint Petersburg State University and the Diplomatic Academy of the Ministry of Foreign Affairs of Russia, October 27 and November 25, 2021. / Ed. by Darya Bazarkina DSc. Edition of the International Center for Social and Political Studies and Consulting. – Moscow: LLC «SAM Polygraphist», 2021. – 52 pp.

ISBN 978-5-00166-511-3
Cover image: Shutterstock.

Signed to print 22.11.2021. Digital printing. Order № 104876. Circulation 16 copies.

Printed in the printing house «OneBook.ru» LLC «SAM Polygraphist».

109316, Moscow, Volgogradsky Avenue, Building 42, Bldg. 5, Technopolis Moscow.

www.onebook.ru

© МЦСПИК, Д. Ю. Базаркина (редактор), 2021
© Авторы, 2021

Оглавление

Предисловие (Александр Райков) 4

Введение (Дарья Базаркина) 6

Challenges of AI-Powered Cyberattacks for Political Stability and Psychological Security in Latin America (Raynel Batista, Javier Anias, and Liomar Rodríguez) 9

Malicious Use of Deepfakes as a Threat to Psychological Security (Valeriia Bulva) 14

Практики использования ИИ для проведения кампаний по управлению восприятием и пропаганды в Северо-Восточной Азии (Юрий Колотаев) 17

MUAI Threats in Taiwan's Cybersecurity Policy: Problems, Vulnerabilities, Prospects (Socio-Psychological Aspect) (Darya Matyashova) 23

Технология дипфейк как вызов информационно-психологической безопасности КНР и Северо-Восточной Азии (Екатерина Михалевич) 27

Malicious Use of Artificial Intelligence – New Psychological Security Risks: an Indian Perspective (Kallakuri Radhakrishna) 32

Страх перед искусственным интеллектом как фактор информационно-психологической дестабилизации (Сергей Себекин) 35

Злонамеренное использование искусственного интеллекта во Вьетнаме в информационно-психологическом аспекте (Нгуен Минь Тыонг) 38

Деятельность НАТО по борьбе со злонамеренным использованием искусственного интеллекта: тенденции и противоречия (Олег Филатов) 41

Дипфейки как угроза информационно-психологической безопасности личности и общества: взгляд Евросоюза (Любовь Шматкова) 45

Заключение (Дарья Базаркина) 48

Информация об авторах 49

Notes on Contributors 50

Основные сокращения, принятые в сборнике:

ИИ (AI) – искусственный интеллект (artificial intelligence)

ЗИИИ (MUAI) – злонамеренное использование искусственного интеллекта (malicious use of artificial intelligence)

ИПБ (PS) – информационно-психологическая безопасность (psychological security)

Предисловие

Об ИИ за 70 лет с момента появления этого термина написаны тысячи книг и статей. История зарождения феномена ИИ началась много раньше, и, скорее всего, ее истоки коренятся в зарождении человеческого знания. Непосредственный научный фундамент ИИ заложен, полагаю, в конце XVIII века развитием термодинамики, а уже позже – логики, когнитивной психологии, феноменологической ветви философии. 40-е годы прошлого века ознаменовались идеями искусственного нейрона, логических паттернов, методов решения обратных задач и пр. За многие годы образ ИИ впитал в себя моменты восхвалений и разочарований, научных побед и практических поражений, радости творчества и переживаний поиска.

Проблема ИИ довольно подробно рассмотрена исследователями под различными углами зрения. Однако основной акцент, как правило, делается на формализованной, технологической стороне. Это мало чем отличает ИИ от алгоритмического инструмента. Вместе с тем, все чаще специалисты говорят о человеко-машинном комплексе, «погружая» ИИ в гибридную реальность социально-технологического развития. Становятся актуальными вопросы этического порядка, предупреждения угроз злонамеренного применения ИИ.

В этом контексте подготовлены доклады, вошедшие в настоящий сборник. Они написаны молодыми учеными – представителями различных областей гуманитарного и технического знания. Человек сегодня не успевает осмысливать все аспекты мировых трендов в экономике, политике, социальной и научно-технической сферах. Тем важнее налаживать сотрудничество специалистов в рамках междисциплинарных проектов. В частности, это нужно для оценки как позитивного, так и негативного воздействия, которое посредством ИИ может быть осуществлено в информационно-психологическом пространстве. Каждый человек должен знать не только то, как можно использовать ИИ на пользу обществу, но и быть предупрежден о том, как злонамеренные субъекты будут тому препятствовать или применять ИИ для манипулирования общественным сознанием.

Молодые исследователи, чьи работы вошли в сборник, бросают свежий взгляд на угрозы информационно-психологической безопасности, обусловленные злонамеренным использованием ИИ, концентрируя внимание не только на опыте Северо-Восточной Азии (что зафиксировано на титуле сборника) – одного из регионов, где ИИ развивается наиболее динамично, – но и других регионов, в которых идет его внедрение, всегда и везде встречая отклик общества.

Чрезвычайно интересными представляются соображения авторов статей о возможных вариантах развития ИИ. Только в рамках междисциплинарного подхода возможно глубокое понимание ИИ как одного из ресурсов, эффективность использования которого будет зависеть от напряженности международных отношений, особенностей экономической и политической системы. Одно дело, если ИИ используется демократическими правительствами в мирно настроенном гражданском

обществе, и совсем другое – если он станет ресурсом авторитарных или тоталитарных систем, какими бы наименованиями они не прикрывали свою античеловеческую сущность.

В завершение хотел бы сказать, что пока что мы имеем дело с так называемым слабым (узким) ИИ, который может только распознавать, прогнозировать и отвечать на вопросы. Ему недоступны чувства, мышление, понимание, объяснение и постановка проблем. Но уже видны контуры сильного ИИ, который, скорее всего, будет представлять собой гибридную систему. В эту систему будут обязательно включены человек и общество как активно влияющие на ситуацию акторы. При построении сильного ИИ принимаются во внимание явления коллективного бессознательного, физического дуализма, термодинамических и релятивистских эффектов и многого того, чего принципиально не может охватить современный узкий ИИ. До создания сильного ИИ, по всей видимости, еще далеко. Вместе с тем, по мере движения к нему, ответы на самые разные вопросы взаимодействия людей с ИИ, затронутые в многих докладах, будут становиться всё более значимыми. Потенциал статей сборника позволяет судить, что у молодежи есть ресурс осилить решение таких вопросов в будущем.

Полагаю, сборник будет полезен большому кругу читателей, и особенно тем, кто желает разобраться, какие выгоды и риски несет в себе повсеместное внедрение ИИ, и что оно означает для сфер национальной и международной безопасности сейчас и в дальнейшем.

Доктор технических наук, профессор,
руководитель центра интеллектуальных технологий
Национального центра цифровой экономики МГУ им. М. В. Ломоносова,
ведущий научный сотрудник
Института проблем управления им. В. А. Трапезникова РАН,
Действительный государственный советник Российской Федерации 3-го класса,
лауреат премии Правительства Российской Федерации в области науки и техники

Александр Николаевич Райков

15 ноября 2021 г.

Введение

Настоящий сборник был подготовлен по итогам двух научных семинаров молодых исследователей из Вьетнама, Индии, Кубы и России. Семинары прошли 27 октября в Санкт-Петербургском государственном университете и 25 ноября 2021 г. в Дипломатической академии МИД России в рамках проекта «Злонамеренное использование искусственного интеллекта и вызовы информационно-психологической безопасности в Северо-Восточной Азии», запланированного на 2021 – 2022 гг. Проект реализуется научными коллективами из России и Вьетнама на основе грантов Российского фонда фундаментальных исследований (проект №21-514-92001, конкурс ВАОН_а) и Вьетнамской академии общественных наук соответственно. В проектный коллектив входят авторы ряда представленных в сборнике работ (Юрий Колотаев, Дарья Матяшова и Екатерина Михалевич).

Главная цель проведенных семинаров – выяснить мнение молодых исследователей об угрозах ИПБ, обусловленных попаданием технологий ИИ в руки злонамеренных акторов (преступных, в том числе террористических, организаций, нечистоплотных политических и бизнес-субъектов и т. д.). В условиях гибридизации международных процессов и беспрецедентной цифровизации практически всех сфер жизни человека такие акторы способны наносить серьезный ущерб обществу в ходе асимметричного противоборства с национальными и наднациональными правительственными структурами, используя ИИ в пропаганде, вербовке, дезинформационных кампаниях, при распространении клеветы, наносящей ущерб репутации общественных, государственных, бизнес-структур, международных организаций. Отдельные виды угроз представляют информационно-психологические последствия взлома ИИ в системах управления физическими объектами (производство, транспорт и т. д.), а также кампании, направленные на дискредитацию самого ИИ и внедряющих его структур. На фоне растущей международной напряженности не исключено использование ИИ для дестабилизации ИПБ противника не только негосударственными акторами, но и государственными или наднациональными – против экономического конкурента или политического оппонента.

Противодействие угрозам злонамеренного использования искусственного интеллекта (ЗИИИ) в сфере ИПБ требует скоординированных действий специалистов в разных областях знания (инженеров, программистов, политологов, юристов, психологов и др.), а также укрепления международного сотрудничества. В силу этого при подготовке семинаров и сборника большое внимание уделялось обмену идеями с зарубежным коллегами, представляющими как технические, так и гуманитарные науки. Представленные доклады охватывают разные виды и аспекты ЗИИИ в сфере ИПБ и опыта противодействия ему.

Целый ряд работ в данном сборнике содержит оценки угроз ИПБ той или иной страны или региона, обусловленных ЗИИИ. Так, коллектив кубинских исследователей (Райнел Батиста, Хавьер Матиас и Омар Родригес) акцентирует внимание на угрозе использования ИИ для осуществления геополитического влияния путем атак на культуру

латиноамериканского общества. Кубинские авторы анализируют масштабные кампании стратегической дезинформации в сочетании с кибератаками на основе ИИ, осуществленные в ходе выборов в Латинской Америке, где проявились манипуляции культурными установками граждан. Юрий Колотаев определяет в своем докладе широкий спектр технологий ЗИИИ, применяемых в настоящее время в Северо-Восточной Азии, выделяя политические и экономические факторы роста масштабов ЗИИИ: став одним из передовых центров развития цифровых технологий, регион сохраняет в себе множество конфликтных очагов разной степени интенсивности. Один из таких очагов – Тайвань, чья политика кибербезопасности в аспекте борьбы с ЗИИИ стала предметом анализа в докладе Дарьи Матяшовой, охватившей вниманием не только правительственные инициативы, но и вызовы общественному доверию к тайваньским институтам управления, способные спровоцировать ЗИИИ.

Индийский взгляд на проблему ЗИИИ в сфере ИПБ представил Каллакури Радхакришна, который на примерах дезинформационных политических кампаний в Индии доказывает актуальность борьбы с гибридными угрозами, усиленными посредством ИИ, в особенности в периоды кризисов, усиливающих эмоциональность восприятия информации гражданами. Нгуен Минь Тыонг выделяет в своей работе проблемы, актуальные для Вьетнама, среди которых не последнее место занимает зависимость молодежи то компьютерных игр, часто усиливающаяся «благодаря» умелому использованию ИИ и снижающая способность к критическому мышлению.

Крайне актуальна проблема использования в манипулятивных целях технологии глубокий фейков, или дипфейков (deepfakes), различные аспекты которой анализируют некоторые авторы сборника. Валерия Булва оценивает угрозу злонамеренного использования дипфейков на фоне факторов, усиливающих информационно-психологическое воздействие на человека: избыточности потока информации, «фильтрующего пузыря» (подбора алгоритмами онлайн-контента на основе анализа данных о пользователе, к примеру, местоположения и истории поисковых запросов), создающего лишь иллюзию свободы и неограниченности выбора информации в цифровой среде. Екатерина Михалевич, анализируя ситуацию в Северо-Восточной Азии с особым вниманием к КНР, выделяет такие угрозы, обусловленные злонамеренным использованием дипфейков, как, к примеру, ущерб репутации бизнес-структур и даже правительству страны их базирования. Любовь Шматкова на основе анализа источников Европейского союза и его государств-членов дает характеристику угроз использования дипфейков, признаваемых в ЕС первостепенными, – от онлайн-преступлений в отношении детей до влияния на результаты расследований.

Доклад Олега Филатова, посвященный подходу НАТО к противодействию ЗИИИ, затрагивает крайне важную проблему использования ИИ военными блоками, которое при нарастании противоречий на международной арене может приобрести злонамеренный характер. Это не только создает военную опасность, но и наносит ущерб ИПБ, по меньшей мере, способствуя нагнетанию в обществе страха перед военным конфликтом. На поиск ответа на вопрос, как страхи, обусловленные ЗИИИ, могут изменить национальный и международный ландшафт безопасности, вплоть до глобальных трансформаций миропорядка, направлен теоретический доклад Сергея Себекина.

Опыт проведения научных семинаров в СПбГУ и Дипломатической академии показал, что в исследовательском сообществе растет стремление к поиску путей преодоления угроз ЗИИИ и понимание важности междисциплинарного подхода и международного сотрудничества на этом направлении. Представляется, что доклады, опубликованные в данном сборнике, станут шагом на пути к углублению научного сотрудничества российских и зарубежных молодых исследователей.

Доктор политических наук,
ведущий научный сотрудник Института Европы РАН

Дарья Юрьевна Базаркина

Challenges of AI-Powered Cyberattacks for Political Stability and Psychological Security in Latin America

Raynel Batista, Javier Anias, and Liomar Rodríguez

Introduction

AI is active in many aspects of our society. The emergence of the big data paradigm has drawn even great attention to computational social science. Data revolution is pushing information-based society to a cognitive new era. Data is creating a global cross-cultural society, supporting the idea of cultural competency as a mechanism of social influence, and setting the distribution of power as an indication of security from a socio-cultural perspective. Disruptive technologies are shaping a new world order and democratic institutions are witnessing the rise of a new way of governance (Barnett and Duvall, 2005). AI-based predictive analytics and prognostic weapons are such technologies to influence public consciousness. AI-based technologies interact with people's cognitive assumptions and provide any individual, group, organization or nation-state to influence on public consciousness as a new kind of weapon in the global system.

The critical role of culture in technology transformation allows to understand how AI-driven technologies are used as a geopolitical weapon targeting people culture to influence regional political stability (Batista et al., 2020). The rise of AI-powered cyber-attacks combined with social media propaganda manipulation in Latin American politics and national economies are dramatically changing the regional balance.

Cyber autonomy and cross-cultural psychology

Cognitive automation and robotics are modelling people assumptions, affecting cognition and handle cultural models (Pashentsev, 2020). Cultures drive technological development and technologies are absorbed into the lives of people, affecting their culture and way of life. However, perceptions and understandings of AI are likely to be profoundly shaped by local cultural and social contexts (Cancialosi, 2018). Cross-cultural psychology looks at how cultural factors influence human behavior and offers alternative explanations for how identities and cultures interact. The fundamental belief in cultural psychology is that not only do human beings shape their cultures, but that cultures also shape human beings. Cultural psychology draws from psychology, anthropology, and sociology, among others, to better understand how people are both shapers of and shaped by their individual cultures.

Recent episodes (2016-2021) in Brazil, Argentina, México, Venezuela, Nicaragua, Uruguay, Panamá and Cuba are proofs that AI-based chatbots applying sentiment analysis and machine learning techniques are related to cyber security threats and political stability in Latin America regional balance. However, what makes it different from other regions is a massive cross-cultural based strategic disinformation campaign combined with AI-powered cyberattacks to influence people cognitive assumptions.

According to AI developing race in region, internet penetration, economies growth perspectives and power distribution patterns, most threats to international psychological security caused by the MUAI are, but not limited to:

- Automated disinformation campaigns;
- Automation of social engineering advertisement practices;
- Automation of hacking;
- Robots' users or fake people;
- Automating influence campaigns.

However, last elections in Mexico, Colombia, Argentina, Venezuela and Brazil are a lesson on digital disinformation spread strategy to change political stability (Radu, 2019). The rise of fake news, cyberattacks and social media manipulation in Latin American politics and national economies are proofs that a new automated propaganda machine is driving politics by leveraging automated emotional manipulation (Bazarkina and Pashentsev, 2019). Big data surveillance meets computational psychology (Davos World Economy Forum, 2020).

Cybersecurity and International cooperation

Advancement in digital technologies produces changes in how information is disseminated. Regional Internet penetration is almost 70% of population and over 90% is connected by mobile devices. The Latin American region is the fourth largest mobile phone global market and the second interest region for AI-powered cyber-attacks up to 2018 (ESET, 2018). Almost 90% of corporate cyberattacks targeted Brazil (56,25%), México (22,81%) and Colombia (10,20%) by adware, phishing, ransomware, malware, man-in-the-middle (MitM), denial-of-service (DOS). Almost 66% of cyberattacks targeted two of three Latin American companies in 2020 (CEPAL, 2020). However, cybercrime rates are growing faster in Costa Rica, Guatemala and Panamá since 2020.

In the first half of 2020, approximately 39 million of human-initiated cyber-attacks originated in Latin America. On the other hand, in the same period, 35 million automated bot attacks were registered as coming from the region. Furthermore, attack rates in Latin America were 24% higher than the global average across all channels in 2019 (Konkel, 2019).

However, just few countries (Colombia 2011, Panamá 2013, Paraguay 2017, Chile 2017, Costa Rica 2017, México 2017) have a national cyber security strategy, which indeed exposes the region to possible attacks and cannot stop growing AI-powered cyberattacks to one of four Latin American companies. Leading companies in providing Internet protection services are outsider region and cover almost 60% of companies and public from Peru, Brazil, Colombia, Chile, Mexico, Venezuela and Argentina. The articulation of cybersecurity with a military focus is evident if one looks at the profile of the companies that are dedicated to providing these services, generally also linked to intelligence or arms sales. Cybercrimes cost rate was up to 90 billion US dollars in 2016 mostly at same countries with missing data protection acts and national cybersecurity strategy. Financial cybercrimes are mainly from Brazil and Mexico, where most of regional FinTech are located (Pashentsev and Bazarkina,

2020). Some countries such as Argentina, Ecuador, Bolivia, Nicaragua, Guatemala, Brazil, Cuba, Peru, are working to enhance actual legislation (Bojalil and Vela-Treviño, 2019).

United Nations initiatives like the Governmental Group of Experts (GGE) and the Internet Governance Forum (IGF) have been established to address issues related to internet governance and cybersecurity (Pantserev, 2020). Regional efforts are closer to succeed due to cultural ties, but Latin states are claiming for funding, investment or capacity building.

Cyber-diplomacy has been recently recognized as an international cooperation instrument to neutralize the proliferation of cyber-attacks and sustain the peaceful use of digital technology in the digital age. However, Latin American countries have differences in interests and application of internet norms and cyber-governance.

The Community of Latin American and Caribbean States (CELAC) was the first regional mechanism to permanently group all 33 countries in Latin America (without the United States and Canada), which have put cybersecurity on the trade cooperation's agendas. However, Latin states already developed a regional cybersecurity strategy in 2004. China-CELAC Summit in 2018 welcomed Belt and Road initiative with benefits to Venezuela, Argentina, Bolivia, Nicaragua, Ecuador, Cuba, becoming China the region's second-largest trading partner. Furthermore, EU-CELAC 2021-2023 Strategic Roadmap declared Cybersecurity as a vital element in building trust in the digital environment.

Progress for the region as a whole has been slow ever since, likely due to a combination of factors – the lack and uneven levels of digital penetration; the low urgency for policymakers to coordinate their cybersecurity responses due to a lack of regional high-profile attacks; the lack of financial means to invest in digital security on a national level; a lack of expertise with policymakers and IT professionals. Cybersecurity industry in Latin America expects to receive 12 billion US dollars in investments over the next few years (Konkel, 2019). Perceptions and understandings of threats to international psychological security caused by the MUAI limit nations' efforts to support multilateral model or cyber-sovereignty common strategy.

The Cuba Files: #SOSCuba

Fake news, disinformation propaganda campaign and AI-powered cyber-attacks fueled a series of protests "seemed" against the Cuban government and the ruling Communist Party. Cuba digital landscape became as a special case of hybrid warfare since July 2021. #SOSCuba was just the origin behind that day's trending hashtag.

Almost two million tweets in just two days (July 10 and 11) fueled an "echo chamber" effect in social media by a swarm of bots with a fake user's profile. Therefore, the background or potential threats of repetition of these practices, connected with psychological warfare operations that come to manipulate public opinion to the benefit of certain interests, have alarmed analysts, political leaders and even the security world.

Cuba is ongoing to develop a national cybersecurity strategy and include issues in several interrelated fields:

- Legal: the new Constitution Act elected for Cuban people include principles for international relations and a Data Protection Law is coming soon;
- Political: last Congress of PCC (Cuban Communist Party) identified these threats as national security priorities;
- Technical: several platforms are implemented to protect national sovereignty (NOVA Operating System for PC and mobile devices, declaration of Telecommunications as public exclusive industry sector, creation of national universities network to cooperate and support Government strategies (devices, infrastructure, operating systems, educational programs, etc);
- Education: Cuban educational programs are focused to promote best practices on develop and use of digital technologies with social responsibility since early stages. Public must be aware of these threats. A new four years educational program of Cybersecurity Engineering started at Universidad de las Ciencias Informáticas;
- Social: Cuban media, TV broadcast system, newspapers, journals, are given special concern to fake news analysis, cyber-attacks and other threats in order to promote a cybersecurity-based culture and safe AI.

References

Barnett, M., & Duvall, R. (2005). Power in International Politics. *International Organization*, *59*(1), 39–75. doi: 10.1017/s0020818305050010

Batista, R., Villar, O., González, H., and Milián, V. (2020). Cultural Challenges of the Malicious Use of Artificial Intelligence in Latin American Regional Balance. In F. Matos, *Proceedings of the 2nd European Conference on the Impact of AI and Robotics, a virtual conference hosted by Instituto Universitario de Lisboa (ISCTE-IUL), Portugal, 22-23 October 2020* (pp. 7–13). Reading, UK: Academic Conferences and Publishing International Limited.

Bazarkina, D. (2021). Advanced Technologies Combating Terrorism in the EU: The Psychological Warfare Aspect. In N. Van der Waag-Cowling & L. Leenen, *Proceedings of the 14th International Conference on Cyber Warfare and Security ICCWS 2019 Hosted by Stellenbosch University and the CSIR, South Africa, 28 February – 1 March 2019* (pp. 23–29). Reading, UK: Academic Conferences and Publishing International Limited.

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

Bojalil, P., & Vela-Treviño, C. (2019). Despuntan las reformas en materia de protección de datos en América Latina - Abierto al Público. Retrieved 3 November 2021, from https://blogs.iadb.org/conocimiento-abierto/es/proteccion-de-datos-gdpr-america-latina/

Cancialosi, C. (2018). The Critical Role Of Culture In Technology Transformation. Retrieved 3 November 2021, from https://www.forbes.com/sites/chriscancialosi/2018/11/06/the-critical-role-of-culture-in-technology-transformation/?sh=2b15fa233807

CEPAL. (2020). Covid-19: The social challenge in times of Covid-19. Retrieved 3 November 2021, from https://www.welivesecurity.com/wp-content/uploads/2018/06/eset_security_report_latam2018.pdf

ESET. (2018). Security Report. Retrieved 3 November 2021, from https://www.welivesecurity.com/wp-content/uploads/2018/06/eset_security_report_latam2018.pdf

Konkel, F. (2019). Report: 2020 is the Year Data Gets Weaponized. Retrieved 3 November 2021, from https://www.nextgov.com/cybersecurity/2019/10/report-2020-year-data-gets-weaponized/160984/

Pantserev, K. (2020). The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability. In H. Jahankhani, S. Kendzierskyj, N. Chelvachandran & J. Ibarra (eds.), *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity* (pp. 37-55). Cham: Springer.

Pashentsev, E., & Bazarkina, D. (2020). *Malicious Use of Artificial Intelligence and International Psychological Security in Latin America. Report by the International Center for Social and Political Studies and Consulting. June*. Moscow: International Center for Social and Political Studies and Consulting.

Radu, S. (2019). Mexico, Colombia and Brazil Faced Heavy Amounts of Disinformation in 2018 Elections. Retrieved 3 November 2021, from https://www.usnews.com/news/best-countries/articles/2019-08-02/mexico-colombia-and-brazilfaced-heavy-amounts-of-disinformation-in-2018-elections

We Are Social. (2020). Digital around the world in april 2020. Retrieved 3 November 2021, from https://wearesocial.com/blog/2020/04/digital-around-the-world-in-april-2020

Malicious Use of Deepfakes as a Threat to Psychological Security

Valeriia Bulva

Deepfakes conceived with the use of AI boosts the development of society. As Pashentsev (2020, p. 13) points out, this technique could serve for developing art, promoting education and treating persons suffering from certain disease. At the same time, a reactionary bellicose state may use deepfakes to influence people consciousness for antisocial purposes (Palmer A., 2020). One of the threats created by synthetic AI product is the harmful treatment of deepfakes aimed at influencing public consciousness. The technique of deepfakes is based on the modification and creation of images, videos and sounds with a view to simulating a real person. The danger consists not in the technology itself but in deliberate attempts to use it for misleading the audience, creating political tensions and undermining personal security. Thus, resorting to deepfakes often leads to the creation of artificial reality, in which the individual is just a puppet ruled by manipulators.

Today, the manipulation of public consciousness has reached a qualitatively new level, which was largely facilitated by scientific and technological progress (Afolabi and Balogun, 2017). For example, modern chatbots employ systems to synthesize the voice of a real person (Pashentsev and Bazarkina, 2020, p. 160). This make a 'fake' more credible and convincing. Though the voice simulation itself is not harmful, the malicious use of deepfakes undermines psychological health of a nation.

The malicious use of deepfakes is compounded by the establishment of conditions for a more active manipulative influence on human consciousness through electronic sources.

First, there are quantitative and qualitative changes in the information environment. Consequently, the amount of information becomes immeasurably large; while on the Internet the user can receive not only a set of information, but also compressed analysis and diagrams on the subject of interest. All this reduces a person's ability to think critically and rationally.

Secondly, the virtualization of the information dissemination space is taking place at a tremendous speed, as a result, now, reality is mixed with fictional images. In such conditions, the vulnerability of a person's consciousness increases, since (s)he cannot distinguish reliable information from fiction. Moreover, a lie is sometimes presented in such a way that it seems more believable than the hidden truth.

Thirdly, the number of contacts is growing, intercultural communication is developing thanks to the universal possibilities of information and communication technologies. The participation in international communication does not know the exclusion of any group, involving terrorist and criminal ones. Moreover, technologies' anonymous nature allow them to remain unnoted.

Various factors increase a person's vulnerability to suggestion. Enormous importance is assigned to his or her personal characteristics (lack of critical thinking, conformism, low self-esteem, etc.). The excessive flow of information, the time and environment chosen for the transmission of information, the technical means that are used in this case, play an additional

role in enhancing the influence on human consciousness. In addition, today it is almost impossible for a person to get out of the “filter bubble”. The illusion of an active Internet user is growing, while, in fact, even when a user tries to independently search for information, Internet sites issue a request based on the analysis of data about the user (his/her location and search history, which determines his/her main preferences).

Today misuse of AI to create dangerous deepfakes poses threats to both national and personal security. Chinese government pays close attention to their dissemination for creating political tensions and sparking instability in a country (Yang, Goh, and Gibbs, 2019). The leadership of Japan focuses on countering porno deepfakes (Bazarkina et al., 2021). The whole world community is concerned with the plausible application of the deepfake techniques by terrorists to promote their ideas and attract new members.

Such a use of deepfakes facilitates the consolidation of the post-truth world, “in which objective facts are less important for the formation of public opinion than appeal to emotions and personal beliefs” (Cubidadi, 2021). In this regard, it is high time to protect ourselves from destructive information and psychological influence through ill-designed ‘fakes’.

The formation of a secure information and cybernetic environment is possible only with an active participation of the state. Control of social networks, prevention of the spread of propaganda (including extremist and terrorist content) are under the jurisdiction of state bodies. In accordance with the new “Fundamentals of the State Policy of the Russian Federation in the Field of Information Security” approved by Russian President Vladimir Putin on 12 April 2021, when developing cooperation in the field of information security, the state relies on interaction with the private sector (President of the Russian Federation, 2021). In the context of countering destructive psychological influence, the formation of the Alliance for the Protection of Children in the Digital Environment on the initiative of nine large Russian companies and the adoption by them of “Voluntary Commitments” is of great importance. One of the areas of activity outlined in the document is the development of content that reflects the main values and foundations of society (constructive content). Such business initiatives have been supported by the state.

The formation of the information culture of society on the basis of high humanitarian technologies seems to be especially important. Close attention should be paid to the observance of ethics in the information environment, since rash, radical publications can lead to an aggravation of the situation. It is necessary to develop the ability of citizens to actively counteract dangerous technologies of communicative influence. Fact checking, forming one’s own opinion, critical thinking are essential prerequisites for ensuring the safety of an individual in a digital society.

References

Afolabi, O., & Balogun, A. (2017). Impacts of psychological security, emotional intelligence and self-efficacy on undergraduates’ life satisfaction. *Psychological Thought*, *10*(2), 247-261. doi: 10.5964/psyct.v10i2.226

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

Bazarkina, D., Dam, V. N., Pashentsev, E., Phan, K. N. A., & Matiashova, D. (2021). The Political Situation in the Northeast Asia and Threats of Malicious Use of Artificial Intelligence: Challenges to Psychological Security. *Social'no-Gumanitarnye Znanija (Social And Humanitarian Knowledge)*, *4*, 212-234. doi: 10.34823/SGZ.2021.4.51655

Cubidadi. (2021). Chto takoe postpravda i kak vyzhit' v more lzhi? (What is post-truth and how to survive in the sea of lies?). Retrieved 3 November 2021, from https://34mag.net/ru/post/kak-vyzhit-v-ehpokhu-postpravdy

Palmer, A. (2018). Experts warn 'deepfakes' videos of politicians could be manipulated. Retrieved 3 November 2021, from https://www.dailymail.co.uk/sciencetech/article-5492713/Experts-warn-deepfakes-videos-politicians-manipulated.html

Pashentsev, E. (2020). Malicious Use of Deepfakes and Political Stability. *Abstracts Of Papers Presented At The European Conference On The Impact Of Artificial Intelligence And Robotics ECIAIR 2020*, 82.

President of the Russian Federation. (2021). Fundamentals of the state policy of the Russian Federation in the field of international information security. Approved by the Decree of the President of the Russian Federation of April 12, 2021 No. 213. Retrieved 5 October 2021, from: http: //www.publication.pravo.gov.ru/Document/View/0001202104120050/

Yang, Y., Goh, B., & Gibbs, E. (2019). China seeks to root out fake news and deepfakes with new online content rules. Retrieved 3 November 2021, from https://www.reuters.com/article/us-china-technology/china-seeks-to-root-out-fake-news-and-deepfakes-with-new-online-content-rules-idUSKBN1Y30VU

Практики использования ИИ для проведения кампаний по управлению восприятием и пропаганды в Северо-Восточной Азии[1]

Юрий Колотаев

Кампании по управлению восприятием являются составной частью современного информационного пространства. Существующие на международной арене противоречия мотивируют различных акторов прибегать к активному информационному воздействию для поддержания своей позиции или искажения позиции оппонентов. Исторически психологическое воздействие и пропаганда являлись составной частью военного или политического противостояния между странами. На сегодняшний день подобные действия получили новую плоскость для воплощения – цифровое пространство.

Социальные сети и другие онлайн-платформы активно превращаются в оружие по всему миру, становясь инструментом для вербовки, радикализации и информационно-психологического противоборства (Marcellino et al., 2020). Отдельные регионы мира при этом имеют особенно значимый потенциал средств по управлению восприятием. Северо-Восточная Азия (СВА) является на сегодняшний день одним из передовых центров развития цифровых технологий. При этом сам регион сохраняет в себе множество конфликтных очагов разной степени интенсивности. Большинство стран ищут способы закрепить своё положение в контексте региональных противоречий и делают ставку, в частности, на ИИ (Asia Pacific Foundation of Canada, 2019). Таким образом ИИ становится составной частью политического влияния и конфликтов в регионе.

На сегодняшний день ИИ является не только наиболее перспективной технологией, определяющей развитие «четвертой индустриальной революции» (Doorsamy, Paul & Marwala, 2020), но одним из наиболее удобных инструментов для манипуляций. Слияние социальных сетей и анализа больших данных на основе ИИ становится источником для новых форм пропаганды, обретающих все большее применение. К ним относятся информационные кампании, основанные на микро-таргитенге и гиперперсонализации, профайлинге и сегментации аудитории, а также дипфейках (Kertysova, 2018). Подобные формы использования ИИ для информационного воздействия особенно подвергают современного человека угрозе быть дезинформированным или стать жертвой манипуляции через основные информационные каналы. Вместе с тем, общая совокупность форм ЗИИИ значительно превышает обозначенные инструменты (Bazarkina & Pashentsev, 2020). ЗИИИ может переходить в смежные сферы, становясь угрозой не только информационного, но и физического пространства. Это делает проблему ЗИИИ особенно актуальной в странах с высоким уровнем развития ИИ, к которым также относятся и государства СВА.

[1] Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 21-514-92001.

Контекст СВА даёт примеры различных форм информационного воздействия. Исходя из политической обстановки можно выделить примеры внутренних и внешних кампаний по управлению восприятием (Corcoran et al., 2019). При этом, характер распространяемой через социальные сети пропаганды зависит от конкретной ситуации.

Наиболее характерные примеры внутренних кампаний информационно-психологического воздействия можно наблюдать в Китайской Народной Республике (КНР), а также в Республике Корея. В каждой из стран информационные кампании – составная часть информационного поля. Вместе с тем, если в КНР основной источник подобных кампаний – это Коммунистическая партия Китая (King et al., 2017), которая стремится к распространению единой партийной линии в информационном поле, то в Республике Корея ситуация с пропагандой и информационными манипуляциями в большей степени децентрализована, т.к. генерируется различными сторонами внутри корейского общества в процессе внутриполитической борьбы (Feigenbaum & Nelson, 2021). Следовательно, внутри стран СВА наблюдаются кампании по управлению восприятием как в состязательной, так и в неконкурентной политической среде.

Внешнее воздействие в регионе происходит чаще в отношении таких акторов как Тайвань (Китайская Республика), а также Гонконг (Cohen et al., 2021). Ключевой стороной, оказывающей подобное давление оказывает КНР, пытаясь распространить своё информационное влияние на свою историческую сферу интересов в регионе. Другие страны, как, к примеру, Корейская Народно-Демократическая Республика (КНДР) также прибегают к распространению пропаганды против ключевых соперников в регионе, однако чаще социальные сети выступают для КНДР вспомогательным средством в процессе кибер-разведки (Cohen et al., 2021). Вместе с тем, на страны СВА наблюдается внешнее воздействие, источники которого находятся за пределами региона. Особая активность отмечается со стороны США и её союзников по блоку НАТО. Наличие американских интересов в регионе обуславливает активное вмешательство в происходящие процессы при помощи информационных кампаний, а также военно-политического взаимодействия с Японией и Южной Кореей (Губин, 2018). При этом в наиболее явной форме подобные действия направлены против КНР, являющейся прямым и наиболее явным конкурентом США в вопросах глобального и регионального влияния (Вартазарова и Кобринская, 2018). Подобное соперничество приводит к тому, что внешнее воздействие на страны СВА происходит как от государств внутри региона, так и от стран, имеющих свои интересы в СВА, но находящихся территориально в других частях света. Новые технологии позволяют более эффективно проводить широкие информационные кампании, вне зависимости от географического расположения.

Вместе с тем, важно отметить, что не каждый пример информационных кампаний в СВА включает в себя применение ИИ. Технологии ИИ в разной степени развиты в странах региона СВА. Среди наиболее активных в этом поле акторов находится КНР, обладающий достаточным уровнем развития ИИ и претендующим на лидерство в этой области (CNews, 2020). Но помимо КНР, ИИ также является потенциальным инструментом информационных кампаний США, обладающих достаточным потенциалом для использования усиленной ИИ пропаганды.

Несмотря на наличие двух сторон, способных к активному применению ИИ в целях управления восприятием, в информационном пространстве преобладает акцент на том,

что именно с КНР связаны наиболее продвинутые кампании в данной сфере (Kumar, 2020). Однако последние годы продемонстрировали и потенциал антикитайских информационных кампаний, связанных с происхождением инфекции COVID-19 (The Global Times, 2021). В подобных кампаниях особенно активную роль занимают информационные каналы США, распространяющие альтернативные сведения о происхождения COVID-19 как внутри СВА, так и за её пределами.

Технически усовершенствованные на основе ИИ кампании строятся не просто на применении ботов, а на интеграции в информационное пространство дипфейков и создание «контент-ферм» для генерации колоссального количества текстовой информации (Insikt Group, 2020). При этом компании по управлению восприятием усложняются и строятся на более усовершенствованных сетях. К примеру, у созданных для распространения пропаганды аккаунтов в профиле находятся фотографии не настоящих людей, а сгенерированные ИИ. Это позволяет усложнить поиск сетей распространения дезинформации или каких-либо других сведений, т.к. подобные изображения могут идентифицироваться системами по распознанию фейковых аккаунтов в качестве уникальных фотографий настоящих пользователей (Carmichael, 2021). Таким образом создаются так называемые, фейковые пользователи.

Установить реальные практики применения ИИ в кампаниях по управлению восприятием с привязкой их к конкретному источнику часто бывает затруднительно (т.к. большая часть исходных данных может быть опубликована анонимными методами). Но, в условиях нарастающего конфликта интересов стран Запада и КНР, каждая из сторон указывает на применение конкурентом ИИ для пропаганды. Американские наблюдатели все чаще обращаются к примерам информационных кампаний КНР в отношении Тайваня и Гонконга (National Security Commission on Artificial Intelligence, 2021). Наиболее крупная связана с президентскими выборами на Тайване в 2020 г., а именно с дискредитацией действующего президента Китайской республики Цай Инвэнь. Предположительно использовалось сочетание информации, генерируемой на контент-фермах, с потенциалом наиболее популярных на Тайване социальных сетей и массового постинга информационных сообщений. Гонконг также находится под давлением аналогичных информационных кампаний, направленных, однако, на разобщение протестующих и снижение их поддержки. При этом активно применяются боты, способные автоматически рассылать информацию (National Security Commission on Artificial Intelligence, 2021).

Сведений о применении ИИ в рамках информационных кампаний со стороны США в СВА в западных информационных ресурсах значительно меньше. Однако масштабы информационных спекуляций на тему происхождения COVID-19 также указывают на потенциальное применение ИИ со стороны США и их союзников против КНР. В китайских информационных ресурсах (например, Baidu) в материалах на тему враждебной антикитайской пропаганды начинает учитываться и применение ИИ. В частности, уделяется внимание интеграции ИИ в сферу национальной безопасности США, прежде всего в рамках ведения информационных войн и противодействия им (人工智能学家, 2019). Таким образом, и США, и КНР начинают фиксировать проникновение ИИ в практику информационного противоборства.

Другие страны СВА также сталкиваются с потоком дезинформации и пропаганды, но в них присутствие ИИ на данный момент регистрируется в меньшей степени, если не брать в расчёт ботов или фейковые изображения различного качества. Это может быть обусловлено приоритетами наиболее продвинутых игроков в области ИИ в регионе. Кроме того, комплексные кампании по управлению восприятием свойственны для одних стран (КНР и КНДР) в большей степени, чем для других (Япония, Республика Корея). Это обусловлено необходимостью активно отстаивать альтернативную информационную повестку первой группе стран, в то время как менее активные в этом отношении государства идут в конформистском ключе с существующим информационным полем, либо ограничиваются развитием своей собственной повестки на национальном уровне.

В регионе существует тенденция к усовершенствованию интеллектуальных систем при проведении кампаний по управлению восприятием. Важную роль играет потенциал страны в развитии ИИ, что говорит о возможных диспропорциях в информационном присутствии тех или иных акторов. Таким образом, лидерство в сфере ИИ будет во многом определять лидерство стран в СВА и других регионах в целом. Следовательно, СВА демонстрирует то, как в рамках региональной системы может распределяться информационное влияние, и как оно может усиливаться при помощи ИИ.

Однако ИИ также может применяться и в ходе информационных кампаний, проводимых из-за пределов региона. В текущих условиях подобные кампании направлены против конкретных стран, в особенности КНР, претендующей на положение лидера в сфере ИИ. Интересы США и союзников в СВА приводят к обострению информационной конфронтации, стимулирующей применение ИИ в пропаганде.

Дальнейший интерес к ИИ во многом будет определять ситуацию в СВА и во всем мире в области информационного влияния. Ключевой задачей стран становится дальнейшее выявление практик ЗИИИ, а также недопущения технической монополии отдельных стран. По этой причине важным является усовершенствование как собственных интеллектуальных систем внутри страны, так и поиск возможностей для глобального регулирования сферы применения ИИ. Но, вне зависимости от прогресса в области международных договоренностей, пример информационных кампаний в СВА демонстрирует необходимость стран быть готовыми к новым, усовершенствованным формам пропаганды и информационно-психологических операций.

Библиография

Губин, А. В. (2018). Изменение роли США в Северо-Восточной Азии: военно-политический аспект. *Проблемы национальной стратегии*, (4), 81-103.

Вартазарова Л.С., Кобринская И.Я. (2018). *США–Китай: борьба двух стратегий и практик мирового лидерства*. М.: ИМЭМО РАН.

Asia Pacific Foundation of Canada. (2019). *Artificial Intelligence Policies in East Asia: An Overview from the Canadian Perspective*. Vancouver, BC: APF Canada. Retrieved 3 November 2021, from https://www.asiapacific.ca/sites/default/files/filefield/ai_report_2019.pdf

Bazarkina, D., & Pashentsev, Y. (2020). Artificial Intelligence and New Threats to International Psychological Security. *Russia In Global Affairs*, *17*(1). doi: 10.31278/1810-6374-2019-17-1-147-170

Carmichael, F. (2021). How a fake network pushes pro-China propaganda. *BBC News*. Retrieved 9 September 2021, from https://www.bbc.com/news/world-asia-china-58062630

CNews. (2020). ИИ двойного назначения. Китай планирует опередить США в области искусственного интеллекта. Retrieved 8 September 2021, from https://www.cnews.ru/articles/2020-02-10_ii_dvojnogo_naznacheniyakitaj_planiruet

Cohen, R., Beauchamp-Mustafaga, N., Cheravitch, J., Demus, A., Harold, S., & Hornung, J. et al. (2021). *Combating Foreign Disinformation on Social Media*. Santa Monica, CA: RAND Corporation. Retrieved 3 November 2021, from https://www.rand.org/content/dam/rand/pubs/research_reports/RR4300/RR4373z1/RAND_RR4373z1.pdf

Corcoran, C., Crowley, B., & Davis, R. (2019). *Disinformation Threat Watch. The Disinformation Landscape in East Asia and Implications for US Policy*. Cambridge, MA: Belfer Center for Science and International Affairs. Retrieved 3 November 2021, from https://www.belfercenter.org/sites/default/files/2019-06/PAE/DisinfoWatch%20-%202.pdf

Doorsamy, W., Paul, B., & Marwala, T. (2020). *The Disruptive Fourth Industrial Revolution*. Cham: Springer International Publishing.

Feigenbaum, E., & Nelson, M. (2021). *The Korean Way With Data: How the World's Most Wired Country Is Forging a Third Way*. Washington, DC: Carnegie Endowment for International Peace. Retrieved 3 November 2021, from https://carnegieendowment.org/files/202108-KoreanWayWithData_final5.pdf

INSIKT GROUP. (2020). *Chinese Influence Operations Evolve in Campaigns Targeting Taiwanese Elections, Hong Kong Protests*. INSIKT GROUP. Retrieved 3 November 2021, from https://www.recordedfuture.com/chinese-influence-operations/?__cf_chl_jschl_tk__=pmd_HLh0IIrIgYXfr1_ohnPm6FCU0i4nnjY1yOq6Ic2SDd4-1631192938-0-gqNtZGzNAjujcnBszQi9

Kertysova, K. (2018). Artificial Intelligence and Disinformation. *Security And Human Rights*, *29*(1-4), 55-81. https://doi.org/10.1163/18750230-02901005

King, G., Pan, J., & Roberts, M. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review*, *111*(3), 484-501. https://doi.org/10.1017/s0003055417000144

Kumar, A. (2020). AI-driven 'deep fakes' next big tool of Chinese disinformation campaign, says study. Retrieved 9 September 2021, from https://www.indiatoday.in/world/story/ai-driven-deep-fakes-next-big-tool-chinese-disinformation-campaign-study-1730903-2020-10-12

Marcellino, W., Cox, K., Galai, K., Slapakova, L., Jaycocks, A., & Harris, R. (2020). *Human-machine detection of online-based malign information* (p. 4). Santa Monica, CA: RAND

Corporation. Retrieved 3 November 2021, from https://www.rand.org/pubs/research_reports/RRA519-1.html

National Security Commission on Artificial Intelligence. (2021). *Final Report*. NSCAI. Retrieved 3 November 2021, from from https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

The Global Times. (2021). Revealing four-step US misinformation campaign against China on virus origins tracing. *The Global Times*. Retrieved 3 November 2021, from https://www.globaltimes.cn/page/202108/1232445.shtml

人工智能学家. (2019). 人工智能技术在美国国家安全领域的应用. *CSDN*. Retrieved 3 November 2021, from https://blog.csdn.net/cf2SudS8x8F0v/article/details/99365013

MUAI Threats in Taiwan's Cybersecurity Policy: Problems, Vulnerabilities, Prospects (Socio-Psychological Aspect)[1]

Darya Matyashova

The recent years are characterized by an emerging focus on cybersecurity in Taiwan. The compelling sights of this process include growing government investments in the field of cybersecurity (ITRI, 2020); applying the "Cybersecurity is a National Security" strategies in 2018 and 2021 (Ya-Chun, 2021), promoting normative initiatives – the "National Cyber Security Programs of Taiwan" of 2017-2020 and 2021-2024 (National Information and Communication Security Taskforce, Executive Yuan, Republic of China (Taiwan), 2021) and the Cybersecurity Management Act; the Global Talent recruitment program (Carnegie Endowment for International Peace & Feigenbaum, 2020) and so forth. The key features of these initiatives are the shared business-state responsibility and the reliance on the international cooperation.

Nonetheless, the given steps reflect the "narrow definition approach" to the cybersecurity. For instance, in the "National Cyber Security Program of Taiwan" AI is regarded in the context of data leakage and hacking, while the communication aspect is omitted (National Information and Communication Security Taskforce, Executive Yuan, Republic of China (Taiwan), 2021). Another case is one of Cybersecurity Management Act that is dedicated to preventing cyberattacks and rising resilience of the critical infrastructure through modifying software and hardware (Cyber Security Management Act – Article Content – Laws & Regulations Database of The Republic of China, 2018). Meanwhile the political situation around Taiwan requires an emphasis on psychological security and its vulnerabilities.

The international situation around Taiwan is determined, firstly, by its place in the global US-Chinese confrontation in the context of the United States' Relations Act, the Chinese initiatives on Taiwan isolation (Carnegie Endowment for International Peace & Feigenbaum, 2020) and the American initiatives to "employ counter-propaganda tools" while raising military capabilities in the region (Gould, 2021) and military rapprochement to Taiwan in the form of weapon contracts and military officials dialogue (Carpenter, 2020); secondly, by the Japanese security steps – parliaments-level talks on possible military cooperation in August 2021 (Reuters, 2021) and work of the "Taiwan's Project Team" seeking to strengthen relations between Japan and Taiwan "amid growing pressure from the side of China" (The Diplomat, 2021); thirdly, by the aggressive information surrounding – in particular, while the US accused China in hacking operations (率, 2021) and spreading fake news, the PRC condemns the USA for "launching a propaganda plan" and spreading false information about China in social media in a global scale (专家：美对中国的信息欺骗行动正火力全开|美国|美国国务院|蓬佩奥_新浪军事_新浪网, 2020).

[1] The reported study was funded by RFBR, project number 21-514-92001.

The domestic political landscape in Taiwan is shaped by two divergent trends of democratic consolidation and deconsolidation. On the one hand, 2000s and 2010s were the period of gradual democracy maturing, enhancement of democratic institutions and procedures as well as of consensual decision-making procedures for Taiwan. Rapprochement of the Kuomintang (KMT) and Democratic Progressive Party (DPP) positions on domestic issues making consensus real and establishing more inclusive mixed electoral system in 2005 contributed to the raising quality of the Taiwanese democracy (Lee, 2018). On the other hand, the row of factors tends to delegitimize formal consolidated institutions, exclude destitute citizens and minorities from collective decision-making and rise preferences of less democratic governance systems (Chull Shin, 2020). These factors include growing political polarization due to the issues of identity (in 2018 approximately 60% regarded themselves as Taiwanese, 34% as Chinese Taiwanese and 3% as Chinese) and rapprochement to the mainland China supported by the KMT (Chatham House et al., 2019); the socio-economic problems – primarily unemployment and wages level. The situation of 2019 was characterized by the "continuing dissatisfaction" caused by "stagnating wages and high youth unemployment" (Chatham House et al., 2019). The situation with unemployment deteriorated – while the general unemployment level in 2019 was 3,7%, it reached around 4,2% in 2020 and 4,76% in the middle of 2021 (Trading Economics, 2021). These determinants are able to challenge the public trust to the Taiwanese governance institutions, which provokes threats to the democracy as a significant element of the Taiwanese identity.

The cybersecurity landscape of Taiwan, despite the initiatives to resolve these problems, is still vulnerable and facing the threat of cyberattacks – including AI-made ones. According to the report of the Netherlands Enterprise Agency, only 14% of Taiwanese businesses have an advanced level of protection from attacks (ITRI, 2020). This vulnerability is critical in the psychological context due to the positioning Taiwan as an economy based on cutting-edge digital technologies.

In the context of the detected vulnerabilities the prior Taiwanese targets of MUAI in information-psychological field include

1) the reputation among potential regional and global partners;

2) the image of democratic institutions and hence politicians and political processes associated with the efficient governance;

3) the identity of the Taiwanese defined through the history, democratic regime, language, etc.;

4) the image of cybersecurity systems robustness.

The first three targets are closely interrelated with the last one since guaranteeing cybersecurity on the national level reflects the resilience of international partner and the efficiency of government initiatives and investments of the recent years as well as the ability of the Taiwanese society to self-identify in the modern world.

The relevant cases of MUAI in Taiwan can be referred to the first (Army University Press, 2020) and the second targets (Macdonald-Laurier Institute, 2020). The basic technologies in this context are AI texts generating, bots spreading fake news and technologies modifying social media algorithms – also due to fake news promotion (Chen & Wen, 2019; Sherwell,

2020). Recently the trend of the malicious use of Search Engine Optimization (SEO) tools is also being established (刘, 2021), aiming at critical information infrastructure and thus challenging the fourth target. This trend is certain to continue in the near future owing to the dynamics of partnership-building in the region as well as the character of local democracy against the global democratic decline. Nonetheless, these targets choice can provoke an active resistance in technologically advanced societies of Taiwan and Japan with the following criticism of the malicious AI user as an irresponsible political actor. Thus, the shift to the third target and more active focus on the fourth one is possible in the near prospect in case of failing the current tactics.

References

Army University Press. (2020). How to Counter China's Disinformation Campaign in Taiwan. Army University Press. Retrieved 3 November 2021, from https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/September-October-2020/Zhang-Disinformation-Campaign/

Carnegie Endowment for International Peace, & Feigenbaum, A. E. (2020). Assuring Taiwan's Innovation Future. Retrieved 3 November 2021, from https://carnegieendowment.org/files/2020-Feigenbaum-Taiwan_Innovation.pdf

Carpenter, T. G. (2020). A Reborn U.S.-Taiwan Military Alliance? Retrieved 3 November 2021, from https://www.cato.org/commentary/reborn-us-taiwan-military-alliance

Chatham House, Brown, K., & Sageman, C. (2019). Taiwan's Geopolitical Challenges and Domestic Choices. Retrieved 3 November 2021, from https://www.chathamhouse.org/sites/default/files/publications/research/2019-04-12-Taiwan%27s%20Geopolitical%20Challenges.pdf

Chen, Y.-N. K., Wen, C.-H. R. (2019). Facebook's Algorithms, Fake News, and Taiwan's 2018 Local Elections, 30th European Conference of the International Telecommunications Society (ITS): "Towards a Connected and Automated Society", Helsinki, Finland, 16th-19th June, 2019, International Telecommunications Society (ITS), Calgary

Chull Shin, D. (2020). Democratic deconsolidation in East Asia: exploring system realignments in Japan, Korea, and Taiwan. Democratization, 28(1), 142–160. https://doi.org/10.1080/13510347.2020.1826438

Cyber Security Management Act - Article Content - Laws & Regulations Database of The Republic of China. (2018, June 6). Laws & Regulations Database of The Republic of China. Retrieved 3 November 2021, from https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=A0030297

Gould, J. (2021). Compromise defense bill confronts a rising China. Retrieved 3 November 2021, from https://www.defensenews.com/congress/2021/03/02/eyeing-china-indo-pacific-command-seeks-27-billion-deterrence-fund/

ITRI. (2020). Research of Cyber Security Industry in Taiwan (RVO-109-2020/RP-INT). Retrieved 3 November 2021, from https://www.rvo.nl/sites/default/files/2020/07/Research-of-Cyber-Security-in-Taiwan.pdf

Lee, W. (2018). Taiwan's Political Re-Alignment and Diplomatic Challenges (Politics and Development of Contemporary China) (Softcover reprint of the original 1st ed. 2019 ed.). Palgrave Macmillan.

Macdonald-Laurier Institute. (2020). Chinese Disinformation in Taiwan: J. Michael Cole in the Taiwan Sentinel. Retrieved 3 November 2021, from https://www.macdonaldlaurier.ca/chinese-disinformation-taiwan-j-michael-cole-taiwan-sentinel/

National Cyber Security Program of Taiwan (2021 to 2024). TWNCert. Retrieved 3 November 2021, from https://www.twncert.org.tw/Download/National%20Cyber%20Security%20Program%20of%20Taiwan%20(2021–2024).pdf

National Information and Communication Security Taskforce, Executive Yuan, Republic of China (Taiwan). (2021).

Reuters. (2021). Taiwan, Japan ruling parties discuss China, military cooperation. Retrieved 3 November 2021, from https://www.reuters.com/world/asia-pacific/taiwan-japan-ruling-parties-discuss-china-military-cooperation-2021-08-27/

Sherwell, P. T. (2020). Artificial intelligence: China 'uses Taiwan for target practice' as it perfects cyber-warfare techniques. World. Retrieved 3 November 2021, from https://www.thetimes.co.uk/article/china-uses-taiwan-for-target-practice-as-it-perfects-ai-cyber-warfare-to-attack-the-west-sdn9qm8jt

The Diplomat. (2021). Prospects and Challenges for Japan-Taiwan Security Cooperation. Retrieved 3 November 2021, from https://thediplomat.com/2021/05/prospects-and-challenges-for-japan-taiwan-security-cooperation/

Trading Economics. (2021). Taiwan Unemployment Rate. 1978–2021 Data. 2022–2023 Forecast. Retrieved 3 November 2021, from https://tradingeconomics.com/taiwan/unemployment-rate

Ya-Chun, C. (2021). Taiwan's president announces new cybersecurity unit. Retrieved 3 November 2021, from https://www.taiwannews.com.tw/en/news/4194694

专家：美对中国的信息欺骗行动正火力全开|美国|美国国务院|蓬佩奥_新浪军事_新浪网. (2020). Sina Finance. Retrieved 3 November 2021, from https://mil.news.sina.com.cn/china/2020-08-05/doc-iivhvpwx9256837.shtml

刘. (2021). 熟骇客的共军将领出列 学者：慎资讯作战(图) - 刘世民 - 在地见闻. 看中国. Retrieved 3 November 2021, from https://www.secretchina.com/news/gb/2021/07/07/977334.html

率. (2021). 专访琼斯：非常规战争已打响，美国如何应对中俄伊三大"危险分子." 美国之音. Retrieved 3 November 2021, from https://www.voachinese.com/a/interview-with-Seth-Jones-how-to-deal-with-Russia-China-Iran-in-irregular-wars-20210908/6215799.html

Технология дипфейк как вызов информационно-психологической безопасности КНР и Северо-Восточной Азии[1]

Екатерина Михалевич

С форсированным развитием в Северо-Восточной Азии таких технологий, как интеллектуальные компьютерные программы, ИКТ, ИИ и т.д., расширяются перспективы их применения в экономике и промышленности. Вместе с тем, увеличивается число вызовов и угроз ИПБ (Базаркина, Дам, Пашенцев, Фан и Матяшова, 2021) как для отдельных государств и их граждан, так и для всего региона в целом. Такая технология на основе ИИ как дипфейк, применяемая антисоциальными акторами, способна оказать значительное влияние на формирование общественного дискурса путем создания ложных новостей и инфоповодов. «Фейк ньюс», зловредные информационные «вбросы», кража персональных данных интернет-пользователей способны подорвать общественное доверие, стать поводом для шантажа государственных чиновников и т.д. В общем, современные технологии являются весьма эффективным информационным механизмом дестабилизации как социально-политической ситуации внутри отдельно взятого государства, так и международных отношений (Bazarkina and Pashentsev, 2019).

Рассматриваемый регион представлен государствами, которые динамично развивают сферу ИКТ и соперничают за превосходство на данном направлении, а, следовательно, потенциально обладают возможностями использования информационных технологий с целью дестабилизации ИПБ в странах-соперниках.

Сама по себе технология дипфейков не представляет собой вызов ИПБ. Изначально подобный алгоритм разрабатывался в развлекательных целях (предшественником технологии можно назвать спецэффекты, активно используемые в киноиндустрии с 80-х – 90-х гг. прошлого века). Сам термин «дипфейк» появился после попадания технологии в руки антисоциальных субъектов. Данная технология привлекает всё большее внимание специалистов по информационной безопасности и международным отношениям, так как ее использование в злонамеренных целях – чаще всего в финансовых махинациях, шантаже политических деятелей, создании фальшивых новостей – представляет собой реальную угрозу международной ИПБ.

Исследование, посвященное влиянию дипфейков на доверие общества к СМИ, показало, что, с одной стороны, со временем люди на собственном опыте учатся отличать фейковые новости и изображения от реальных, с другой – знание, что дипфейки могут проникнуть даже в новостные сводки государственных агентств, дает обратный эффект: уровень доверия общества СМИ с течением времени неумолимо снижается (Vaccari and Chadwick, 2020).

Важнейшим социально-психологическим воздействием дипфейков является изменение человеческого восприятия, чаще всего неосознанное для самой личности, на

[1] Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 21-514-92001.

которую это воздействие направлено. Одно недавнее исследование показало, что воздействие фейковых видео с изображениями политических деятелей значительно ухудшило отношение электората к данным личностям (Dobber et al, 2019). Проблема потери доверия к политическим лидерам становится более актуальной, если учесть, что фейк-контент распространяется преимущественно (и зачастую неосознанно) через социальные сети обычными пользователями – поиск первоисточника ложной информации становится практически невыполнимой задачей, и при этом фейк может трактоваться людьми по-разному, обрастая новыми подробностями, вводя в заблуждение все более широкие круги пользователей. Все это приводит к дефолту доверия и нигилизму (Levine, 2019) – ситуации, когда общество настолько привыкло к постоянному обману, что старается фильтровать всю получаемую информацию и не доверять даже официальным источникам.

Китайское приложение Zao, созданное в 2019 г. и в рекордные сроки оказавшееся в топе скачиваний интернет-пользователями, вызвало неконтролируемую волну создания дипфейков, начиная от, казалось бы, безобидного контента с лицами пользователей, подставленных в видеофайлы вместо киногероев, и заканчивая созданием, по сути, компромата на первых лиц КНР (ZAO火不火, 2019). Одно из свойств такой ситуации, которое во многих исследованиях остается без должного внимания, – использование фото- или видеоизображения личности без ее согласия. Учитывая возможности визуального изменения контента, а также особенности восприятия человеком подобного материала, очевидно, что дипфейк может оказать пагубное влияние на карьеру и личную жизнь жертвы (Hancock, 2021).

Еще одной скрытой опасностью, которую несет в себе Zao, является утечка конфиденциальных данных пользователей. Исходя из условий пользовательского соглашения, разработчикам программы переходило бессрочное право на безвозмездное пользование контентом, созданным пользователями, а также на фото- и видеоматериал, загружаемый пользователями в программу с целью последующей обработки. Кроме того, согласно условиям пользовательских соглашений, разработчики Zao имели право передавать разрешение на пользование контентом третьей стороне без уведомления и дополнительного разрешения пользователя (唏嘘！曾经爆红的ZAO APP, 2021). Данные пункты прямо нарушают законодательство КНР в области защиты персональных данных (个人信息保护法草案, 2021).

После скандала в китайской прессе, разразившегося по поводу возможной утечки персональных данных и их использования в злонамеренных целях (ZAO出事, 2021), компания Momo, осуществившая релиз Zao, изъяла пункты, противоречащие законодательству КНР, из пользовательского соглашения. Тем не менее, популярная социальная сеть (а по совместительству и платежная система) Weixin (WeChat), распространенная не только в КНР, но и в Южной Корее, Японии, России и ряде других стран, начала удаление фейкового видеоконтента со своей платформы для предупреждения возможных рисков безопасности. Крупнейшая в мире платежная система Alipay, пользователями которой являются все страны Северо-Восточной Азии, за исключением Северной Кореи (всего более 900 млн. человек), также не могла не отреагировать на скандал. Представителям Alipay пришлось заверить пользователей в

безопасности проводимых платежей через смартфон, для чего достаточно лишь посмотреть в камеру. По заявлению Alipay, система использует сложные алгоритмы на основе ИИ, которые распознают фейковые фото и видео и блокируют проведение платежа (ZAO出事, 2021). Реакция Alipay является примером упреждающей кризисной коммуникации, цель которой – восстановление пошатнувшейся репутации.

Кроме пагубного влияния на репутацию самих китайских IT-гигантов (Alibaba, WeChat, Baidu), активное и бесконтрольное развитие дипфейков в КНР отрицательно сказывается на доверии международных компаний к стране. Уровень доверия западных инвесторов (как к китайским компаниям, так и к правительству КНР – поскольку власти пока не в силах пресечь ЗИИИ) в последнее десятилетие значительно снизился (Nash, 2021). Это приводит к оттоку иностранных инвестиций из Китая и сокращению числа пользователей китайских программ, как следствие – снижению влияния Китая на цифровую экономику стран-потребителей его услуг (Ohara, 2021).

Эксперты приходят ко мнению, что злонамеренное использование дипфейков в ходе информационных атак представляет серьезную угрозу всему международному сообществу (Пашенцев, Фан и Дам, 2020). IT-аналитики считают, что дипфейк может оказаться самой опасной технологией ИИ за последние десятилетия. В области политики дипфейки могут быть использованы в качестве оружия против отдельных деятелей, партий, правительств с целью манипулирования общественным восприятием и мнением. Хани Фарид в качестве одного из самых ярких примеров применения технологии дипфейк в злонамеренных целях называет вирусный видеоролик, героем которого стал Дональд Трамп, заявляющий, что применил ядерное оружие против Северной Кореи. «Нарезанные» звуковые дорожки из публичных речей бывшего президента США, наложенные на видеофрагмент одного из выступлений, вызвали кратковременную истерию среди граждан стран Северо-Восточной Азии, пока данная новость не была официально развенчана новостными агентствами (Experts fear face swapping tech, 2018).

Директор центра медиа-криминалистики при Университете Колорадо в США Каталин Григорас заявил, что эксперты в области криминалистики обладают достаточными знаниями и технологиями, чтобы за короткий промежуток времени отличить поддельный высокотехнологичный контент от реального изображения. В то же время, нужно учитывать тот факт, что технологии ИИ также не стоят на месте. Новые продукты ИИ обладают потенциалом синтеза новых высказываний, которые человек никогда не произносил, на основе лишь нескольких звуковых дорожек с голосом (Experts fear face swapping tech, 2018). В настоящее время, по мнению профессиональных криминалистов, выявить звуковые подделки намного сложнее, чем графические.

Чтобы снизить информационно-психологическое воздействие технологии дипфейк на общество, необходимо выработать ряд рекомендаций. Самым первым шагом в этом направлении должно стать просвещение граждан: необходимо внедрять в процесс обучения в учебных заведениях занятия по цифровой грамотности и цифровому этикету, на государственном уровне – проводить пропаганду ответственного поведения в Интернете, повышать осведомлённость общества о дезинформации и распространении

дипфейков. Кроме образовательно-просветительской деятельности, власти должны задуматься об ужесточении законодательства в области хранения и использования персональных данных. Кроме того, сфера ИИ, не смотря на активное развитие технологий, практически не регулируется законодательно. Вопрос, кто должен нести ответственность за факт злонамеренного использования ИИ, остается открытым. Тем не менее, в ряде регионов КНР в качестве эксперимента вводятся в действие локальные нормативно-правовые акты, регулирующие функционирование ИИ с точки зрения этики и законности. Так, в Шэньчжэне в июле 2021 г. был принят свод правил использования и развития технологий на основе ИИ, предусматривающий реальные санкции за их нарушения (Koty, 2021). Шэньчжэнь считается китайским лидером в области ИИ, следовательно, его пример, в случае положительных результатов апробации, может оказать влияние на установление единого стандарта на территории всей страны.

В заключение следует отметить, что развитие новых технологий на основе ИИ, таких как дипфейк, и возможность их применения в злонамеренных целях диктуют необходимость более внимательного изучения потенциала самих технологий и путей нейтрализации новых угроз. Налицо тенденция, когда технологии ИИ, создаваемые в целях ускорения технологического прогресса, становятся оружием против общества. Экспертам стоит выработать инструменты и протоколы предупреждения информационных вбросов и научиться эффективно и своевременно реагировать на них.

Библиография

Базаркина Д.Ю., Дам В.Н., Пашенцев Е.Н., Фан К.Н.А., Матяшова Д.О. (2021). Политическая ситуация в Северо-Восточной Азии и угрозы злонамеренного использования искусственного интеллекта: вызовы информационно-психологической безопасности. Социально-гуманитарные знания. Вып. 4. С. 212-234. DOI 10.34823/SGZ.2021.4.51655.

Пашенцев Е.Н., Фан К.Н.А., Дам В.Н. (2020). Злонамеренное использование искусственного интеллекта в Северо-Восточной Азии и угрозы международной информационно-психологической безопасности. Retrieved 30 September 2021, from: https://cyberleninka.ru/article/n/zlonamerennoe-ispolzovanie-iskusstvennogo-intellekta-v-severo-vostochnoy-azii-i-ugrozy-mezhdunarodnoy-informatsionno/viewer

Bazarkina D. & Pashentsev E. (2019). Artificial Intelligence and New Threats to International Psychological Security. Russia in Global Affairs. Vol. 17. № 1. P. 147-170. DOI: https://doi.org/10.31278/1810-6374-2019-17-1-147-170.

Dobber T. et al. (2019). Do (microtargeted) deepfakes have real effects on political attitudes? The International Journal of Press/Politics. № 26. P. 69–91.

Experts fear face swapping tech could start an international showdown. (2018). Retrieved 30 September 2021, from: https://theoutline.com/post/3179/deepfake-videos-are-freaking-experts-out

Hancock J.T. The Social Impact of Deepfakes. (2021). Retrieved 30 September 2021, from https://stanfordvr.com/mm/2021/03/cyber_deepfakes_2021.pdf

Koty A.C. (2021). Artificial Intelligence in China: Shenzhen Releases First Local Regulations. Retrieved 30 September 2021, from: https://www.china-briefing.com/news/artificial-intelligence-china-shenzhen-first-local-ai-regulations-key-areas-coverage/

Levine T.R. (2019). Duped: truth-default theory and the social science of lying and deception. Tuscaloosa, AL: University Alabama Press.

Nash J. (2021). AI researchers won't be playing 'trust fall' with Facebook. Or China. Retrieved 30 September 2021, from: https://www.biometricupdate.com/202108/ai-researchers-wont-be-playing-trust-fall-with-facebook-or-china

Ohara B. (2021). How China is undermining digital trust – and how democracies can respond. Retrieved 30 September 2021, from: https://ecfr.eu/article/how-china-is-undermining-digital-trust-and-how-democracies-can-respond/

Vaccari C. & Chadwick A. (2020). Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media and Society. № 6. P. 1–13.

ZAO出事，陌陌露底 (ZAO has an accident, Momo reveals the bottom) (2019). Retrieved 30 September 2021, from: https://baijiahao.baidu.com/s?id=1644344007383017858&wfr=spider&for=pc (In Chinese).

ZAO火不火，取决于动机纯不纯 (*ZAO: danger or not, depends on the pure impure motivation*) (2019). Retrieved 30 September 2021, from: https://www.thepaper.cn/newsDetail_forward_4334350 (In Chinese).

个人信息保护法草案 (Personal Information Protection Law) (2021). Retrieved 30 September 2021, from: https://m.thepaper.cn/baijiahao_12598997 (In Chinese).

唏嘘！曾经爆红的ZAO APP下架被约谈 (*The once-popular ZAO APP was removed from the shelves and was interviewed*) (2021). Retrieved 30 September 2021, from https://m.mydrivers.com/newsview/644907.html (In Chinese).

Malicious Use of Artificial Intelligence – New Psychological Security Risks: An Indian Perspective

Kallakuri Radhakrishna

"We are living through the largest unregulated social experiment of all time – a generation of youth has been exposed to extreme content online."
Michael Seto (Aiken, 2017, pp. 16–17)

While analyzing the threats to psychological security caused by MUAI both at a national and international level, these threats can be considered at three levels. The first level is directed towards generating a feeling of negative attitudes to the development of AI where the aim is to slow down the progress of its development, which can also consist of creating a malicious image of institutions and professionals involved in the development of AI technology. The second level of threats directly involves the MUAI, where the main goal is not an attack on public consciousness but the goals are more direct and kinetic like using of unmanned aerial vehicles by terrorists, using AI to increase the effectiveness of cyber-attacks leading to direct threats to cyber infrastructure. Crypto currency manipulations by supranational organizations may lead to artificially induced changes to crypto currency markets (Pashentsev and Bazarkina, 2020).

Threats emanating from MUAI designed primarily to cause psychological damage is the third and highest level of PS threats. Using tactics and strategies from psychological warfare may lead to significant amplification of perceived threats leading to public responses becoming spontaneous and irrational. Using of AI based technologies in psychological warfare makes covert perception management campaigns more dangerous.

Serious damage to the reputation of people, organizations or even country can be achieved through using AI-based phishing, use of deepfakes and smart bots in psychological warfare campaigns (Pashentsev and Bazarkina, 2020).

With China and Russia, leading actors in the research area of MUAI and PS, India can pursue joint research activities along the BRICS framework where policy documents state readiness for joint build-up of capacity and capability to counter threats from MUAI. The more practical solution and likely scenario taking into consideration current geopolitical challenges faced by India, bilateral agreements with individual countries may play the leading role for development of artificial intelligence cooperation amongst BRICS countries.

Based on current capability analysis done by leading experts in the field of MUAI and international PS, current national regulation pertaining to AI and policies to counter its malicious use are at fledging state in most countries. Due to the complex nature of technology and its dual use characteristics current lack of regulation may lead to significant challenges to the PS of people. This makes it a matter of urgent concern for each of BRICS member states to develop international cooperation and share their experiences.

The most significant threats to public consciousness arise during crises with longer the crisis the more difficult it becomes to defend against threats from aggressive actors. As seen

during the current pandemic, misinformation designed using the time tested methods of classical psychological warfare were amplified using the new available tools provided by AI. Experts have widely documented the use of deepfakes and smart bots by various state sponsored actors to highlight their country's response in a positive manner while showing the others in a negative light.

Targeted psychological warfare campaigns designed to destabilize the decision making process were always present, it is the new dimension provided by AI which makes these campaigns very swift and decisive. These AI-based psychological warfare tools provide significant advantages to those who decide to employ them as both speed and efficiency are delivered at a very low cost. The current coronavirus pandemic has led to a major downturn in global economic output, which were last seen during the great depression. Targeted information warfare campaigns in such an atmosphere leads to amplification of risks which could bring current socio-political tensions under even more strain. When seen from the perspective of AI-based emerging technologies the current coronavirus pandemic has necessitated the transition to distant forms of working undoubtedly leading to many advantages. However, what remains poorly studied is the new opportunities it provides to actors with malicious intentions.

One such recent attack which could have been more dangerous had it been integrated with AI-based technologies is that of fake letter circulating on social media allegedly written by national security advisor Ajit Doval (NDTV, 2021). The fake letter crucially timed during the holy *Kumbh Mela*[1] celebrations appreciates the performance of Uttarkhand chief secretary in successfully organizing the festival at Haridwar. The fake letter compliments his performance and ends by saying that his efforts have helped maintain order and promote *Rashtriya swayamsevak sangh*'s ideology[2].

The MUAI in this case may help provide aggressive actors to portray government institutions as torchbearers of a certain political ideology trying to guarantee the interests of one community, which when employed on a mass level may lead to social tensions amongst various communities. Usage of deepfakes and other such applications in my opinion provides an array of new opportunities that will add to India's asymmetric warfare capabilities, but to protect the larger public consciousness increase in regulation is very much necessary. Despite significant threats posed by malicious use of AI-based applications such as deepfakes, there are wide range of positive applications such as creation of false targets for criminals and better identifying potential victims (Pashentsev and Bazarkina, 2020).

Conclusion

As Great War's in the current world will lead to end of humanity, the focus has shifted to hybrid warfare. The integration of MUAI with traditional psychological warfare will be very

[1] Kumbh Mela is a major pilgrimage and festival in Hinduism which attracts millions of people.

[2] Rashtriya Swayamsevak Sangh ("National Volunteer Organisation"), abbreviated as RSS, is an Indian right-wing nationalist organization. The ideological focus of RSS has been to strengthen the core ancient Indian values, which celebrates unity in diversity. To achieve this RSS has continued to strengthen and build relationships with their allies who all face a common enemy.

decisive to win these battles. Indian security apparatus should continue to build their capacity and capability in this regard.

References

Aiken, M. (2017). *The Cyber Effect: A Pioneering Cyberpsychologist Explains How Human Behavior Changes Online*. London: John Murray Publishers.

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

NDTV. (2021). Letter From Ajit Doval Praising Kumbh Efforts "Fake", Say Officials: Report. (2021). Retrieved 13 May 2021, from https://www.ndtv.com/india-news/nsa-doval-has-not-written-any-letter-lauding-kumbh-mela-conduct-say-govt-officials-2418291

Pantserev, K. (2020). The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability. In H. Jahankhani, S. Kendzierskyj, N. Chelvachandran & J. Ibarra (eds.), *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity* (pp. 37-55). Cham: Springer.

Pashentsev, E. (2020). Malicious Use of Deep fakes and Political Stability. In F. Matos, *Proceedings of the 2nd European Conference on the Impact of AI and Robotics, a virtual conference hosted by Instituto Universitario de Lisboa (ISCTE-IUL), Portugal, 22-23 October 2020* (pp. 100–107). Reading, UK: Academic Conferences and Publishing International Limited.

Страх перед искусственным интеллектом как фактор информационно-психологической дестабилизации

Сергей Себекин

Появление любой дизруптивной технологии[1], как правило, влечёт за собой существенные изменения в жизни общества: кардинально меняются системы производства и управления, экономика, социальная жизнь и политика, а иногда и характер межгосударственного стратегического соперничества. Ярким историческим примером появления дизруптивной технологии стало появление ядерного оружия. Именно то, что ядерным оружием обладали сначала два государства – СССР и США – определило расстановку сил и биполярность системы международных отношений после Второй мировой войны. Две страны получили мощные рычаги влияния на мировую политику. Но, что более важно, в результате была сформирована система взаимного *сдерживания*, которая основывалась на *психологическом страхе* взаимно-гарантированного уничтожения (Kennan, 1947; Rothschild, 1995, p. 58). Сдерживание действовало (и действует) в сознании стратегических соперников (Nye, 2017, p. 53).

Предполагается, что дизруптивной технологией является и ИИ. Президент России В. В. Путин уже неоднократно заявлял о том, что страна, добившаяся лидерства в создании ИИ, «будет властелином мира» (ТАСС, 2019). Вероятно, как и в случае с ядерным оружием, обладание ИИ способно вызвать естественное чувство страха или опасения у того, кто отстает в развитии данной технологии или вовсе не сможет обладать ей. Таким образом, уже сегодня, на стадии активной разработки, ИИ способен оказать негативное влияние на сложившуюся архитектуру стратегической стабильности в силу самих опасений, что он сможет существенно изменить расстановку сил на международной арене. Сам страх перед наличием у кого-либо передовых систем ИИ способен дестабилизировать стратегическую стабильность даже вне условий его реального деструктивного применения.

Предметом данного доклада будут не конкретные проявления ЗИИИ против международной ИПБ, а то, к каким последствиям для этой ИПБ, и, как следствие, для современного миропорядка и сложившейся системы стратегической стабильности, может привести именно *страх* от обладания и потенциального злонамеренного применения ИИ в военно-политических целях.

Основные страхи и тревоги в отношении ЗИИИ связаны с тем, что одно или группа государств в определённый момент первыми получат доступ к передовым системам ИИ, либо превосходящим аналоги в других странах, либо вовсе уникальны в своей сфере. Другие страны станут аутсайдерами, что сегодня вызывает основную тревогу в силу ряда факторов. Страна с самым передовым ИИ получит существенное

[1] Дизруптивные технологии (от англ. «disruptive» – подрывной) – технологии, которые приходят на смену «классическим» технологиям и влекут за собой коренные изменения ситуации в политике, экономике, международных отношениях и т.д.

технологическое преимущество, а, следовательно, способность значительно изменить баланс сил – как в обороне, так и в наступлении – что не может не вызывать психологической тревоги у других. Что более важно, как показывает история, страны, первыми получившие доступ к новой дизруптивной технологии и закрепившие «монополию» на обладание ею, задают «правила игры» в самом использовании новой технологии. Первопроходец получает право на первоочередное нормотворчество в этой сфере. После Второй мировой войны США и СССР, первыми создав ядерное оружие, сформировали биполярный мир. Большинство ключевых соглашений, составляющих основу стратегической стабильности, были подписаны ими же.

Соответственно, те страны, которые не получат доступ к технологии ИИ или будут обладать более слабыми ИИ-системами, рискуют превратиться из субъекта в объект мировой политики или же вовсе утратить статус великой державы (если такой у них имелся) и лишиться рычагов глобального влияния. Здесь снова уместен пример с ядерным оружием: в Совет безопасности ООН в качестве постоянных членов, обладающих правом вето, вошли именно страны «ядерной пятёрки». Соответственно, те, кто не обладал, не вошли. По этой причине такие страны, как США, Китай и Россия, сегодня стремятся разработать продвинутые ИИ-системы, чтобы не просто стать лидерами отрасли, но и создать себе определённые механизмы влияния на мировую политику. Это не может не вызывать тревоги у других стран: аутсайдеры будут вынуждены пассивно принимать правила игры, написанные инсайдерами.

В условиях растущей международной напряженности лидерство определённого государства или государств само по себе может оказывать негативное психологическое влияние на национальном и глобально-стратегическом уровнях. Какие последствия может иметь статус аутсайдера для страны в сфере информационно-психологической стабильности на национальном уровне? Такое положение может сложиться в сегодняшних сверхдержавах, если ИИ обеспечит их соперникам качественное преимущество в экономике, военной и других сферах.

При сохранении международной напряженности нельзя исключить, что те из соперничающих государств, кто первым получит передовые технологии ИИ, попытаются (в том числе через спецслужбы) использовать его в злонамеренных деструктивно-психологических целях. В будущем вероятно проведение широкомасштабных деструктивно-психологических кампаний против граждан страны-соперника – как отдельных групп (молодежь, пенсионеры, представители определённых профессий, жители отдельного региона и т. д.), так и против всего населения с самыми разными целями, вплоть до свержения правительств. Государство-соперник может использовать недовольство населения по определённым вопросам внутренней политики, чтобы спровоцировать волнения. Например, в России жители Сибири обеспокоены вырубкой лесов. Как на них повлияет дипфейк, на котором глава одного из регионов оглашает указ, позволяющий вырубать леса в ещё больших масштабах, или изображения, сгенерированные ИИ, на которых якобы зияют пустоши там, де раньше был лесной массив?

Один из футуристичных сценариев – гибридная атака, при которой определённая часть населения помещается в информационный вакуум, теряя доступ к привычным новостным ресурсам, чтобы затем транслировать информационные потоки,

сгенерированные посредством использования той же дипфейк-технологии. Уже сегодня боты на основе ИИ продвигают в социальных сетях повестку дня и мнения, выгодные своим разработчикам. Деморализация может происходить и посредством вывода из строя или взятия под контроль ИИ, управляющего физическими объектами. Невозможность для страны противостоять этому не прибавит оптимизма гражданам.

Другой уровень – глобально-стратегический. Стратегическая стабильность во многом зависит именно от степени психологической тревоги по поводу возможности применения какой-либо технологии в злонамеренных целях, и, соответственно, включает в себя поведенческие факторы (поведение субъектов политики в определённой ситуации). Тотальное недоверие к стратегическому сопернику и постоянные опасения по поводу наличия у него передового ИИ могут привести к массовому психозу, как это было в США во времена «красной истерии». «Раздражённый» субъект, как правило, склонен более импульсивно реагировать на угрозы безопасности, как реальные, так и мнимые. Более того, подобное недоверие приводит к информационной изоляции, нежеланию обмениваться информацией в области ИИ между странами. Такая тревожность чревата новой «гонкой вооружений» в сфере ИИ, а одержимость в стремлении к стратегической инициативе может подорвать стратегическую стабильность.

Любая дизруптивная технология – значимый фактор в системе стратегической стабильности. ИИ способен стать ещё одним инструментом сдерживания (Geist and Lohn, 2018), подорвав существующую систему стратегической стабильности, гарантом которой является ядерное сдерживание. Опасения одних, связанные с наличием ИИ у других, повышают риск принятия военно-политическим руководством той или иной страны необдуманных решений, что может привести к серьезному кризису.

Библиография

Kennan, G. (1947). The Sources of Soviet Conduct. *Foreign Affairs*, *25*(4), 566. doi: 10.2307/20030065

Rothschild, E. (1995). What Is Security?. *Daedalus*, *124*(3), 53–98. doi: http://www.jstor.org/stable/20027310

Nye, J. (2017). Deterrence and Dissuasion in Cyberspace. *International Security*, *41*(3), 44-71. doi: 10.1162/isec_a_00266

ТАСС. (2019). Путин: монополист в сфере искусственного интеллекта может стать властелином мира. Retrieved 3 November 2021, from https://tass.ru/ekonomika/6489864

Geist, E., & Lohn, A. (2018). How Might Artificial Intelligence Affect the Risk of Nuclear War?. Retrieved 3 November 2021, from https://www.rand.org/pubs/perspectives/PE296.html

Злонамеренное использование искусственного интеллекта во Вьетнаме в информационно-психологическом аспекте

Нгуен Минь Тыонг

В современном мире ИИ присутствует практически во всех областях жизни человека и является ключевым элементом промышленной революции. ИИ во Вьетнаме получил развитие в ряде экономических и социальных областей. В 2020 г. Вьетнам занял 76-е место из 172 стран по индексу развития ИИ. В начале 2021 г. правительство опубликовало Национальную стратегию исследований, разработки и применения ИИ до 2030 г. (Пашенцев, Фан и Дам, 2020).

Вьетнам занимает важное геостратегическое положение в современном мире. Располагаясь на важнейших транспортных маршрутах, морских путях, Вьетнам является одной из самых экономически активных стран в Азии, страной с большим потенциалом в одном из наиболее динамично развивающихся регионов в мире. В последнее время с быстрым подъемом Китая, увеличением присутствия и роли США политическая ситуация и ландшафт безопасности в Южно-Китайском море стали всё более сложными. Присутствует конкуренция между Вьетнамом, Китаем и другими странами вокруг спорных островов Южно-Китайского моря (Hoa Nguyễn, 2020).

В силу устремлений различных антисоциальных акторов, ИИ может служить все более эффективным инструментом злонамеренного использования, в том числе оказывать деструктивное воздействие на информационно-психологическую безопасность ИПБ путем целевого воздействия на сознание граждан. Среди возможных угроз ЗИИИ для ИПБ во Вьетнаме можно назвать следующие (Bộ KH&CN, 2021):

Злонамеренное использование глубоких фейков (дипфейков). Приложения Avatartify, Reface, основанные на глубоком обучении, широко используются во Вьетнаме для создания фальшивых видеороликов с «участием» реальных людей. Общей чертой этих приложений является использование техники дипфейков для совмещения лица одного человека с выражением лица и голосом другого. Во Вьетнаме многие люди по-прежнему небрежно делятся личными фотографиями с лицами крупным планом и в хорошем разрешении в социальных сетях, размещая их в открытом доступе. Происходит это из-за распространенного мнения, что злонамеренное использование дипфейков – нечто далекое, и не затрагивающее тебя лично, потому что ты не знаменитость (Tạp chí điện tử Thông tin và Truyền thông, 2020). Но такая опрометчивость может обернуться разнообразными угрозами для общества. В настоящее время злонамеренное использование дипфейков в основном осуществляется для создания порнографического контента, дискредитации, оскорблений, преследования и шантажа.

Злонамеренное использование ботов. Ботов для социальных сетей можно использовать в различных безобидных или вредоносных целях. В целом боты, как правило, специализируются на одной из трех задач:

а) Увеличение количества подписчиков. Самые простые аккаунты ботов следят за другими пользователями, чтобы повысить их популярность. Эти боты коммерчески доступны, хотя покупка подписчиков является нарушением условий обслуживания большинства социальных сетей. Многие общественные деятели имеют большие базы последователей-ботов в социальных сетях, но часто не знают, кто контролирует этих ботов.

б) Распространение информации и дезинформации. Боты способны распространять информацию и создавать ложный консенсус. Хотя все общество пытается справиться с пандемией COVID-19, отдельные индивиды и группы демонстрируют безразличие к боли других людей, распространяя дезинформацию ради личной выгоды. Те, кто распространяет фейковые новости, причиняющие вред и вызывающие панику в обществе, не только нарушают закон, но и имеют моральные недостатки.

в) Взаимодействие с другими пользователями. Некоторые боты могут разумно реагировать и влиять на поведение человека. Отдельное исследование показало, что боты могли выявлять пользователей социальных сетей, недовольных коррупцией в правительстве, и привлекать их к определенным действиям (Phú Lữ, 2020). Боты для социальных сетей могут принести пользу общественности, но их также можно использовать в злонамеренных политических целях. Если предприятия малого и среднего бизнеса могут поднять популярность знаменитостей, то они могут укрепить авторитет диктаторов; они могут распространять истину, но также и ложь; могут продавать продукты, но также и идеологию.

ЗИИИ в компьютерных играх. С помощью фреймворка General Video Game AI теперь игроки получают доступ к широкому и, в принципе, неограниченному диапазону игр с применением ИИ. Конкретно, фреймворк решает проблему разработки алгоритма, который позволяет играть в любую игру, даже если игра неизвестна априори. Эту область исследований можно рассматривать как приближение к общему ИИ, в котором очень мало места для эвристики, зависящей от игры.

Игры уже давно являются популярной средой для применения ИИ. Многие исследователи изучали алгоритмы и методы для того, чтобы оптимизировать компьютерные игры, такие как шахматы, автомобильные гонки, «Мисс Пакман», стратегии в реальном времени (RTS) и “Super Mario Bros”.

Вклад, внесенный исследованиями компьютерных игр, крайне важен для понимания проблем ИПБ важен. Сам характер соревнований в конкретных играх, их неотъемлемая структура может создавать определенные информационно-психологические риски: большинство предлагаемых решений имеют узкую область применения. Другими словами, характер вызова, бросаемого игроку, или самой борьбы за победу в соревнованиях побуждает участников действовать в соответствии с алгоритмом с тщательно продуманной эвристикой, которая будет работать только в игре. Так разработчики игр могут повысить эффективность информационно-психологического воздействия, особенно на детские и подростковые аудитории. В настоящее время игровая зависимость, особенно от онлайн-игр среди молодежи, неуклонно растет, что приводит ко многим негативным последствиям для жизни

человека. Помимо того, что онлайн-игры отнимают у людей время и энергию, не всегда компенсируя их полезными навыками работы с компьютером или эмоциональной разгрузкой, игры также напрямую влияют на изменение эмоций и мыслей, на другие аспекты социальной жизни (это влияние может быть и крайне негативным) (Long, 2020). Власти Вьетнама давно поднимали проблему зависимости подростков от компьютерных игр (Ngọc Anh, 2010).

Контроль над развитием технологий ИИ – это дело всего общества (от государства до частных компаний). Людям необходимо повышать свои знания об ИИ. Также необходимо совместно развивать законодательную базу и технические нормативы для дальнейшего использования и развития ИИ, вместе с тем повышая способность общества противостоять конкретным угрозам ИПБ.

Библиография

Пашенцев, Е. Н., Фан, К. Н. А., Дам, В. Н. (2020). Злонамеренное использование искусственного интеллекта в Северо-Восточной Азии и угрозы международной информационно-психологической безопасности. *Государственное Управление. Электронный Вестник*, *80*, 178-199. doi: 10.24411/2070-1381-2020-10069

Bộ KH&CN. (2021). Toàn cảnh Trí tuệ Nhân tạo Việt Nam: Hiện trạng và thách thức. Retrieved 3 November 2021, from https://congnghiepcongnghecao.com.vn/tin-tuc/t24134/toan-canh-tri-tue-nhan-tao-viet-nam-hien-trang-va-thach-thuc.html

Hoa Nguyễn. (2020). Vị thế Việt Nam trong cục diện mới của khu vực. Retrieved 3 November 2021, from https://www.tapchicongsan.org.vn/web/guest/tin-binh-luan/-/asset_publisher/DLIYi5AJyFzY/content/vi-the-viet-nam-trong-cuc-dien-moi-cua-khu-vuc

Long, N. (2020). Dấu hiệu nhận biết người nghiện game. Retrieved 3 November 2021, from https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/dau-hieu-nhan-biet-nguoi-nghien-game/

Tạp chí điện tử Thông tin và Truyền thông (Electronic Journal of Information and Communication). (2020). Deepfake – mối nguy hại mới trong thế giới Internet. Retrieved 3 November 2021, from https://ictvietnam.vn/deepfake-moi-nguy-hai-moi-trong-the-gioi-internet-20201116154635751.htm

Phú Lữ. (2020). Nhận diện rõ những thủ đoạn dùng mạng xã hội chống phá Đảng. Retrieved 3 November 2021, from https://cand.com.vn/Chong-dien-bien-hoa-binh/Nhan-dien-ro-nhung-thu-doan-dung-mang-xa-hoi-chong-pha-Dang-i570727/

Ngọc Anh. (2010). Vì sao thanh thiếu niên sa đà vào game?. Retrieved 3 November 2021, from https://cand.com.vn/Xa-hoi/Vi-sao-thanh-thieu-nien-sa-da-vao-game-i163585/

Деятельность НАТО по борьбе со злонамеренным использованием ИИ: тенденции и противоречия

Олег Филатов

Значимость ИИ на сегодняшний день трудно переоценить. В военной сфере его внедрение может значительно способствовать снижению человеческих потерь. К примеру, в военной разведке ИИ уменьшает необходимость участия человека в опасных для жизни заданиях. Применение ИИ при распознавании целей повышает точность действий артиллерии и авиации. С другой стороны, ИИ способен заменить человека во многих сферах, что чревато социальными кризисами глобального масштаба (Catania, Ganga, Milardo, Morabito and Mursia, 2018). Это и многое другое говорит нам о том, что развитие ИИ должно иметь контролируемый характер.

Имея одну из наиболее многочисленных армий, военный блок НАТО активно модернизирует свои вооруженные силы (Willemo, 2019). Характеристика внедрения ИИ в деятельность НАТО подробно представлена в докладе заместителя главы НАТО по вопросу инноваций Эдварда Кристи (2020), который на основе классификации Министерства обороны США подразделяет ИИ на три категории: а) корпоративный ИИ – автоматизированные системы управления финансами и персоналом, применяемые в строго контролируемой среде, в которой негативные последствия возможных технических сбоев минимальны (с точки зрения возникновения угрозы для жизни и потенциальных жертв); б) оперативный ИИ – автоматизированные системы, которые могут быть применены в рамках боевых действий, то есть в значительно менее контролируемой среде (последствия технических сбоев и ошибок могут привести к критическим последствиям и жертвам среди личного состава) например, стационарные комплексы вооружения или беспилотная техника; в) ИИ обеспечения миссий (применяется как в контролируемой, так и в неконтролируемой среде) – системы ИИ в сфере логистики, технического обслуживания, разведки.

Опираясь на подход Э. Кристи, НАТО осуществляет машинное обучение ИИ (Christie, 2020). Согласно ресурсу Стратегического командования НАТО по трансформации, внедрение ИИ в военном блоке происходит в рамках проекта «Межнациональная кампания по развитию потенциала». Инициатива исходит от США и направлена на упрощение взаимодействия между различными группировками войск в ходе совместных операций НАТО. Помимо ИИ в рамках проекта предполагается проведение модернизации вооруженных сил НАТО посредством масштабной автоматизации североатлантического ВПК и активного внедрения робототехники (NATO Allied Command Transformation, 2020). Приоритетными направлениями для внедрения и развития вышеуказанных технологий для НАТО являются разработка и применение электромагнитного оружия, системы ПВО и ПРО, логистика, космическое пространство, киберпространство, военно-воздушные, военно-морские и сухопутные войска.

Несмотря на энтузиазм НАТО в развитии ИИ, данный процесс может затянуться или происходить неравномерно. Одной из наиболее очевидных причин для этого является

существенная разница в развитии вооруженных сил между странами-членами альянса (Thiele, 2020). Несмотря на то, что значительная часть государств НАТО обеспечивают свои армии вооружением американского и собственного производств (например, Великобритания, Франция, Италия, Германия), странами, присоединившимися к альянсу в рамках пятой волны и далее, до сих пор используется вооружение советского производства, сохранившееся со времен пребывания данных государств в ОВД (Болгария, Румыния, Венгрия, Чехия, Польша). Переход на вооружение стандарта НАТО сопряжен с большими финансовыми затратами для новых членов альянса, этот разрыв препятствует также внедрению ИИ. В качестве второй, не менее значимой причины затягивания модернизации НАТО, можно назвать ЗИИИ (Kasapoglu and Kirdemir, 2019).

НАТО уделила значительное внимание проблеме ЗИИИ. Так, 3 июня 2020 г. на официальном сайте альянса было опубликовано заявление Североатлантического совета о злонамеренных кибернетических действиях в рамках пандемии коронавируса. В данном заявлении альянс осуждает кибератаки, направленные на дестабилизацию работы медицинских учреждений и нарушение поставок необходимых лекарств. Помимо этого, совет заявляет о готовности оказывать помощь государствам, пострадавшим от данных атак. В июле 2021 г. НАТО обвинил Китай во взломе серверов Microsoft Exchange. Ряд организаций, находящихся на территории стран Европейского Союза, США и Великобритании пожаловались на взлом и атаку ботов на почтовый сервис со стороны неизвестных хакеров. В результате рейда компании понесли экономические потери, а доверие к Microsoft было подорвано. В ответ на это НАТО также выпустил заявление, осуждающее Китай и выражающее солидарность по отношению к пострадавшим.

Так или иначе, говоря о мерах, применяемых для борьбы с ЗИИИ, стоит сказать, что на данный момент в открытых источниках НАТО ограничивается предупреждениями и письменными заявлениями. Однако об озабоченности данной проблемой со стороны альянса говорит тот факт, что на данный момент в его рамках продолжается разработка Стратегии по ИИ. По заявлениям представителей НАТО, цель стратегии – стандартизация использования ИИ. Официальная публикация должна была состояться этим летом, но Заместитель Генерального секретаря НАТО Мирча Джоанэ заявил, что работа над документом по-прежнему ведется.

Обсуждение угрозы ЗИИИ в сфере ИПБ ведется на уровне Европейского центра по противодействию гибридным угрозам, совместного органа ЕС и НАТО. Данный вызов стал одной из главных тем для обсуждения на встрече экспертов по кибербезопасности в феврале 2020 г. В ходе данной встречи эксперты обсудили необходимость обеспечения защиты от ЗИИИ в таких жизненно важных областях, как медицина и энергетика, а также пагубные последствия, к которым может привести бездействие. Здесь же эксперты отмечают возникновение каскадного эффекта и перетекание последствий ЗИИИ из одной сферы в другую.

В военной сфере особое внимание экспертов НАТО и ЕС вызывает потенциал применения дронов, значительно меняющий представление о традиционных методах ведения боевых действий. Отсутствие необходимости постоянного человеческого контроля увеличивает точность ударов, сводит на «нет» влияние человеческого фактора, а также приводит к уменьшению людских жертв для применяющей стороны.

Помимо этого, данное вооружение весьма доступно и дешево в производстве. С другой стороны, решения ИИ не строятся на морально-человеческих соображениях, что по мнению экспертов, может оказать значительное влияние не только на традиционные стратегии конфликтов, но и на нормативно-правовые базы (добавим – и на психологические установки операторов дронов).

Наконец, еще одной уязвимой для ЗИИИ сферой, отмечаемой экспертной группой, является информация, где выделяются спам-рассылки, а также технологии дипфейк. Данные технологии позволяют эффективно влиять на общественное мнение в интересах применяющей их стороны. Отдельного внимания заслуживает тот факт, что в качестве стран, активно применяющих данные технологии, эксперты НАТО и ЕС отметили Россию и Китай (Hybrid CoE Trend Report, 2021).

На уровне заявлений альянс отрицательно относится к агрессивному применению технологий ИИ. Однако здесь возникает явное противоречие, связанное с тем, что одним из приоритетных направлений для внедрения ИИ в НАТО, опираясь на данные указанного выше Стратегического командования НАТО по трансформации являются вооруженные силы, включая как их оборонительный, так и наступательный компонент в том числе системы ПВО и ПРО (Hill and Marsan, 2018) ВВС и ВМС. Данные системы применяются Североатлантическим альянсом на фоне и в качестве средств агрессивного отстаивания своих геополитических интересов на территории Европы, начиная с 2007 г. Система ПРО НАТО с того момента претерпела значительное расширение и продолжает модернизироваться, в том числе посредством активного внедрения ИИ.

Другим примером провокационного применения средств ИИ со стороны альянса является создание и развитие системы воздушной разведки и наблюдения AGS (Alliance Ground Surveillance). В рамках данной системы уже закуплено 5 БПЛА нового поколения, а также проведена модернизация наземного и космического разведывательных сегментов, которые будут обеспечивать работу данной системы. Проект AGS предполагает постоянное патрулирование данных беспилотников в международном воздушном пространстве и в приграничном воздушном пространстве НАТО. Учитывая регулярные патрульные полеты самолетов AWACS вблизи от российских границ, создание и закрепление еще одной разведывательной системы, действующей, к тому же, на основе ИИ, может внести дополнительные военные риски. Также это идет вразрез с позицией НАТО относительно применения ИИ для достижения политических интересов.

Так или иначе, Североатлантический альянс скорее положительно воспринимает внедрение ИИ, который находится в списке «новых и прорывных технологий», наравне с автономными и квантовыми технологиями. Помимо этого, реализуются различные проекты альянса по внедрению ИИ в финансовую сферу, логистику и военную отрасль. С другой стороны, в этом году началась работа по подготовки Стратегии ИИ, целью которой является формирование «правил» применения данного новшества, а также противодействия ЗИИИ со стороны потенциальных противников. При этом, несмотря на негативное отношение к использованию технологий ИИ для достижения внешнеполитических интересов, НАТО нередко прибегает к данному средству, в том числе для выполнения своих геополитических задач.

Библиография

Catania, G., Ganga, L., Milardo, S., Morabito, G., and Mursia, A. (2018). An AI-Assisted Cyber Attack Detection Framework for Software Defined Mobile Networks. *NATO Science and Technology Organization MP-IST-160-S2-1,* 1-12

Hill, S., Marsan, N. (2018). Artificial Intelligence and Accountability: A Multinational Legal Perspective. *NATO Science and Technology Organization MP-IST-160-PP-4,* 1-10

Christie, E. H. (2020). *Artificial Intelligence at NATO: dynamic adoption, responsible use.* Retrieved 3 November 2021, from https://www.nato.int/docu/review/articles/2020/11/24/artificial-intelligence-at-nato-dynamic-adoption-responsible-use/index.html;

Kasapoglu, C., Kirdemir, B. (2019). *Artificial Intelligence and the Future of Conflict.* Retrieved 3 November 2021, from https://carnegieeurope.eu/2019/11/28/artificial-intelligence-and-future-of-conflict-pub-80421;

NATO Allied Command Transformation. (2020). *Military Uses of Artificial Intelligence, Automation, and Robotics (MUAAR).* Retrieved 3 November 2021, from https://www.act.nato.int/application/files/5515/8257/4725/2020_mcdc-muaar.pdf

Pauwels, E. (2021). *Cyber-biosecurity: How to protect biotechnology from adversarial AI attacks.* Retrieved 3 November 2021, from https://www.hybridcoe.fi/wp-content/uploads/2021/05/20210503_Hybrid_CoE_Strategic_Analysis_26_Cyber_biosecurity_WEB.pdf;

Hybrid CoE Trend Report. (2021). *The future of cyberspace and hybrid threats*. Retrieved 3 November 2021, from https://www.hybridcoe.fi/wp-content/uploads/2021/04/20210407_Hybrid_CoE_Trend_Report_6_The_future_of_cyberspace_and_hybrid_threats_WEB.pdf

Thiele, R. (2020). Artificial Intelligence – A key enabler of hybrid warfare. Retrieved 3 November 2021, from https://www.hybridcoe.fi/wp-content/uploads/2020/07/WP-6_2020_rgb-1.pdf

Willemo, J. (2019). Trends and Developments in the Malicious Use of Social Media. Retrieved 3 November 2021, from https://stratcomcoe.org/pdfjs/?file=/cuploads/pfiles/nato_report_-_trends_and_developments.pdf?zoom=page-fit

Дипфейки как угроза информационно-психологической безопасности личности и общества: взгляд Европейского союза

Любовь Шматкова

Традиционно под дипфейком понимают технологию создания с помощью ИИ реалистичного аудио- и видеоконтента. В самом понятии отражена комбинация двух терминов: «глубокое обучение» (deep learning) и «подделка» (fake). Первое является подтипом машинного обучения, в рамках которого искусственные нейронные сети обучаются на больших объемах данных, второе указывает на характеристику итогового продукта.

Несмотря на то, что технология является относительно новой и используется преимущественно для создания рекламного и развлекательного контента, широко обсуждается практически полезный потенциал дипфейков для других областей. К возможным сферам применения данной технологии относят образование (например, наглядное представление исторических событий в рамках школьного обучения или моделирование человеческого поведения при изучении социальных наук), искусство (прежде всего, в области кинематографии и сатиры), медицина (научные исследования биотехнологий), социальная сфера (для реконструкции лиц или создания собственного голоса у людей, получивших увечье или страдающих некоторыми формами паралича) (Chesney and Citron, 2019).

Однако создание дипфейков все чаще рассматривается как одна из потенциально наиболее опасных форм ЗИИИ. Это связано как со спецификой самой технологии, так и с особенностями современного информационного развития. Во-первых, беспрецедентной скоростью передачи информации в современном мире: благодаря быстрому и доступному Интернету обмен сообщениями происходит практически мгновенно. Во-вторых, высоким уровнем охвата: социальные сети и мессенджеры обеспечивают распространение информации среди десятков миллионов пользователей. В-третьих, объемами передаваемой информации: наличие плотного потока постоянно обновляющихся новостей снижает способность большинства людей к их критическому восприятию. В то же время качество и доступность инструментов для создания дипфейков продолжает повышаться. Как следствие, технология дипфейков является угрозой не только отдельным людям, которые могут стать жертвами обмана, но обществу и государству в целом, так как число введенных в заблуждение людей может быть критически велико.

Кроме того, как отмечается в докладе нидерландского Института Ратенау, технология дипфейков не только является потенциально опасной на разных социальных уровнях (индивидуальном, организационном, национальном), но и несет риски причинения вреда разного характера: психологического (клевета, запугивание, подрыв доверия), финансового (вымогательство, кража личных данных, мошенничество, репутационный ущерб), социального (ущерб экономической стабильности, демократии,

правосудию, национальной безопасности, международным отношениям) (Tackling deepfakes, 2021).

Всё это открывает возможность для совершения ряда злоупотреблений и противоправных действий с помощью рассматриваемой технологии: от нарушения авторских прав до разжигания социальных волнений. Согласно анализу Европейского полицейского агентства (Европол), можно выделить несколько потенциальных сценариев использования дипфейков для противоправных целей (Malicious Uses and Abuses, 2020).

Во-первых, это применение дипфейков для проведения кампаний дезинформации, а также с целью вмешательства в выборы, разжигания социальных конфликтов или общественных беспорядков. Примерами такого использования может служить кампания дезинформации с помощью массового распространению правдоподобного фальшивого аудиовизуального контента с участием высокопоставленных государственных чиновников.

Во-вторых, возможно осуществление крупного мошенничества с ценными бумагами, например, с использованием фальшивых видео- и аудиозаписей с руководителями крупных компаний, что может приводить к резким колебания цен на фондовой бирже. Кроме того, фальшивые аудио- и видео сообщения могут быть использованы преступниками для получения коммерческой тайны компании или конфиденциальных данных путем введения в заблуждения сотрудников. Будучи опубликована, такая украденная информация может серьезно повлиять на стоимость акций компании. Масштабное использовании дипфейков мошенниками может нанести серьезную угрозу стабильности мировой финансовой системы.

Еще одной формой противоправного использования технологии дипфейков может стать совершение онлайн-преступлений в отношении детей. Несмотря на наличие на международном уровне консенсуса о необходимости согласования международных механизмов защиты детей от наиболее изощренных угроз онлайн (например, сексуальной эксплуатации и буллинга), эффективные меры противодействия этим преступлениям все еще разработаны. Использование дипфейков существенно повышает риски в этой сфере: возможность генерирования контента путем преобразования взрослых лиц в детские может использоваться для шантажа, а также затрудняет привлечение к ответственности преступников.

Кроме того, серьезной потенциальной угрозой является более сложная форма вымогательства, когда под угрозой распространения поддельного видео, которое может нанести ущерб репутации жертвы, преступники получают возможность требовать не только денежный выкуп, но и ценную информацию. Эта форма вымогательства может быть совмещена с атакой программы-вымогателя. В случае, если речь идет об угрозах в отношении публичной фигуры или государственного деятеля, последствия могут иметь глобальное значение.

Аудиовизуальный контент, созданный с помощью дипфейков, может быть злонамерено использован в качестве доказательств для влияния на результаты уголовного расследования и судебного разбирательства. Распространение подобных

фальшивых материалов будут ставить под сомнение аудиовизуальные доказательства как целую категорию.

Следует, однако, принять во внимание, что результат воздействия дипфейков, особенно используемых в целях дезинформации, не является долгосрочным и имеет кратковременный эффект (так называемое «окно эффективности»). Это качество способствует тому, что дипфейки могут иметь решающие значение при немедленной эмоциональной реакции, или в обстоятельствах, когда имеет значение время, например, в последний день перед выборами (Malicious Uses and Abuses, 2020). Следовательно, принципиально важен момент появления подобного контента по отношению к значимым социальным и политическим событиям или в сферах, где скорость распространения точной информации может оказать существенное влияние на процессы (как на фондовом рынке).

Хотя ключевым инструментом противодействия злонамеренному использованию дипфейков остаются методы их технического выявления и распознавания, а также гибкое правовое регулирование (Meskys et al., 2019), рекомендуется уделять внимание социально-психологическим аспектам. Так как уровень информированности населения об угрозах такого рода все ещё остается невысоким, решающие значение может иметь активизация усилий по повышению осведомленности о дипфейках.

Библиография

Chesney, R. and Citron, D. K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review 1753*. Retrieved 3 November 2021, from https://ssrn.com/abstract=3213954 or http://dx.doi.org/10.2139/ssrn.3213954

Malicious Uses and Abuses of Artificial Intelligence. (2020). Retrieved 3 November 2021, from https://www.europol.europa.eu/publications-documents/malicious-uses-and-abuses-of-artificial-intelligence

Meskys, E., Kalpokiene, J., Jurcys, P. and Liaudanskas, A. (2020). Regulating Deep Fakes: Legal and Ethical Considerations. *Journal of Intellectual Property Law & Practice,* Volume 15, Issue 1, January 2020, pp. 24–31.

Tackling deepfakes in European policy. (2021). Retrieved 3 November 2021, from https://www.rathenau.nl/en/digital-governance/tackling-deepfakes-european-policy

Заключение

Представленные доклады и научная дискуссия в ходе проведенных семинаров выявили комплексность проблемы ЗИИИ, требующую комплексных решений с опорой на междисциплинарные исследования. На основе изучения ЗИИИ как угрозы ИПБ в Северо-Восточной Азии (как авторами докладов сборника, так и более широким кругом специалистов – коллективом проекта №21-514-92001 РФФИ и ВАОН, международной группой по исследованию угроз международной ИПБ посредством ЗИИИ (Research MUAI)), можно выделить целый ряд направлений будущих исследований. Ниже – только некоторые из них:

- оценка влияния особенностей современных культур (в том числе в политическом, деловом аспектах), вкупе с актуальными экономическими, политическим и социальными условиями, на восприятие информации аудиториями в разных странах;
- оценка психологической уязвимости аудиторий в разных странах, которыми могут воспользоваться субъекты ЗИИИ;
- определение вероятности и сценариев ЗИИИ на основе уровня развития и доступности ИИ в той или иной стране или регионе, а также в глобальном масштабе;
- оценка инструментария, который уже доступен агрессивным политическим акторам и появится в будущем для осуществления ЗИИИ против ИПБ;
- выявление возможностей и способов предотвращения такого ЗИИИ и противодействия ему, минимизации его негативного информационно-психологического влияния;
- разработка рекомендаций по противодействию угрозам ИПБ, обусловленным ЗИИИ, для государственных органов, институтов гражданского общества, бизнес-структур.

Для решения этих и других задач необходимо создание крупных международных лабораторий, в которых учитывались бы новейшие достижения стран-лидеров, как в сфере технических, так и в сфере гуманитарных наук. Работа по созданию таких лабораторий уже активно проводится. Частью ее стал и проект, финансируемый РФФИ и ВАОН, в рамках которого подготовлен этот сборник, положивший основу дальнейшей совместной работе молодых специалистов из разных стран. Научное сотрудничество России и Вьетнама, способное обеспечить взвешенный «взгляд со стороны» на происходящее в других странах, может стать прологом к более широкому международному сотрудничеству в борьбе с угрозами ЗИИИ в сфере ИПБ.

Информация об авторах

Хавьер АНИАС (Куба) – студент магистратуры, инженер-программист в Лаборатории по изучению искусственного интеллекта, Университет информатики, Гавана.

Райнел БАТИСТА (Куба) – главный преподаватель и руководитель Национальной инженерной программы по кибербезопасности, Университет информатики, Гавана.

Валерия Игоревна БУЛВА (Россия) – эксперт, помощник директора Центра международной информационной безопасности и научно-технологической политики, Московский государственный институт международных отношений (университет) МИД России.

Юрий Юрьевич КОЛОТАЕВ (Россия) – аспирант, Санкт-Петербургский государственный университет.

Дарья Олеговна МАТЯШОВА (Россия) – студентка магистратуры, Санкт-Петербургский государственный университет.

Екатерина Андреевна МИХАЛЕВИЧ (Россия) – аспирантка, Санкт-Петербургский государственный университет.

Каллакури РАДХАКРИШНА (Индия) – студент магистратуры, Международная лаборатория прикладных сетевых исследований, Национальный исследовательский университет «Высшая школа экономики», Москва.

Лиомар РОДРИГЕС (Куба) – инженер-программист, Центр телематики, Университет информатики, Гавана.

Сергей Александрович СЕБЕКИН (Россия) – кандидат исторических наук, преподаватель кафедры политологии, истории и регионоведения, Иркутский государственный университет.

Нгуен Минь ТЫОНГ (Вьетнам) – кандидат технических наук, кафедра систем связи и телекоммуникаций, МИРЭА – Российский технологический университет, Москва.

Олег Владимирович ФИЛАТОВ (Россия) – аспирант, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Москва.

Любовь Павловна ШМАТКОВА (Россия) – преподаватель кафедры региональных проблем мировой политики факультета мировой политики, Московский государственный университет им. М. В. Ломоносова.

Научный редактор сборника

Дарья Юрьевна БАЗАРКИНА – доктор политических наук, ведущий научный сотрудник Отдела исследований европейской интеграции Института Европы Российской академии наук. Профессор кафедры международной безопасности и внешнеполитической деятельности России Института права и национальной безопасности Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации. Член международной группы специалистов по исследованию угроз международной ИПБ посредством ЗИИИ (Research MUAI). Координатор программ по коммуникационному менеджменту и стратегической коммуникации Международного центра социально-политических исследований и консалтинга (МЦСПИК).

Notes on Contributors

Javier ANIAS (Cuba) is Master's student and Software Engineer at the AI Research Lab, University of Informatics Sciences, Havana.

Raynel BATISTA (Cuba) is a Principal Professor and Chair of Lectures at National Cybersecurity Engineering Program, University of Informatics Sciences, Havana.

Valeriia BULVA (Russia) is an expert and assistant director of the Center for International Information Security, Science and Technology Policy at Moscow State Institute of International Relations of the Ministry for Foreign Affairs.

Oleg FILATOV (Russia) is a PhD student at Russian Presidential Academy of National Economy and Public Administration, Moscow.

Yury KOLOTAEV (Russia) is a PhD student at Saint Petersburg State University.

Darya MATYASHOVA (Russia) is a Master's student at Saint Petersburg State University.

Ekaterina MIKHALEVICH (Russia) is a PhD student at Saint Petersburg State University.

Kallakuri RADHAKRISHNA (India) is a Master's student at the International Laboratory for Applied Networks Research at the National Research University Higher School of Economics, Moscow.

Liomar RODRÍGUEZ (Cuba), Software Engineer, Center for Telematics, University of Informatics Sciences, Havana.

Sergey SEBEKIN PhD (Russia) is a Lecturer at the Department of Political Science, History and Regional Studies at the Irkutsk State University.

Liubov SHMATKOVA (Russia), Lecturer at the Department of Regional Problems of World Politics at the Faculty of World Politics, Lomonosov Moscow State University.

Nguyen Minh TUONG PhD (Vietnam) works at the Department of Communication Systems and Telecommunications, MIREA — Russian Technological University, Moscow.

Editor

Darya BAZARKINA DCs is a Leading Researcher at the Department of European Integration Research, Institute of Europe of the Russian Academy of Sciences. She is a Professor at the Department of the International Security and Foreign Policy of Russia, Institute of Law and National Security, Russian Presidential Academy of National Economy and Public Administration. Darya is also a member of the International Group of Experts on Threats for International Psychological Security through the Malicious Use of Artificial Intelligence (Research MUAI) and the Research Coordinator on Communication Management and Strategic Communication at the International Center for Social and Political Studies and Consulting (ICSPSC).

International Center for Social and Political Studies and Consulting (ICSPSC)

The International Center for Social and Political Studies and Consulting (ICSPSC) was founded in March 2002 as an association of researchers and consultants from different countries. Over the years, the ICSPSC has organized hundreds of international academic conferences, roundtable discussions, and workshops concerning the issues of national and international security and strategic communication, and published about 30 books and different reports. Monographs and collections of articles published by the ICSPSC in Russian and English include:

- Armies and Politics (in English);
- Russia and Latin America (in Russian);
- Russia and India – Strategic Partners(in English);
- Public Relations Training Courses (in Russian);
- Avenir Khanov – a Person, a Citizen, and a Diplomat (in Russian);
- India – Russia: A Dialogue between Civilizations(in English);
- India – Russia: Trade and Economic Relations (in English);
- Genesis of Russia's Market Reforms (in Russian);
- Mass Media and PR in Bulgaria (in Russian);
- Hugo Chavez and the Bolivarian Revolution (in Russian);
- Communication Management. Consulting in Public Relations (in Russian);
- Public Relations and Communication Management: The Foreign Experience (in Russian);
- The Foreign Policy of the USA: The Communication Aspect (in Russian);
- Communication Management in World Politics and Business (in Two Volumes, in Russian);
- The Rising Role of Communication Management in World Politics and Business (in English);
- Ultra-Left Terrorism in Germany: Major Trends in the Activity of the Red Army Fraction (RAF) and its Communication Maintenance (in Russian);
- Communication Management in the Foreign Policy of France in the Late 20th Century (in Russian);
- Communication Management and Strategic Communication(in Russian);
- Crisis, Army, Revolution (in Russian);
- The Presidents in Media Focus: The Practice of Psychological Warfare in Latin America;
- Hugo Chavez and Psychological Warfare in Venezuela (in Russian);
- Communication Management and Strategic Communication: The Modern Forms of Global Influence and Control (in Russian);
- "Ukraine" Strategic Provocation (in Russian);
- Communication and Terrorism (in Russian),
- Strategic Communication in EU-Russia Relations: Tensions, Challenges, and Opportunities (in Russian)
- Malicious Use of Artificial Intelligence and International Psychological Security in Latin America (in English);
- Experts on the Malicious Use of Artificial Intelligence and Challenges to International Psychological Security (in English).

Among the authors of these books are more than 102 researchers from 28 countries in Europe, Asia, and North and South America.

One of the most recent projects of the ICSPSC is the development of international associations that work in various fields of strategic studies and strategic communication. Leading scholars, CEOs, and employees of public and private structures and non-governmental organizations from Asia, Oceania, Africa, Europe, and South and North America are taking part in the activities of these associations (See more at GlobalStratCom: globalstratcom.ru/globalstratcom-eng/).

E-mail: icspsc_office@mail.ru, icspsc@mail.ru

GlobalStratCom

Russia is developing cooperation with different regions of the world. The GlobalStratCom platform aims to develop five associations in various fields of strategic studies and strategic communication. The following are currently in progress:

- European – Russian Communication Management Network (EU-RU-CM Network)
- Russian – Latin American Strategic Studies Association (RLASSA)

Leading scholars, heads, and responsible employees of public and private structures and non-governmental organizations from Asia, Oceania, Africa, Europe, and South and North America are taking part in the activities of these associations.

Research Areas:

- Challenges and threats to national and international security: joint interests and possible areas of collaboration between Russia and other countries;
- Armed Forces and politics;
- Conflict resolution and crisis management;
- Participation in peace missions;
- Malicious use of artificial intelligence and psychological security;
- Participation in wars and military conflicts;
- Prospective models of social and political development;
- New technologies and their influence on social development and security issues;
- Activities of law enforcement agencies;
- Terrorism and communication;
- Armed Forces, State, and Society;
- Strategic communication;
- Military history;
- Strategic studies as an area of cooperation between Russia and other countries;
- War and peace studies.

For more information, see the website of GlobalStratCom (http://globalstratcom.ru/).

Malicious Use of Artificial Intelligence as a Threat to Psychological Security: Northeast Asia and the Rest of the World

Proceedings of the seminars of young researchers at Saint Petersburg State University and the Diplomatic Academy of the Ministry of Foreign Affairs of Russia, October 27 and November 25, 2021

Edited by Darya Bazarkina

Edition of the International Center for Social and Political Studies and Consulting

The collection presents the papers of young researchers on the issues of the malicious use of artificial intelligence to undermine the psychological security of society. The papers, presented at research seminars at Saint Petersburg State University and the Diplomatic Academy of the Russian Ministry of Foreign Affairs in the fall of 2021, contain both an analysis of politically motivated threats to the psychological security of individual countries and regions (paying special attention to the situation in Northeast Asia) and assessments of the current and future threats of the malicious use of certain artificial intelligence technologies, such as deepfakes.

Злонамеренное использование искусственного интеллекта как угроза информационно-психологической безопасности: Северо-Восточная Азия и остальной мир. Материалы научных семинаров молодых исследователей в Санкт-Петербургском государственном университете и Дипломатической академии МИД России, 27 октября и 25 ноября 2021 г.

ISBN 978-5-00166-511-3

Адреса для комментариев: icspsc_office@mail.ru, bazarkina-icspsc@yandex.ru.

Издано в Российской Федерации издательством «OneBook.ru» LLC «SAM Polygraphist».