Malicious Use of AI and Challenges to Psychological Security of BRICS Countries

Report

Research Coordinator: Evgeny PASHENTSEV

Edition by the International Center for Social and Political Studies and Consulting with the help of the International Research Group on Threats to International Psychological Security through Malicious Use of Artificial Intelligence (Research MUAI)

April 2024, Moscow

Malicious Use of AI and Challenges to Psychological Security of BRICS Countries

Report

Research Coordinator: Evgeny PASHENTSEV

Edition of the International Center for Social and Political Studies and Consulting
with the help of the International Research Group on Threats to International
Psychological Security through Malicious Use of Artificial Intelligence (Research MUAI)

April 2024, Moscow

УДК 004.89.056

ББК 16.6

М21

This report is published in English, Chinese and Russian under the general editorship of Evgeny Pashentsev.

Malicious Use of AI and Challenges to Psychological Security of BRICS Countries. Research coordinator: Evgeny Pashentsev. Edition of the International Center for Social and Political Studies and Consulting with the help of the International Research Group on Threats to International Psychological Security through Malicious Use of Artificial Intelligence (Research MUAI). – Moscow: LLC «SAM Polygraphist», 2024. – 120 pp.

ISBN 978-5-00227-204-4

Artificial intelligence (AI) technologies have enormous potential for transformative power and already led to numerous positive changes in the BRICS countries, but they also involve great possible risks, not least related to the activities of various malicious actors. The malicious use of AI is growing in the modern world in quantitative and qualitative terms, which is posing a serious threat to human life, health and well-being. It can both increase the risks of using all modern technologies, and create new ones that cannot be fully foreseen. This report focuses on the threats of malicious AI influence on the human psyche, and through this on political, economic, social processes and the activities of state and non-state institutions in ten BRICS countries. The report is based on a three-level classification of threats of malicious use of AI against psychological security.

Cover image: Freepik.

Signed to print 08.04.2024. Digital printing. Order № 28988.

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

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

www.onebook.ru

Contents

Introduction: The Malicious Use of AI – Growing Threats

Evgeny PASHENTSEV

The BRICS countries are experiencing rapid development and implementation of artificial intelligence (AI) technologies, which brings great economic and social benefits and leads to radical changes in production, finance, trade, transport, education, medicine, and leisure. The functioning of government institutions, political parties, and public organizations is also being increasingly influenced by AI. These advanced AI capabilities offer immense transformative social power and have the potential to bring about numerous positive changes in society, but there are great risks too.

"BRICS countries have agreed to launch the AI Study Group of BRICS Institute of Future Networks at an early date… We need to jointly fend off risks, and develop AI governance frameworks and standards with broad-based consensus, so as to make AI technologies more secure, reliable, controllable and equitable," (CGTH 2023) Chinese President Xi Jinping told the 15th BRICS Summit in South Africa. In 2024 under its leadership term, Russia plans to put AI cooperation firmly on the BRICS agenda for "detailed discussion," according to President Vladimir Putin. He envisions a technological future where AI safeguards align among member states, ensuring both opportunities and risks are managed responsibly (Think BRICS 2023). Speaking at the Global Partnership on Artificial Intelligence Summit in December 2023, Prime Minister of India Narendra Modi said AI can become the biggest tool for development in the 21st century but it also equally be a force in destroying the 21st century. "We have to work together to prepare a global framework for the ethical use of AI," (PTI 2023) he emphasized.

The malicious use of AI (MUAI) in the BRICS countries is on the rise, reflecting a general global trend, and representing a leading risk, since it can both amplify other AI risks and create new ones that cannot be fully foreseen, but which are becoming more and more real. The potential for MUAI today is astounding. Noting this, researchers from Collaborations Pharmaceuticals in cooperation with European scientific institutions conducted a conceptual experiment. Instead of synthesizing new drugs, they asked the opposite of the MegaSyn AI neural network: to identify substances that are the most toxic to the human body. The neural network correctly understood the task and, in under six hours, generated a list of 40,000 substances that are optimal components of chemical and biological weapons. The AI independently designed not only many known chemical warfare agents, but also many new ones that are more toxic. This simple inversion of the ML model turned a harmless generative model from a useful tool into an enabler of mass murder (Urbina et al., 2022).

It is reasonable to suspect that this inversion approach could be applied to other areas, such as finding optimal ways to have negative psychological impacts on the public consciousness. It is important to adequately assess the real threat of MUAI as a means of psychological confrontation. MUAI contributes to the destabilization of human consciousness, which contributes to the destabilization of society, which contributes once again to the destabilization of human consciousness—a dangerous, unsustainable cycle—aiming to "win" in the global redistribution of assets among very narrow circle of superrich. At the same time, the objective problems facing social development—selfish calculations, the actions of anti-social actors, inertia, and the inability of a majority of society to easily enact social change—are closely intertwined.

Military – political aggressive blocks have led the world to destructive wars more than once in history including two world wars in the 20th century. Other associations of countries, however, are looking for an alternative to the present-day social and economic world order threatened by

unrestrained arms races fraught with nuclear Armageddon. The second type of interstate associations can be BRICS, connecting the current fragile and dangerous situation with a better future not through empty promises but through great opportunities at the global level (Pashentsev and Miao, 2023). Even in the current configuration, the BRICS unites countries with more than 45% of the population, 32% of world GDP in terms of PPP (compared to 30% held by the G7 countries), great research and technological potential, and the largest natural resources, which makes it possible to solve global problems in the interests of all mankind. It should also be taken into account that there are about three dozen countries that are ready to join BRICS in one form or another (TV BRICS 2024).

The improvement and implementation of AI technologies can play an important role in solving the numerous and complex tasks of the development of BRICS countries, as well as in ensuring their security from malicious state and non-state actors, including an area of psychological security.

MUAI and Three Levels of Threats to Psychological Security

MUAI is an intentional antisocial action that comes in both explicit and implicit forms. Anti-social circles (ranging from individual criminals to powerful egoistic interest groups, corrupt state institutions) are already leveraging MUAI in pursuit of their interests. Recent years have revealed great potential for MUAI in the psychological field. Although there are a significant and rapidly growing number of academic publications on the technical aspects of MUAI, its general socio-economic and political implications, and the first attempts to classify MUAI (Brundage et al. 2018; Caldwell et al. 2020; Malicious Uses 2020; Blauth et al. 2022), there are relatively few publications on specific MUAI issues in the context of psychological security, and even less the systemic consideration of MUAI to psychological security. The notion of psychological security can be found in many studies (Grachev, 1998; Roshhin and Sosnin, 1995; Afolabi and Balogun, 2017). The renowned U.S. psychologist Abraham Maslow believed that, once the basic physiological needs are met, the need for security moves to the forefront. In more specific terms, it is the need for protection, stability, confidence about the future, good health, etc. (see more: Maslow, et al., 1945). National psychological security is understood as the protection of citizens, individual groups, social groups, large associations of people and the entire country's population from negative psychological influences (Barishpolets, 2013, p. 63; see more: Barishpolets, ed., 2012). Based on the above definitions, the author considers it possible to define international psychological security as protection of the international relations system from negative psychological influences that are associated with various factors of international development. This includes protection from targeted efforts by various state, non-state and supranational actors to achieve partial/complete, local/global, short-term/long-term and latent/open destabilization of the international system in order to gain competitive advantage, even through the physical elimination of enemies.

For all of its importance, separate analysis of malicious psychological impact from deepfakes, bots, predictive analytics and so on, does not take into account the synergy of such an impact, nor does it provide a systemic idea of psychological security risk growth or the risks to the entire national and international security system. This lack of comprehensive analysis is explained by the novelty of the issue: the practice of MUAI cannot outpace the progress of AI.

The first step is to consolidate the efforts of scholars of different countries in the new field, which was done in 2019 with the founding of the international group on studies of MUAI threats for international psychological security (Research MUAI) and the cooperation (e.g., joint research and publications, international conferences, scientific seminars) between researchers from seven countries that followed. Tens of academic articles prepared by the group researchers on different issues of MUAI and PS finally led to the publishing of *The Palgrave Handbook of Malicious Use of AI and Psychological*

Security, which is the first book in this new area. The 23 contributors represent 11 countries across Asia, Europe, and North America (The Palgrave Handbook, 2023). In recent years, both the first systemic studies on MUAI threats to psychological security in BRICS countries have been published (Bazarkina, Pashentsev 2020, PP. 154-177; Pashentsev, Bazarkina 2023), and publications on the analysis of MUAI threats to psychological security in individual BRICS member countries have been appeared too (Bazarkina, Matyashova 2022, PP. 14-20; Bazarkina, Mikhalevich et al. 2023; Cai, Zhang 2023; Gupta, Guglani 2023, etc.).

This report focuses on threats to psychological security through MUAI against the human psyche, and through this on the political, social, economic, cultural processes and the activities of state and non-state institutions in BRICS countries taking into account the expansion of the association.[1] This report does not present the response of the BRICS countries to these new threats, it is at the stage of formation at the national level (from an initial and fragmentary understanding of the severity of these threats in some countries to the adoption of the first legal and technical countermeasures in others) and requires independent consideration.

The report is based on a three-level classification of MUAI threats to psychological security (see more: The Palgrave Handbook 2023, PP. 23-46).

At the first level, these threats are associated with deliberately distorted interpretations of the circumstances and consequences of AI development for the benefit of antisocial groups. In this case, AI itself at this level is not directly involved in the destabilization of psychological security. The destructive (open or hidden) impact imparts a false image of AI in the minds of people. Excessive, artificially created negative reaction to AI development (for example, horror stories that robots and AI will soon force all people out of work, workers will become slaves of AI etc.) has certain political and economic goals and is not as harmless as it may seem. Such a negative reaction can slow down the implementation of progressive, nearly all-encompassing, AI technologies and cause socio-political tensions and conflicts. Inflated public expectations about AI may also manifest, which at a certain stage could result in a natural collapse in the value of high-tech companies and the market as a whole. These expectations can maliciously be used and strengthen to disorient the general public, interested commercial and non-profit structures, and public authorities, and, ultimately, turn into disappointments, wrong decisions, and social and political conflict.

At the second level the field for malicious use is wide open: the unjustified use of drones, cyberattacks on vulnerable infrastructure, the reorientation of commercial AI systems, the use of AI technologies to disrupt decision-making or modify it in a latent way, and much more. But an attack on public consciousness is not its main goal at this level.

The MUAI is designed primarily to cause psychological damage belongs to the *third, and highest, level* of psychological security threats. Synthetic AI products (combining a number of technologies, which can increase the damage from their hacking or malicious use) create a whole range of new risks and threats. Professional use of the means and methods of psychological warfare can shift the threat perception level above or below what is appropriate. Moreover, the use of AI in psychological warfare makes hidden (latent) campaigns of perception management more dangerous, and this will only increase in the future. Therefore, MUAI that is aimed primarily at causing damage in the psychological sphere, deserves independent and very close attention. The first two levels of threat affect human

[1] The BRICS founding countries of Brazil, Russia, India, and China held the first summit in Yekaterinburg in 2009, with South Africa joining the association a year later. BRICS invited Saudi Arabia, the UAE, Argentina, Egypt, Iran, and Ethiopia to join the association during the 15th summit in August last year. Later, the newly-elected President of Argentina, Javier Milei, rejected BRICS membership. Egypt, Ethiopia, Iran, Saudi Arabia, and the United Arab Emirates joined the organization.

consciousness and behavior to varying degrees, and may even be catastrophic for humankind, as would be the case in the event of World War Three. However, the impact of the third level at a certain stage of development can facilitate the influence or even control of antisocial groups over public consciousness; this can result in the sudden destabilization of a particular country or the international situation as a whole. Ultimately, if at the third level there is reliable control over an enemy psychologically, the role of the other two MUAI levels to psychological security becomes auxiliary.

MUAI threats can occur at one level of impact, or at multiple levels at once, as part of a single perception management campaign. The use of a drone by terrorists or the organization of an attack on a civilian population will be second-level threats, which has a communication effect (panic and shock after the attack). However, if criminals accompany their actions with broad information support (also with the help of AI), the threat will reach the third level.

AI is not a single technology. There are many AI technologies applied to numerous functions through various applications in different environments and modalities under different circumstances. The authors of this report take into account how different technologies under the general AI umbrella help create a particular product, seriously changing the technological level and practical capability of any particular type of activity.

Threats from the MUAI are gaining new relevance globally and in BRICS countries at all three levels with the growth of geopolitical rivalries, with the activity of different state and non-state antisocial actors, and with the development and growing affordability of various AI technologies, making them more accessible. This cannot lead anywhere but to attempts by various interest groups to use AI to influence the public consciousness for their own purposes. Such attempts to manipulate the public consciousness are particularly destructive during historical moments of crisis. The inhumanity of fascism became apparent to the absolute majority of humankind after the deaths of over 50 million human beings in the flames of the Second World War. Before the war, however, manipulation of the public consciousness ensured Hitler's victory in the 1933 Reichstag elections. This not-so-distant history remains highly instructive for those alive today. It is understandable that modern governments and political figures in BRICS, many other countries, are exhibiting rising concern over the threat of high-tech disinformation on the Internet and the role of private leading media platforms that use AI technologies.

MUAI threats to psychological security in the BRICS countries arise both for internal reasons and are the result of external factors. Therefore, it makes sense here to give some general idea of the nature and dynamics of threats at three levels in the global dimension.

MUAI at the First Level of Threats to Psychological Security

At the first level of MUAI threats to psychological security, a favorable ground for such threats is created due to the increase in negative perception of the further development and use of AI technologies in the modern world, which is especially characteristic of Western countries. Consistent with the Pew Research Center survey conducted in August 2023, 52% of US citizens say they feel more concerned than excited about the increased use of AI. Just 10% say they are more excited than concerned, while 36% say they feel an equal mix of these emotions. The share of Americans who are mostly concerned about AI in daily life is up 14 % since December 2022, when 38% expressed this view. The rise in concern about AI has taken place alongside growing public awareness. Nine-in-ten adults have heard either a lot (33%) or a little (56%) about AI. The share who have heard *a lot* about AI is up 7 points since December 2022. Those who have heard a lot about AI are 16 points more likely now than they were in December 2022 to express greater concern than excitement about it. Among this most

aware group, concern now outweighs excitement by 47% to 15%. In December, this margin was 31% to 23% (Tyson, Kikuchi 2023).

In collaboration with the University of Queensland, KPMG Australia led the world-first deep dive into trust and global attitudes towards AI across 17 countries. They surveyed over 17,000 people from 17 countries covering all global regions: Australia, Brazil, Canada, China, Estonia, Finland, France, Germany, India, Israel, Japan, the Netherlands, Singapore, South Africa, South Korea, the UK, and the USA. These countries are leaders in AI activity and readiness within their region. According to the KPMG study in the BICS[2] countries, most people (56–75%) trust in AI systems, with people in India reporting the greatest willingness to trust, followed by China. In contrast, a minority of people in other countries report trusting AI, with the Finnish reporting the lowest trust (only 16%). The higher trust and acceptance of AI in the BICS countries is likely due to the accelerated uptake of AI in these countries, and the increasingly important economic role of emerging technologies. The people in the BICS countries are the most positive about AI, perceive the most benefits from it, and report the highest levels of AI adoption and use at work (Gillespie et al. 2023, P. 14). People in the western countries and Japan are particularly unconvinced that the benefits of AI outweigh the risks. In contrast, the majority of people in the BICS countries and Singapore believe the benefits outweigh the risks (The Global Risks Report 2024, P. 3).

Distrust or fear of AI, apparently, is due to an extreme drop in trust into the major societal institutions. For instance, in the USA Gallup recorded in 2022 significant declines in public confidence in 11 of the 16 institutions it tracks annually, with the presidency and Supreme Court suffering the most. The share of Americans expressing a great deal or fair amount of confidence in these fell 15 and 11 percentage points, respectively. Neither score recovered appreciably in the latest survey conducted June 1-22, 2023, with confidence in the court now at 27% and the presidency at 26%. The five worst-rated institutions -- newspapers, the criminal justice system, television news, big business and Congress -- stir confidence in less than 20% of Americans, with Congress, at 8%, the only one in single digits. Most of the institutions rated in 2023 are within three points of their all-time-low confidence score, including four that are at or tied with their record low. These are the police, public schools, large technology companies and big business (Saad 2023).

Is it possible to expect the public to believe in the ability of the authorities to achieve socially oriented use of AI with such indicators of trust in the major societal institutions? The answer is obvious. It is not AI that is not trusted (today it is only machine intelligence), but the authorities, big business, large technology companies that direct its development. An almost irreconcilable political split (especially obvious and strategically dangerous in the United States, up to the threat of a coup d'etat and civil war (Marche 2022a, Pashentsev 2022, Walter 2022a), the degradation of the ruling establishment, the disruptive role of a ruling oligarchy (Collins et al. 2021; Gilens, Page 2014), foreign policy aggressiveness (Abelow 2022, Sachs 2018), low economic growth rates and acute social contradictions are objective indicators of the inability of Western elites to secure the development and application of AI, to counteract MUAI. It is logical to assume that institutions with such a level of trust themselves generate malicious actors, and, increasingly, are at least partly carriers of MUAI.

Fears about the future of AI have only increased over the past year with the rapid progress of generative AI amid the deepening global crisis. Many top business leaders are seriously worried that AI could pose an existential threat to humanity in the not-too-distant future. Forty-two percent of CEOs surveyed at the Yale CEO Summit June 2023 say AI has the potential to destroy humanity five to ten years from now, according to survey results shared exclusively with CNN. The survey included

[2] The authors of the KPMG study did not include Russia in their sampling, and therefore use the acronym BICS in this report. For information on attitudes towards AI in Russia and the new BRICS members, see the relevant chapters of this report.

responses from 119 CEOs from a cross-section of business, including Walmart CEO Doug McMillion, Coca-Cola CEO James Quincy, the leaders of IT companies like Xerox and Zoom as well as CEOs from pharmaceutical, media and manufacturing. “It’s pretty dark and alarming,” Yale professor Jeffrey Sonnenfeld said in a phone interview, referring to the findings (Egan 2023).

The Global Risks Report 2024 presents the findings of the Global Risks Perception Survey (GRPS), which captures insights from nearly 1,500 global experts. Surveyed in September 2023, the majority of respondents (54%) anticipate some instability and a moderate risk of global catastrophes, while another 30% expect even more turbulent conditions. The outlook is markedly more negative over the 10-year time horizon, with nearly two-thirds of respondents expecting a stormy or turbulent outlook (The Global Risks Report 2024, P. 6). In Global risks ranked by severity over 10-year period adverse outcomes of AI technologies come to the six position (The Global Risks Report 2024, P. 8). The Global Risks Report 2024 draws attention to the fact that “Technological power in the hands of the unelected is seen …to be a bigger concern than power concentrated in government. The influence of Big Tech companies is already transnational, competing with the likes of nation states, and generative AI will continue to catalyze the power of these companies and associated founders.” (The Global Risks Report 2024, P. 54).

The largest companies in the field of high technologies actively use AI according to their narrow corporate interests, which rather often go against the interests of society. It is clear that companies with access to large amounts of data to power AI models are leading AI development. Key groups within AI include GAFAM—Google (Alphabet), Apple, Facebook (Meta), Amazon, and Microsoft, also known as the Big Five, which is a name given to the five largest, most dominant, and most prestigious companies in the information technology industry of the United States, early-mover IBM, and hardware giants Intel and NVIDIA (Lee, 2021). Of course, there are serious questions to the largest private technology companies in other countries, but their current global role is significantly lower than that of US companies, whose rapid enrichment, enormous influence and possible existential risks of their narrow corporate control over promising advanced forms of AI are causing growing concern around the world.

It is hardly accidental that among the individuals with the ten largest fortunes in the world, six represented in 2021 Amazon (1), Microsoft (2), Google (2), and Facebook (1) (Forbes, 2021). $7.5 trillion: that was the combined market capitalization of GAFAM at the end of 2020, according to an analysis by the Wall Street Journal. At the end of 2019, these firms’ combined market capitalization was $4.9 trillion, which means they increased in value by 52% in a single year. As of November 12, 2021, the capitalization of these companies has grown by another $2.5 trillion and reached approximately $10 trillion (Statista, 2021a). That was nearly a quarter of the combined $41.8 trillion market capitalization of all companies in the S&P 500 (La Monica, 2021). It is appropriate to recall that the United States’ nominal GDP in 2020 was around $21 trillion. Japan, the world’s third-largest economy, had a GDP of about $5 trillion, and Russia had one of only about $1.5 trillion.

An extremely dangerous role of the Western technology companies became even more obvious after the start of Russia's special military operation in Ukraine. In addition to the sanctions imposed by national governments on Russia, tech firms have emerged as additional geopolitical actors, capable of actively punishing a global power for a military campaign. To demonstrate support for Ukraine, a growing number of tech vendors suspended business in Russia, including Accenture, Adobe, Cisco, Oracle, Dell, IBM, Microsoft and many others (NS Business, 2022; Fried, 2022). This has, of course, caused serious damage to the Russian IT sector and the economy as a whole, but the moves have also negatively affected the companies that have withdrawn from the Russian market.

The first months of 2022 turned out to be difficult for US Big Tech, which rely heavily on digital advertising. Surging inflation, the Ukrainian crisis and other unfavorable macro factors forced

advertisers to slash marketing budgets, which translated into lower profits for platforms like YouTube, Google and Facebook (Cao, 2022). The military hostilities in Ukraine has shattered the myth of neutrality. For most of their existence, Internet companies have argued that they are only neutral content distribution platforms—they are not responsible for the content that is distributed (Feldstein 2022).

Having dispensed with the veneer of neutrality, even after suffering significant losses, Big Tech has actually gained quite a lot.

First, Big Tech has avoided an international confrontation between nation states and the general public on the one hand, and Big Tech on the other. Such a confrontation could have emerged simply from the fact that Big Tech is taking on state and non-state actors that have very different aspirations. However, now and in the near future, Big Tech might not be afraid of political initiatives or international coalitions that might emergence from the United Nations or other international bodies to limit its independence.

Second, Big Tech has demonstrated that it is a powerful tool for confronting Russia in cyberspace. Brad Smith, president and vice chair of Microsoft, writes in no uncertain terms about the role of his company in Ukraine. He notes, “Ukraine’s government has successfully sustained its civil and military operations by acting quickly to disburse its digital infrastructure into the public cloud, where it has been hosted in data centers across Europe. This has involved urgent and extraordinary steps from across the tech sector, including by Microsoft. While the tech sector’s work has been vital, it’s also important to think about the longer-lasting lessons that come from these efforts” (Microsoft, 2022). General Paul Nakasone, director of the National Security Agency, confirmed for Sky News in June 2022 that for the first time, the United States was conducting offensive hacking operations in support of Ukraine: “We’ve conducted a series of operations across the full spectrum; offensive, defensive, [and] information operations” (Martin 2022). Such operations are impossible without support from Big Tech. Therefore, high-tech agenda setting in the United States, which today is unthinkable without the full use of AI technologies, has turned out to be openly subordinated to military-political interests and the waging of psychological warfare. “They are actually ‘shooting’! This is extraordinary,” exclaimed Matthew Schmidt, a professor of national security at the University of New Haven, accusing Western technology companies of accelerating their involvement in military conflicts (Global Times, 2022).

Third, a resource that is an unqualified necessity for protecting national security, and that is already actively being used involved in war, cannot a priori be antinational, which reduces the democratic public’s ability to criticize Big Tech.

Fourth, any US government will need information and analytical support during the “Cold Hot War,” which Big Tech can provide based on AI developments, including support against “internal” enemies and “disinformation.” More than fifty officials in President Biden ‘s administration across a dozen agencies have been involved in efforts to pressure Big Tech companies to crack down on alleged misinformation, according to documents released on August 31, 2022. The documents were part of a preliminary hearing in a lawsuit against the government brought by the attorneys general of Missouri and Louisiana, later joined by experts maligned by federal officials. “When the federal government colludes with Big Tech to censor speech, the American people become subjects rather than citizens,” Louisiana Attorney General Jeff Landry said in a statement (Stieber, 2022).

Fifth, Big Tech’s close cooperation with the military-industrial complex during the new Cold War can fully compensate for the losses from leaving the Russian market. Under Cold War conditions, it is easier to avoid public scrutiny and scandals from promising developments that carry great risks for humanity, along with great profits.

In the United States, due to the size of the economy, military allocations exceed the spending of the next nine largest economies combined (PGPF, 2022). Combined with the global role of digital platforms, the level of AI technology development and acute political confrontation, the transformation of agenda setting into a tool of psychological warfare is the most obvious outcome. Unfortunately, similar processes are developing with varying intensity in other countries as well. However, the militarization of agenda setting and its "legalization" as a tool of psychological warfare is unlikely to be applied to humanity's social needs; on the contrary, these developments push those needs further into the background.

The deterioration of the international situation, the continuation of military operations in Ukraine for more than two years, the bloody conflict in Gaza, as well as other military conflicts, the energy crisis, the alternation of recession with weak growth in the European Union, disruption of supply chains and other factors led to a decrease in the market value of major technological companies, but did not eliminate their leading role in big business. In the context of the current paradigm of building larger- and larger-scale AI systems, there are limited opportunities to develop AI without Big Tech. With vanishingly few exceptions, every startup, new entrant, and even AI research lab is dependent on GAFAM. All rely on the computing infrastructure of Microsoft, Amazon, and Google to train their intellectual systems, and on those same firms' vast consumer market reach to deploy and sell their AI products (Kak et al. 2023). Even Elon Mask decision to buy Twitter in October 2023 was largely motivated by his intention to develop his own AI startups based on Twitter big data capabilities. The same year Mask announced the formation of what he's calling xAI, a company with a mission to "understand the true nature of the universe" (Metz et al. 2023). But pragmatically the main companies of Mask: Tesla, SpaceX, Twitter, Neuralink are deeply interconnected by the rising role of AI in their progress and seems xAI will be a central command team for that.

Big Tech has accumulated, in a contradictory way, scientific and technical power, intellectual resources that are necessary for everyone, and impressive financial opportunities that contribute to economic expansionism. These are all part of the tools of global governance, as well as a growing and more obvious involvement in a geopolitical struggle with an end that has not yet been formed by independent interests. The system-forming elements of global communications and development, such as the current leading digital platforms, cannot be eliminated, but it is certainly necessary to put them under more effective international control to reduce the possibility of their technological potential being used for antisocial purposes. However, only united, socially-oriented actors can control these elements, and these actors in the current socially and geopolitically split world only partially include modern states, leading corporate structures and political parties, opening the door to further MUAI in the agenda setting context.

Further, MUAI already exists on a global scale as a game based on inflated expectations of benefits from the incorporation of AI. This game is played through a versatile psychological impact on target audiences who are particularly susceptible and vulnerable to perception management in a crisis situation. In whose hands are the most advanced tools of global psychological influence and whose financial interests are at stake? Not only is there enough objective data to answer this question; there is an abundance. Therefore, the possible and specific scenarios of combined, targeted impact—not only with the help of specific AI technologies, but also of the very perception of AI—on the public consciousness for the purpose of speculative enrichment and the destabilization of public order requires the most serious attention and comprehensive study by specialists from different countries and with different scientific specializations.

The use of AI technologies, of course, has its risks. Among the risks, one of the first is the risk of mass unemployment due to the widespread introduction of AI technologies and robotization. According to many reports, done 5-10 years ago such as from the UN, the World Economic Forum, the

Bank of America, Merrill Lynch, the McKinsey Global Institute, Oxford University and others (Mishra et al. 2016, Bank of America and Merrill Lynch 2015, Frey and Osborne 2013, 2016; Manyika et al. 2017, UN Conference on Trade and Development 2016, World Economic Forum 2016, Pol and Reveley 2017), 20-30% of jobs or more will disappear in the coming two to three decades as a result of the robotization of manufacturing, finance, services and management; this also includes high-paying positions. In 2016, the World Bank published a report stating that in the coming decades more than 65% of the jobs in developing countries would be threatened by the accelerating development of technology (Mishra et al. 2016, p 23).

More recently, in March 2023, a Goldman Sachs study made a forecast based on data on occupational tasks in both the US and Europe: "If generative AI delivers on its promised capabilities, the labor market could face significant disruption…roughly two-thirds of current jobs are exposed to some degree of AI automation, and that generative AI could substitute up to one-fourth of current work. Extrapolating our estimates globally suggests that generative AI could expose the equivalent of 300mn full-time jobs to automation" (Hatzius et al. 2023).

Catastrophic forecasts of mass unemployment due to the introduction of AI have not yet come true. Moreover, in the coming years, AI together with major labor cuttings may stimulate some increase in labor demand. In place of the departed professions, new ones will appear, including those related to the development and implementation of AI, which are generally more creative in their content. This transformation of the labor market will require great efforts to retrain old and train new personnel. But we are for the first time in history on the way towards the complete (but far from instantaneous) extinction of uncreative activities. However, the system of mass education is far from being ready to provide mass training for specialists in innovative technology development. Many questions arise in this respect. Would it be possible at all to provide such training? Are all equally gifted with abilities for this kind of activity? Even the vast majority of "white-collar" activities are by no means related to innovation.

In addition, many creative jobs are already under increasing pressure from AI. Thus, there is every reason to believe that over time, unemployment problems will sharply worsen with the further development of AI technologies combined with a sharp drop in their cost. The most daring and, perhaps, ultimately, the only correct decisions will be related to the qualitative development of human intellectual and physical capabilities and the creation of hybrid forms of intelligence.

Different aspects of AI security, and many other serious problems of AI development can be used by different malicious actors in their в perception management campaigns against people.

The progress in AI robotics adds more concerns. For example, the *1X's* humanoids use Embodied Learning, integrating AI software directly into their physical forms for advanced capabilities. The overarching objective is to equip humanoids with the ability to comprehend and execute tasks via voice commands, catering to diverse applications ranging from household chores to industrial settings (Malayil 2024). The more complex the AI robot and its activities, the more unexpected fluctuations in the learning and self-learning ANI adaptation process will be, both in a positive and negative sense. Of course, the positive will prevail with the all-round development of humans and their ability for socially necessary work. Otherwise, total robotization will only indicate the uselessness of "consumer humanity," along with its rapid degradation and sad end. This time may come even before the triumph of mass robotization.

The malicious use of AI robots has not only physical but also psychological aspects. This can include provoking fear of a "robot uprising" at the first level of MUAI threats to psychological security, manipulating people with targeted and non-targeted rational-emotional reactions by hacking AI robots at the second level, and issuing false, disorienting information in the interests of a malicious actor at

the third level. The more “human” the appearance and inner world of AI is, the more successful it will be in influencing a person, both positively and negatively.

MUAI at the Second Level of Threats to Psychological Security

At the second level, there is also an increase of threats to psychological security through MUAI. INTERPOL Secretary General Jürgen Stock said at the 90th INTERPOL General Assembly in October 2022 that “Cyber vulnerabilities are increasing: for citizens, governments, industry, for police agencies. Experts estimate cyber-related crime to cause over 10 trillion US dollars in damages by 2025. For INTERPOL – wanted fugitives for digitally-enabled crimes are already the fastest growing dataset” (90th INTERPOL General Assembly 2022). Such a volume of damage (with a global GDP of about 105 trillion dollars in 2023) (Rao P 2023) indicates the almost inevitable impact of organized cybercrime on the governments, and socio-political processes in the global dimension, which is implemented at all levels of threats to PS through MUAI.

Several tools created in 2023 on the dark web are WormGPT and FraudGPT. These models were created specifically for malicious activities and were trained on a large array of data sources, particularly concentrating on malware-related data. FraudGPT, identified in July 2023, does not have built-in controls preventing it from answering questions on criminal activities. This would allow criminals to easily create malicious emails, phishing attacks and provide information to hackers allowing them to choose their victims (Eurojust 2023). FraudGPT is available on a subscription basis, with pricing ranging from $200 per month to $1,700 per year, providing hackers with an AI-driven resource to facilitate their malicious objectives. Moreover, the developer highlighted the 3,000+ confirmed sales and reviews for FraudGPT on the forum and Telegram to lure threat actors (Subhra Dutta T 2023).

As stated on the WormGPT website, this tool “guides hackers through the darkest and most clandestine techniques, promoting immoral, unethical, and illegal behavior” (WormGPT V3.0). The researchers gained access to these malicious AI tools and tested them with various prompts. In a prompt requesting to draft phishing email, FraudGPT even suggested where to place the malicious link for a more efficient attack (Eurojust 2023). The researchers were able to use WormGPT to "generate an email intended to pressure an unsuspecting account manager into paying a fraudulent invoice." The team was surprised at how well the language model managed the task, branding the result "remarkably persuasive [and] also strategically cunning" (Osborne 2023).

AI technologies play a leading role in cybercrime. According to *Arkose Labs* it’s understatement to say the rapid proliferation of generative AI (GenAI) is transforming the landscape of cybersecurity. In fact, GenAI has lowered the barrier to entry for attackers (Arkose Labs 2023, p. 3) For example, only in dating services the threat researchers observed a more than 36,000% increase in fake account creations in Q3 over Q2 in 2023. They also noted a 4,992% increase in bot attacks – both intelligent and basic bots – on dating sites in Q3 over the second quarter. From Q1 2023 to Q2 2023, intelligent bot traffic nearly quadrupled—far outpacing basic bots and heavily contributing to a total increase of approximately 167% for all bot attacks (Arkose Labs 2023, p. 12).

In November 2023 Sumsub, a full-cycle verification platform, has released its third annual *Identity Fraud Report* which provides analysis of identity fraud across industries and regions based on millions of verification checks across 28 industries and over 2,000,000 fraud cases between 2022-2023. According to this report AI-driven fraud remaining the most prominent challenge across various industries, crypto is the main target sector (representing 88% of all deepfake cases detected in 2023), followed by fintech (8%). Deepfakes pave the way for identity theft, scams, and misinformation campaigns on an unprecedented scale. There’s been a significant 10x increase in the number of

deepfakes detected globally across all industries from 2022 to 2023, with notable regional differences: 1740% deepfake surge in North America, 1530% in APAC, 780% in Europe (inc. the UK), 450% in MEA and 410% in Latin America. The country attacked by deepfakes the most is Spain, the most forged document worldwide is UAE passport, whereas Latin America is the region where fraud increased in every country (Sumsub Research 2023).

Deepfakes as an important element of social engineering operations, with a short-term psychological impact on specific people in a specific situation, is an example of borderline threats between the second and third levels. However, when it comes to the explicit or implicit impact of deepfakes on mass audiences with the formation of psychological reactions and actions based on them in the interests of malicious actors, the use of deepfakes belongs to the third level of threats.

MUAI at the Third Level of Threats to Psychological Security

2024 is a year of more than 40 elections worldwide – more fearful than ever for the malicious use of deepfakes. According to The Global Risks Report 2024 misinformation and disinformation is a new leader of the top 10 rankings over the two-year horizon (The Global Risks Report 2024, P.18). Even before the latest Generative Pretrained Transformer (GPT) tools were introduced (e.g. GPT-4, ChatGPT), it was predicted that 90% of online content will be generated by artificial intelligence (AI) by 2026 (Johnson et al.2024). No longer requiring a niche skill set, easy-to-use interfaces to large-scale AI models have already enabled an explosion in falsified information and so-called 'synthetic' content. A growing distrust of information, as well as media and governments as sources, will deepen polarized views – a vicious cycle that could trigger civil unrest and possibly confrontation. New classes of crimes will also proliferate, such as non-consensual deepfake pornography or stock market manipulation (The Global Risks Report 2024, P.18).

Modern AI capabilities already enable one to influence public consciousness. Back on January 1, 2019, the video of Ali Bongo, the president of Gabon, was mistakenly taken for a deepfake, which became the reason for the unsuccessful military coup attempt in that country. Three years later, the use of deep focus technology has a significant impact on the election results. The president-elect of South Korea, Yoon Suk-yeol, used an unusual strategy during his 2022 election campaign. His campaign team used deepfake technology to make an "AI avatar" that helped him win the election. This technology is helpful for appealing to younger voters and to get them more involved (Vastmindz 2022). AI Yoon's creators believe he is the world's first official deepfake candidate—a concept that is gaining traction in South Korea, which has the world's fastest average internet speeds (France 24 2022).

AI technology transformed Yoon Suk-Yeol into a more modern candidate than his competitors from the perspective of younger voters. With neatly combed, black hair and a smart suit, the avatar looks nearly identical to the real candidate but instead, used salty language and meme-ready quips in a bid to engage younger voters who get their news online (Forbes India 2022). Some alarms were raised when the avatar politician used humor to try and deflect attention from Yoon's past scandals (The Times of India 2022). AI Yoon's pronouncements made headlines in the South Korean media, and seven million people visited the "Wiki Yoon" website to question the avatar (France 24 2022). At first glance, AI Yoon could pass for an actual candidate—an apt demonstration of how far artificially generated videos have come in the last few years. "Words that are often spoken by Yoon are better reflected in AI Yoon," said Baik Kyeong-hoon, director of the AI Yoon team (France 24 2022). However, there is a question about what should be done if the avatar of a statesman, politician or business person is a false representation, reinforcing in the public's conscious and subconscious mind *inflated* qualities, creating the illusion of attributes that the real person does not possess. The experience of South Korea's last presidential campaign may have shown, in part, the initial form of a new method of rather dangerous political

manipulation. An increasingly adaptive avatar that does not need rest creates an image that a real person will be able to compete with less and less in the public space. This prompts the question of whether “televised presidents” will soon be replaced by “deepfake presidents”?

In March 2023, Eliot Higgins, the founder of the open-source investigative outlet Bellingcat turned to an AI art generator, giving the technology simple prompts, such as, “Donald Trump falling down while being arrested.” He shared the results — images of the former president surrounded by officers, their badges blurry and indistinct — on Twitter. “Making pictures of Trump getting arrested while waiting for Trump’s arrest,” he wrote. Two days later, his posts depicting an event that never happened have been viewed nearly 5 million times, creating a case study of the power of deepfakes to create confusion in volatile news environments (Stanley-Becker, Nix2023). The pictures were very obviously fake; to see them at all, though, was to have a strong emotional reaction to them (Garber 2023).

In May 2023 a fake image purporting to show an explosion near the Pentagon was shared by multiple verified Twitter accounts, causing confusion and leading to a brief dip in the stock market. Local officials later confirmed no such incident had occurred. The image, which bears all the hallmarks of being generated by AI, was shared by numerous verified accounts with blue check marks, including one that falsely claimed it was associated with Bloomberg News (O’Sullivan, Passantino 2023).

The MUAI as separate episodes or targeted campaigns of socio-political destabilization, may occur both in the most developed and powerful countries with the highest level of development of AI technologies, such as the United States, and in countries with a lower level of development and implementation of these technologies. But if Big Tech behaves more cautiously in large and relatively economically, militarily and technologically developed countries, then small and poor countries as a whole are more vulnerable. So, while the leading social networks have started to engage in content moderation around the world, they appear comparatively inactive on the African continent. In 2019, upon request from governments, courts, civil society organizations, and members of the Facebook community’(Facebook), Facebook removed content from its platform in several thousands of instances in countries like Pakistan (N=7’960), Mexico (N=6’946), Russia (N=2’958), or Germany (N = 2’182), but it hardly removed any content in Africa. In fact, Morocco had the highest number of content removals, with N = 6 Twitter’s transparency reports suggest similar figures for African countries (Garbe, Selvik & Lemaire 2023). Rather, this approach is associated with the disadvantage of content moderation, where there are relatively few solvent people, and the authorities are not very demanding in their relations with social networks.

It is necessary to understand well why there are problems with biased AI, discriminatory AI. As noted Safiya Noble, an internet studies scholar and professor of gender studies and African American studies at UCLA, the author of the book, "Algorithms of Oppression: How Search Engines Reinforce Racism," (Noble Umoja 2018) over-policing and the over-arresting happens in Black and Latino communities of the USA. “That's just a fact. So, if that is a main factor in whether you are likely to commit - in predicting whether you're likely to commit another crime because lots of people in the zip code you live in have been arrested…, then you are more likely to be considered a risk. That has nothing to do with you. That has to do with the history of structural racism in policing in the United States” (Scott 2023).

More broadly, the search and analytical AI systems created in the West, to one degree or another, inevitably bear the imprint of the social diseases of the society where they were created. The information content for AI models is got primarily from the most accessible English-language data set, which reduces and distorts the processes of learning and self-learning of models and leads to their gaps of ignorance and false conclusions regarding non-Western countries in general and BRICS in particular. Moreover, ideologically and politically motivated adjustments are sometimes made to the learning and functioning of models, examples of which are given in various chapters of the current report. Machine

learning data is also clogged with low-quality synthetic texts, images, videos created by other intelligent systems, including those compromised by various malicious actors. All this strengthens the neocolonial model of the digital space, which is dangerous not only for the population of non-Western countries, but also for Western ones, since distorted information contributes to dangerous cognitive and psychological deformations in the perception of the world by different social groups, primarily among young people in the West.

The destruction of personal, group, and social consciousness with the growing use of AI technologies is a key aspect of malicious influence, since it opens the way to the domination of antisocial actors in the desired form and for the desired purposes (or additionally supports already existing domination). And this is not someone's centralized plan, a conspiracy, but the process of high-tech "devouring" of an *already* sick social organism by parasites, who, feuding (sometimes fatally) with each other, lead society to a catastrophe, which until a certain moment is not fully realized and is not felt, but will eventually affect everyone, even its temporary beneficiaries.

Global MUAI Threats and the BRICS High-Tech Response

There is a growing danger of AI being used to destabilize economies, political situations, and international relations through targeted, high-tech, psychological impacts on people's consciousness. Meanwhile, crisis phenomena have been rapidly increasing in frequency, number, and severity worldwide. In 2020, the Doomsday Clock was set to 100 seconds to midnight for the first time in history since the clock was created in 1947, and remains unchanged in 2021 – 2022. And there is no need to explain here why the growth in the world's billionaires' fortunes from 8 to 13 trillion dollars in the crisis year of 2020 when COVID-19 pandemic started (Dolan, Wang & Peterson-Withorn, 2021)—against the backdrop of record economic decline in recent decades, hundreds of millions of newly unemployed people, and, according to the UN, the growth in the number of hungry people in the world from 690 million in 2019 (Kretchmer, 2020) to 811 million in 2020 (World Health Organization, 2021)—did not contribute to solving these and other acute problems of our time.

In January 2023, the Doomsday Clock was moved to 90 seconds to midnight, and it has remained this close in 2024 (O'Neill 2024). "In 2023, Earth experienced its hottest year on record, and massive floods, wildfires, and other climate-related disasters affected millions of people around the world. Meanwhile, rapid and worrisome developments in the life sciences and other disruptive technologies accelerated, while governments made only feeble efforts to control them" (Mecklin 2024) concluded the members of the Science and Security Board of the *Bulletin of the Atomic Scientists*. Economic problems, military conflicts, the degradation of democratic institutions, social polarization, internal political and interstate conflicts, all under the conditions of rapid AI development, create extremely favorable ground for the MUAI.

In the context of the growing global crisis, the leading Western and Chinese AI scientists have issued a stark warning that tackling risks around the powerful technology requires global co-operation similar to the cold war effort to avoid nuclear conflict. A group of renowned international experts met in Beijing in March 2024, where they identified "red lines" on the development of AI, including around the making of bioweapons and launching cyberattacks. The academics warned that a joint approach to AI safety was needed to stop "catastrophic or even existential risks to humanity within our lifetimes". The experts also discussed threats regarding the development of "artificial general intelligence", or AI systems that are equal to or superior to humans (Criddle, Olcott 2024).

Such contacts are extremely important, but in the absence of radical changes in Western countries (for example, as a result of anti-oligarchic transformations), it is unlikely that the ruling circles of these countries will abandon their course towards global domination, this also applies to the field of

AI, as an increasingly important tool for their technological, economic and military dominance. The response to sanctions and military-political pressure from the West is the growing desire, both at the level of individual countries and at the level of independent international associations, to pursue a nationally oriented policy.

As part of the course on technological sovereignty, several BRICS countries are actively developing a base for the production of semiconductors, without which the successful development of the AI industry is impossible. Russia and China are doing this in the face of Western sanctions.

The microelectronics industry was actually destroyed during the reforms of the last President of the USSR Mikhail Gorbachev and the disastrous privatization of the 1990s. In 1962, industrial production of microchips started in the USSR almost simultaneously with the United States, later the USSR was among two leaders in the field and now Russia is trying to make up for lost decades.

In 2023, Chinese chip manufacturers were forecasted to grow capacity by 12%. This growth is projected to accelerate to 13% in 2024 and do the main contribution in the rise of chip production globally. A total of 18 new fabs are expected to start operations in China in 2024 (globally 11 in 2023 and 42 in 2024). China is in the process of raising more than $27 billion for its largest chip fund to date, accelerating the development of cutting-edge technologies to counter a US campaign to thwart its rise (Cao, Gao 2024).

Some other BRICS countries, without breaking with the West, but taking into account the risks of the future, are striving to achieve greater autonomy in the field of conductors. In February 2024 India’s government has approved $15.2 billion worth of investments in semiconductor fabrication plants, including a Tata Group proposal to build the country’s first major chipmaking facility (Phartiyal 2024).

The new BRICS members, primarily Saudi Arabia and the UAE, have very ambitious plans for the development of the semiconductor industry. Saudi Arabia’s Alat which is a sustainable technology manufacturing company backed by Saudi Arabia’s Public Investment Fund (PIF), announced in 2024 four partnerships with global technology companies – Softbank Group, Carrier Corporation, Dahua Technology and Tahakom – to boost the country’s technology sector. Alat will initially focus on the manufacturing of products in 34 categories in seven business units, including semiconductors, smart devices, smart buildings, smart appliances, smart health, etc. (Finance Middle East 2024).

In 2011, as part of the UAE’s effort to diversify its economy away from energy production, Abu Dhabi's Mubadala Investment Company, acquired Advanced Technology Investment Co., the parent company of the California-based semiconductor manufacturer GF. GF is one of the top five chipmakers globally, producing advanced semiconductors for companies such as Apple, Intel, and Amazon. It is the third-largest semiconductor producer, behind only TSMC and Samsung. In 2021, GF announced plans to expand by building a new $4 billion manufacturing plant in Singapore (Soliman 2022). In March 2024 MGX, the newly-established Abu Dhabi technology investment company, is reportedly in talks to invest billions of dollars in Sam Altman’s OpenAI CEO visionary plan to build a network of AI chip factories worldwide. This potential partnership could change the global AI landscape and position Abu Dhabi one of key players in the development and deployment of advanced AI technologies. MGX plans to have $100 billion in assets under management. This plan includes AI infrastructure, semiconductors, and core AI technologies, with the aim of driving innovation and fostering economic growth on a global scale (Abu Dhabi Startups 2024).

Thus, in the field of semiconductors (as in many other areas), the foundations are being laid for the confident development of the AI industry in the BRICS countries. Further into the future, the West's overall technological superiority over the BRICS countries will decrease, which will allow the association to more effectively protect its information space with the help of AI technologies.

Dr Greg Simons from Turiba University considers “the current state of international affairs is that the old order has not disappeared, and the new order is in the process of consolidating… BRICS will be a challenger to the geo-economic institutional structure…They need to offer an alternative vision of global relations and interactions that is superior to the existing model, which can be achieved through a viable and resilient geo-economic vision…” (Simons, 2024) . Such a new vision cannot but include as an integral part the achievement of technological sovereignty on the basis of a new technological order, the key component of which is the development of AI technologies in BRICS.

One of new proposals in this direction suggests that a unified internet service for BRICS countries could diminish the technological dominance of the US. Dmitry Gusev, the Deputy Chairman of the State Duma Control Committee has proposed BRICS to develop an alternative internet service not dependent on US communications. Gusev suggests in the proposal that creating an internet service solely for BRICS countries will weaken the US control of the global news narrative. The official submitted a request to work on creating *“a single inclusive BRICS+ cyberspace”* to Maksut Shadaev, the head of Russia's Ministry of Digital Development, Communications and Mass Media (CGS 2023).

Back in 2020 and later, the author of this Introduction suggested in different publications (Bazarkina, Pashentsev 2020, Pashentsev, Bazarkina 2023) the idea of creating a BRICS communication network (based on intelligent text recognition) with the function of high-quality online translation of leading media and research journals of BRICS into the language of the addressee. This idea was not quite possible at that time, but now with rapid progress underway in machine translation can be implemented rather soon. This will significantly improve mutual understanding between BRICS countries, and give an alternative, which will be independent from the United States and will save residents of the BRICS countries with their numerous languages from communicating mainly through English and Big Tech tools. The already existing wide range of portable voice translator devices can facilitate understanding on tourist trips, business negotiations in the BRICS countries (however, the universal translator from the Star Trek series, alas, still remains the subject of dreams), as well as other AI-based communication tools.

Of course, the author does not idealize the possibilities of developing AI technologies in the BRICS countries, insofar as acute social and political contradictions exist in the countries of the association, it is also impossible to exclude the deformation of the work of local intelligent systems in favor of internal and external malicious actors. But the diversity of interests and cultures of the member countries of the association, the absence of a monopoly center of power claiming world domination, turns the BRICS into a broad international community that, without being a military bloc, can become a real alternative to the hegemony of the West. The latter is based on the largest expansionist military-political alliance in history, NATO, which, with the growing degradation of civil society institutions, the intensification of intra-elite contradictions, and the strengthening of corporate rulers of Big Tech, poses a threat to all mankind, especially given the potential of AI technologies.

The population of the BRICS countries does not want to break with the West at all, but has no intention of being subjected to both traditional and AI powered tools of Western elites’ propaganda any more. In this desire, the BRICS people have similar aspirations to the citizens of the United States and the EU, who do not have confidence in mainstream media. The survey, released by Gallup and the Knight Foundation in February 2023, goes beyond others that have shown a low level of trust in the media to the startling point where many believe there is an intent to deceive. Asked whether they agreed with the statement that national news organizations do not intend to mislead, 50% said they disagreed. Only 25% agreed, the study found (Bauder 2023).

This report is only the first attempt at systemic analysis of MUAI and challenges to psychological security after the expansion of the BRICS. It does not yet cover all forms and methods of MUAI against psychological security, but it already allows to identify some general trends in their development, and

on this basis, it is better to understand the nature, scope and consequences of the activities of various malicious actors in the countries of the association.

As research coordinator of this report, I would like to express my gratitude to its authors, who jointly tried to present a three-level vision of MUAI threats to psychological security. I would like to express special gratitude to Greg Simons, Associate Professor at Turiba University, for his help in editing the Introduction and Conclusion, as well as the thematic chapter on Iran of this report; the chapter on China, to Peter Mantello, professor /AI Researcher at Ritsumeikan Asia Pacific University; the chapter on India to Arvind Gupta, the Head and Co-Founder of Digital India Foundation and Aakash Guglani, a policy associate at the Digital India Foundation; chapter on Egypt to Deborah Sellnow-Richmond, Associate Professor at the Department of Applied Communication Studies at Southern Illinois University; chapter on South Africa, to Sergei A. Samoilenko, an Assistant Professor at the Department of Communication at George Mason University.

Evgeny Pashentsev April 8, 2024

References

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

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

Abu Dhabi Startups (2024) Abu Dhabi's MGX in Talks to Invest Billions in Sam Altman's Chip Venture. https://www.abudhabistartup.com/startup-news/2024/03/abu-dhabis-mgx-in-talks-to-invest-billions-in-sam-altmans-chip-venture/. Accessed 29 Mar 2024

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

Arkose Labs (2023) Breaking (Bad) Bots: Bot Abuse Analysis and Other Fraud Benchmarks

Bank of America. Merrill Lynch (2015) Creative disruption

Barishpolets VA (2013) Informatsionno-psikhologicheskaya bezopasnost': osnovnye polozheniya [Informational and psychological security: main principles]. Radioelektronika. Nanosistemy. Informatsionnye tehnologii [Radionics. Nanosystems. Information Technology], Vol. 2. Pp. 62-104.

Barishpolets VA (ed.) (2012) Osnovy informatsionno-psihkologicheskoy bezopasnosti [Fundamentals of the psychological security]. Moscow, Znanie.

Bauder D, The Associated Press (2023) Trust in media is so low that half of Americans now believe that news organizations deliberately mislead them. In: Fortune. https://fortune.com/2023/02/15/trust-in-media-low-misinform-mislead-biased-republicans-democrats-poll-gallup/ Accessed 29 Mar 2024

Bazarkina D, Mikhalevich EA, Pashentsev E, Matyashova D (2023) The Threats and Current Practices of Malicious Use of Artificial Intelligence in Psychological Security in China. In: Pashentsev, E. (eds) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham. 2023.

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

Bazarkina DY Matyashova DO (2022) 'Smart' Psychological Operations in Social Media: Security Challenges in China and Germany. In: ECSM, 9th European Conference on Social Media proceedings. Reading. P. 14-20.

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

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

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

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

Cao D, Gao Y (2024) China Readies $27 Billion Chip Fund to Counter Growing US Curbs In: Bloomberg. https://www.bloomberg.com/news/articles/2024-03-08/china-readies-27-billion-chip-fund-to-counter-growing-us-curbs. Accessed 29 March 2024

CGS (2023) BRICS should create their own internet – MP In: ChinaGoSmart. https://chinagosmart.com/brics-should-create-their-own-internet-mp. Accessed 29 March 2024

CGTH (2023) Full text: Xi Jinping's speech at the 15th BRICS Summit. https://news.cgtn.com/news/2023-08-23/Full-text-Xi-Jinping-s-speech-at-the-15th-BRICS-Summit-1mvxFMvuFLW/index.html. Accessed 29 March 2024

Collins C, Fitzgerald J, Flannery H, Ocampo O, Paslaski S, Thomhave K (2021) Silver spoon oligarchs: How America's 50 Largest Inherited-Wealth Dynasties Accelerate Inequality. In: Institute for Policy Studies. https://ips-dc.org/report-americas-wealth-dynasties-2021/. Accessed 29 March 2024

Criddle C, Olcott E (2024) Chinese and western scientists identify 'red lines' on AI risks. In: Financial Times. https://www.ft.com/content/375f4e2d-1f72-49c8-b212-0ab2a173b8cb. Accessed 29 March 2024

Dolan K, Wang J, Peterson-Withorn C (2021) The Forbes World's Billionaires list. In: Forbes. https://www.forbes.com/billionaires/. Accessed 29 March 2024

Egan M (2023) Exclusive: 42% of CEOs say AI could destroy humanity in five to ten years. https://edition.cnn.com/2023/06/14/business/artificial-intelligence-ceos-warning/index.html. Accessed 29 March 2024

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

Finance Middle East (2024) Saudi PIF company Alat to invest $100 billion in the country's tech sector. https://www.financemiddleeast.com/saudi-pif-company-alat-to-invest-100-billion-in-the-countrys-tech-sector/. Accessed 29 March 2024

Forbes (2021) The World's Real-Time Billionaires. https://www.forbes.com/real-time-billionaires/#1d7a52b83d78. Accessed 29 March 2024

Forbes India (2022) Deepfake Democracy: South Korean Presidential Race Candidate Goes Virtual For Votes. https://www.forbesindia.com/article/lifes/deepfake-democracy-south-korean-presidential-race-candidate-goes-virtual-for-votes/73715/1. Accessed 29 March 2024

France 24 (2022) Deepfake democracy: South Korean candidate goes virtual for votes. https://www.france24.com/en/live-news/20220214-deepfake-democracy-south-korean-candidate-goes-virtual-for-votes. Accessed 29 March 2024

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

Fried I (2022) Inside tech companies' unprecedented move to suspend sales in Russia. In: Axios. https://www.axios.com/2022/03/07/tech-companies-suspend-sales-russia. Accessed 29 March 2024

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

Garber M (2023) The Trump AI Deepfakes Had an Unintended Side Effect. In: The Atlantic. https://www.theatlantic.com/culture/archive/2023/03/fake-trump-arrest-images-ai-generated-deepfakes/673510/. Accessed 29 March 2024

Gilens M, Page IB (2014) Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens. In: Cambridge University Press. https://www.cambridge.org/core/journals/perspectives-on-politics/article/testing-theories-of-american-politics-elites-interest-groups-and-average-citizens/62327F513959D0A304D4893B382B992B. Accessed 29 March 2024

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

Global Times (2022) From commercial satellites to social media, Western tech companies are deeply involved in the Russia-Ukraine conflict. https://www.tellerreport.com/news/2022-11-02-from-commercial-satellites-to-social-media--western-tech-companies-are-deeply-involved-in-the-russia-ukraine-conflict.HJSuXB1Bo.html. Accessed 29 March 2024.

Grachev GV (1998) Informatsionno-psikhologicheskaya bezopasnost' lichnosti: sostoyanie i vozmozhnosti psikhologichesko? zastchity [Information and psychological security of the person: the state and possibilities of psychological protection]. Moscow, RAGS.

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

Hatzius J, Briggs J, Kodnani D, Pierdomenico G. (2023) The Potentially Large Effects of Artificial Intelligence on Economic Growth. In: Goldman Sachs. https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf. Accessed 29 March 2024

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

Kak A, Myers West S, Whittaker M (2023) Make no mistake – AI is owned by Big Tech. In: MIT Technology Review. https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech/. Accessed 29 March 2024

Kretchmer H. (2020) Global hunger fell for decades, but it's rising again. In: World Economic Forum. https://www.weforum.org/agenda/2020/07/global-hunger-rising-food-agriculture-organization-report/. Accessed 29 March 2024

La Monica P (2021) The race to $3 trillion: Big Tech keeps getting bigger. In: CNN. https://edition.cnn.com/2021/11/07/investing/stocks-week-ahead/index.html. Accessed 29 March 2024

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

Malayil J (2024) OpenAI backed 1X's humanoid robots showcase an advanced neural network. In: Interesting Engineering. https://interestingengineering.com/innovation/openai-backed-1xs-humanoid-robots-showcase-an-advanced-neural-network. Accessed 29 March 2024

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

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

McKinsey Global Institute (2017) A Future that Works: Automation, Employment, and Productivity. January 2017 Executive Summary. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works-Executive-summary.ashx?trk=public_post_comment-text. Accessed 29 Jan 2024

Mecklin J (2024) A moment of historic danger: It is still 90 seconds to midnight. In: The Bulletin. https://thebulletin.org/doomsday-clock/current-time/. Accessed 29 Jan 2024

Metz R, Mcbride S, Bloomberg (2023) Elon Musk unveils A.I. startup with execs from DeepMind and Microsoft, with goal to 'understand the true nature of the universe.' In: Fortune. https://fortune.com/2023/07/12/elon-musk-ai-startup-xai-deepmind-microsoft-executives/. Accessed 29 Jan 2024

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

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

NS Business (2022) Oracle, SAP, and Accenture suspend business operations in Russia. https://www.ns-businesshub.com/technology/oracle-sap-accenture-suspend-russian-operations-ukraine/. Accessed 29 March 2024

O'Neill A (2024) Minutes to midnight on the Doomsday Clock every year from 1947 to 2024. In: Statista. https://www.statista.com/statistics/1072256/doomsday-clock-development/. Accessed 29 Jan 2024

O'Sullivan D, Passantino J (2023) 'Verified' Twitter accounts share fake image of 'explosion' near Pentagon, causing confusion. In: CNN Business. https://edition.cnn.com/2023/05/22/tech/twitter-fake-image-pentagon-explosion/index.html. Accessed 29 Jan 2024

Osborne C (2023) WormGPT: What to know about ChatGPT's malicious cousin. In: ZD net. https://www.zdnet.com/article/wormgpt-what-to-know-about-chatgpts-malicious-cousin/. Accessed 29 Jan 2024

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

Pashentsev E (2022) U.S.: On the Way to Right-Wing Coup and Civil War? In: Russian International Affairs Council. https://russiancouncil.ru/en/analytics-and-comments/analytics/u-s-on-the-way-to-right-wing-coup-and-civil-war/. Accessed 29 Jan 2024

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

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

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

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

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

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

PTI (2023) PM Modi calls for global framework for ethical use of AI. In: The Economic Times. https://economictimes.indiatimes.com/news/india/pm-modi-calls-for-global-framework-for-ethical-use-of-ai/articleshow/105939251.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst . Accessed 29 Jan 2024

Rao P (2023) Visualizing the $105 Trillion World Economy in One Chart. In: Visual Capitalist. https://www.visualcapitalist.com/visualizing-the-105-trillion-world-economy-in-one-chart/. Accessed 29 Jan 2024

Roshhin SK, Sosnin VA (1995) Psikhologicheskaya bezopasnost': novy? podhod k bezopasnosti cheloveka, obstchestva i gosudarstva [Psychological security: a new approach to human, social and state security]. In: *Rossiski monitor* [Russian Monitor].

Saad L (2023) Historically Low Faith in U.S. Institutions Continues. In: Gallup. https://news.gallup.com/poll/508169/historically-low-faith-institutions-continues.aspx. Accessed 29 Jan 2024

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

Scott B, Woods J, Chang A (2023) How AI could perpetuate racism, sexism and other biases in society. In: NPR. https://www.npr.org/2023/07/19/1188739764/how-ai-could-perpetuate-racism-sexism-and-other-biases-in-society. Accessed 29 Jan 2024

Simons G (2024) BRICS and the Geo-Economic Aspects of Engineering a New Global Order. In: TPQ. http://turkishpolicy.com/article/1245/brics-and-the-geo-economic-aspects-of-engineering-a-new-global-order. Accessed 29 Jan 2024

Soliman M (2022) Strategic Start-Ups: The UAE Is Betting Big on Semiconductors. In: The National Interest. https://nationalinterest.org/blog/techland-when-great-power-competition-meets-digital-world/strategic-start-ups-uae-betting-big. Accessed 29 Jan 2024

Stanley-Becker I, Nix N (2023) Fake images of Trump arrest show 'giant step' for AI's disruptive power. In: The Washington Post. https://www.washingtonpost.com/politics/2023/03/22/trump-arrest-deepfakes/. Accessed 29 Jan 2024

Statista (2021) S&P 500: largest companies by market cap 2021. https://www.statista.com/statistics/1181188/sandp500-largest-companies-market-cap/. Accessed 29 Jan 2024

Stieber Z (2022) Over 50 Biden Administration Employees, 12 US Agencies Involved in Social Media Censorship Push: Documents. In: The Epoch Times. https://www.theepochtimes.com/over-50-biden-administration-employees-12-us-agencies-involved-in-social-media-censorship-push-documents_4704349.html?welcomeuser=1. Accessed 29 Jan 2024

Subhra Dutta T (2023) FraudGPT: New Black Hat AI Tool Launched by Cybercriminals. In: Cyber Security News. https://cybersecuritynews.com/fraudgpt-new-black-hat-ai-tool/. Accessed 29 Jan 2024

Sumsub Research (2023) Global Deepfake Incidents Surge Tenfold from 2022 to 2023. https://sumsub.com/newsroom/sumsub-research-global-deepfake-incidents-surge-tenfold-from-2022-to-2023/. Accessed 29 March 2024

The Times of India (2022) Deepfake democracy: South Korean candidate goes virtual for votes. https://timesofindia.indiatimes.com/world/rest-of-world/deepfake-democracy-south-korean-candidate-goes-virtual-for-votes/articleshow/89556568.cms. Accessed 29 Jan 2024

Think BRICS (2023) BRICS Nations Map an AI Future on a Parallel Digital Track. https://thinkbrics.substack.com/p/brics-nations-map-an-ai-future-on. Accessed 29 Jan 2024

Trend Micro, UNICRI and Europol (2020) Malicious Uses and Abuses of Artificial Intelligence. Trend Micro Research

TV BRICS (2024) Vladimir Putin announces that 30 countries are ready to join BRICS. https://tvbrics.com/en/news/vladimir-putin-announces-that-30-countries-are-ready-to-join-brics/?ysclid=lu71w68ybm40618341. Accessed 29 Jan 2024

Tyson A, Kikuchi E (2023) Growing public concern about the role of artificial intelligence in daily life. In: Pew Research Center. https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/. Accessed 29 Jan 2024

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

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

Vastmindz (2022) South Korea's presidential deepfake. https://vastmindz.com/south-koreas-presidential-deepfake/. Accessed 29 Jan 2024

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

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

World Economic Forum (2024) The Global Risks Report

World Health Organization (2021). UN report: Pandemic year marked by spike in world hunger. https://www.who.int/news/item/12-07-2021-un-report-pandemic-year-marked-by-spike-in-world-hunger. Accessed 29 Jan 2024

WormGPT V3.0 (2024) https://flowgpt.com/p/wormgpt-v30. Accessed 29 Jan 2024

The Malicious Use of AI: Challenges to Psychological Security in the Arab Republic of Egypt

Evgeny PASHENTSEV, Vladilena CHEBYKINA, Julia SHEMETOVA

Introduction

In 2022, Egypt ranked second in Africa after Mauritius, according to the report on the readiness of world governments to implement AI technologies. In comparison with the report for 2019, where Egypt ranked eighth among African countries and 111th out of 194 countries in the world considered in the report, there is clear progress. The Egypt Human Development Report 2021 revealed its progress by 55 places in the "Government Readiness for Artificial Intelligence" index. According to the World Knowledge Index, Egypt advanced from 72nd out of 138 countries in 2020 to 53rd out of 154 countries in 2021(Draya Egypt 2023). This positive trend indicates the country's serious efforts in the field of technological development and contributes to the creation of a favorable environment for innovation and economic development.

The latest technologies are also used to build smart cities and transform existing cities into smart ones in accordance with international standards. For example, Hawa Dawa (AI technology aimed at combating environmental problems) in Egypt combines sensor technologies of the Internet of Things and satellite images with machine learning algorithms to collect and analyze high-quality data on outdoor air pollution (Sayed M K 2018).

The Government of the country is actively developing its legislation in the field of AI and digital development. It is represented by two main documents: Egypt's Digital Development Strategy and The National Strategy for Artificial Intelligence. The first phase of the strategy will end in May 2024, and it aims to exploit artificial intelligence technologies to support the achievement of Egypt's sustainable development goals. The duration of the second phase of the strategy is 3 years (Business Today Egypt 2023). According to Amr Talaat, Communications and Information Technology Minister of Egypt, the second phase of its national AI strategy will commence in the second quarter of 2024 and encompass several key economic sectors: the government will introduce initiatives across governance, human resources, technology, information infrastructure, data, and environment (Pessarlay W 2024). The AI market in Egypt is projected to reach US$785.20m in 2024. The market size is expected to show an annual growth rate of 17.18%, resulting in a market volume of US$2,033.00m by 2030 (Statista Egypt 2024).

The presidential elections that took place in December 2023 confirmed President El-Sisi with nearly 90% of votes. But serious economic challenges, high youth unemployment, low purchasing power and currency depreciation, as well as potential misalignment with other Arab countries over the relationship with Israel and the protection of Palestinians, may trigger revolts and institutional changes (Allianz 2024). The rapid growth in AI industries, along with Egypt's complex socio-economic and political problems, the risks of further aggravation of the situation in the Middle East, create favorable conditions for MUAI's growth in this country.

The First Level of Threats of MUAI Against Psychological Security

The main threat of the first level is the risks of malicious interpretation of AI implementation into different areas of life. The fear of Egyptians losing their jobs due to robotization, and automation of

processes is most often associated with employee concerns about the possible replacement of human labor with machines (this has a lot in common with other countries). Kaspersky Corporation's research is an important indicator for understanding the current state of the impact of technological development on labor relations. According to one such report: "About half (44%) of employees in Egypt are afraid of losing their jobs because of robots, and one in four employees (25%) reported having heard about cybersecurity incidents with robots or automated systems in their company" (Daily News Egypt 2023a). Consequently, statistics show that there are widespread concerns among employees of Egyptian companies about the impact of the introduction of AI technologies on their employment and personal safety. Companies and the Egyptian government may face challenges related to the need to ensure job stability and staff retraining programs. As for improving IT competencies, the Kaspersky survey showed that "41% of employees in Egypt feel the need to improve their digital skills, and 33% are afraid of losing their jobs due to a lack of IT competencies. Some believe that this may happen in the next 5 years (15%), others suggest that it may happen at some point later (18%). Only 40% are confident that they are not in danger of losing their jobs due to poor IT knowledge" (Daily News Egypt 2023b). A variety of malicious actors, both external and internal, can actively play on the risks of rising mass unemployment or lower wages as a result of the introduction of AI technologies in the future: from religious fanatics to supporters of destabilizing the government for selfish interests, which requires special attention to the timely occurrence of employment problems with the introduction of AI technologies (quantitatively and qualitatively the growth of such problems seems inevitable all over the world).

In Egypt (as in other countries), predictions about the future relationship between man and machine are extremely gloomy. In particular, Mohammed Javdat, Google's Former Egyptian AI Expert, warned that one day AI may start treating people like “scum” and create its own “killing machines” (Blunt P 2023a). Javdat warns that current AI-based language learning models read the negative information we place in virtual space, which in the future may allow machines to think of humanity as something negative and evil that poses a threat. Such statements about possible risks, which must be taken seriously with their appropriate targeted promotion, can sow panic among the population and cause people to have a negative attitude towards the further development and use of AI in everyday life.

In 2019, the article “Egypt Is Using Apps to Track and Target Its Citizens, Report Says” was published in The New York Times based on research conducted by cybersecurity experts, according to which the Egyptian government may be associated with attackers who conducted a series of cyber-attacks on a number of Egyptian journalists, opposition activists, human rights defenders, etc. (Bergman R, Walsh D 2019). The actions themselves originated back in 2016. The specialists of the IT security company Check Point Technologies discovered that hackers used the official Google Play Store to distribute programs that collected geolocation information, email data, calls, etc. One of the necessary requirements for installation was to provide access to the user's call history and contacts. Check Point researchers have found that the program can be used to benefit the government. “Coordinates embedded in one of the HTML phishing pages pointed to a government building in Cairo. The registrant of a domain used by the attackers is listed as MCIT, which the researchers said could be Egypt’s Ministry of Communications and Information Technology (MCIT)” (Lyngaas S 2019). However, it is highly likely that one of the attackers could have framed the Egyptian government structure for the subsequent anti-government influence campaign. Thus, the unproven use of AI technologies at the second level of threats to psychological security became the reason for creating, without sufficient justification, an image of the antisocial use of AI by the Egyptian government, i.e. the first level of threats, which is indicated by the title of the article in the newspaper, and which leaves no doubt about the "guilt" of the Egyptian authorities.

Thus, it can be concluded that MUAI at the first level is possible in Egypt already today, which requires appropriate reflection and advanced analysis, both at the level of the expert community and at the level of public administration.

The Second Level of Threats of MUAI Against Psychological Security

Phishing is a widespread type of fraud in Egypt. According to a Kaspersky study in the Middle East and Africa region, phishing is used in the field of courier deliveries. The attackers send letters to their victims containing a link to pay for delivery, because of which the goods cannot be delivered. By clicking on the link, they go to a website posing as the official website of the delivery service, enter the details of a bank card, from which funds are debited by fraudsters (Daily News Egypt 2023d). Another Kaspersky study for 2022 on electronic payments showed that “57% of users in Egypt encountered phishing attempts when using online banking services or digital wallet services. It also said that 54% of users encountered fake websites and 57% of phishing attempts occurred through text messages or phone calls using social engineering methods” (Daily News Egypt 2022e). Thus, not only the government, but also the banking sector and companies connected in one way or another with electronic payments, should raise awareness among citizens on the topic of digital security.

The main aspect of online fraud is social engineering, which helps the attacker gain the trust of the victim and force it to act recklessly. For example, in August 2022, attackers in Egypt launched an online platform and promised customers "great financial benefits" due to commission from cryptocurrency mining and trading services. 29 people were arrested, almost half of whom were foreign nationals who stole more than $600,000 using the HoggPool fraud network (Helou E A 2023). This example shows that cybercriminals are increasingly diligent in using psychological methods to manipulate the minds of their victims.

Amin Hasbini, head of Kaspersky's global research and analysis group for the Middle East, Turkey and Africa and specializing in cybersecurity solutions and services, said that Kaspersky tracked and stopped about 13 million electronic attacks in Egypt during the first quarter of 2023 (Daily News Egypt 2023c). Hasbini said in an interview with Daily News Egypt that the number of attacks targeting bank accounts and customer data has increased by 186% compared to 2022. It was emphasized that the number of hacker attacks on the information system in the retail banking sector in Egypt is actively growing. At the same time, phishing attacks via e-mail and SMS messages have intensified. About 75,000 users in Egypt were subjected to phishing attacks in the first quarter of 2022, and according to statistics, from 17 to 70% of users receiving these fraudulent emails click on links and fall into an electronic trap (Daily News Egypt 2023c).

Concerns that some AI systems may pose a serious threat to national and international security have increased since the advent of chatbots. According to a recent report by Group-IB, countries in the Middle East and North Africa are most susceptible to cyber-attacks aimed at stealing accounts, crypto wallets, browser histories and other confidential information. Egypt occupies a leading position in the region in terms of the number of stolen ChatGPT accounts – about 4,500 in the period from June 2022 to May 2023. The hacked data includes login credentials and search queries (Ahram Online 2023b). Even though the use of AI bots in Egypt is not as widespread when compared to other countries, the country has a predominantly young population – “60% of the Egyptian population are citizens from 10 to 49 years old, and more than 69.4 million people use the mobile Internet” (Salah A 2023). This suggests that the popularity of this technology in Egyptian society will only grow in the coming years.

The presentation by Dr. Mohamed El-Guindy (Egyptian researcher, cybersecurity expert, specialist and cybercrime consultant for international organizations) at the International Consortium called Global AI Ethics Network for Social Good (GAIEN4SG) on the topic "Malicious use of artificial

intelligence: legal and ethical consequences" proves that in the near future, our face may become a trigger for the introduction of malware by detractors. Upon visual identification of the target, the corresponding malware will be launched (ISSA Egypt 2022).

There are two hacker groups in Egypt: Horus Group and Anubis. Their goal is to obtain classified information about geopolitical rivals. For example, their organization is known to have contributed to cyber attacks on Ethiopia, which were provoked in connection with the GERD (Great Ethiopian Renaissance Dam) under construction since 2012 (Munawer Q 2020). The activities of these organizations indicate that cybercriminals may be involved in espionage and cyber espionage in the interests of certain States or organizations. Of course, in the process of cheapening and spreading technologies, the latter will be used more and more often by hackers.

Thus, the problem of cybercrime with the growing use of AI technologies is acute in Egypt, suggesting the further growth of MUAI in this country.

The Third Level of Threats of MUAI Against Psychological Security

At the third level of psychological security threats in Egypt, there concern regarding the malicious use of deepfakes. In 2023, KnowBe4 conducted a study among 800 employees aged 18 to 54 living in Egypt, South African countries, and Kenya. Direct communication via email and video calls was organized, and the main point of this study was to communicate not with a human, but with a special bot created using AI technologies, namely deepfake technology. According to the results of this study, 74% of employees failed to recognize that they were interacting with a bot, not a real person (Shankar A 2023). This reflects how sophisticated deepfake technology has become, making it quite difficult and sometimes impossible for most people to detect fakes online. The study also explicitly points out that the lack of citizen awareness of the proliferation of deepfakes in African countries is acute, putting millions of people at risk. According to Anna Collard, SVP Content Strategy & Evangelist at KnowBe4 Africa, the "... deepfake platforms are capable of creating civil and societal unrest when used to spread mis- or dis-information in political and election campaigns, and remain a dangerous element in modern digital society" (Shankar A 2023). Consequently, the Egyptian government needs to pay increased attention to combating the threat posed by deepfake technology and its potential use for spreading misinformation not only in political and electoral campaigns but also in business campaigns, as well as to protect ordinary citizens at the legislative level and through educational initiatives on recognizing deepfake content they may encounter online.

Currently, another area experiencing a crisis of trust related to the spread of deepfakes is the media sphere. Imad Eddine Adib, an Egyptian journalist and businessman, speaking at the Arab Media Forum in Dubai in September 2023, warned that the future of the media industry is not secure. According to him, the world has already entered an era of misleading news that spreads at the speed of light. Adib highlighted an example of a deepfake video sent to him by friends that falsely claimed former US President Donald Trump had been given a new heart belonging to a Muslim man and had converted to Islam (Al-Faour N 2023). The problem lies in people’s tendency to believe what they can see with their own eyes, even if the content is not authentic.

In early 2022, the Egyptian House of Fatwas (Dar Al-Ifta) issued a statement that it is unacceptable to use AI technology to create fake videos or audio recordings of people unrelated to them. "The fabrication of these clips using deepfake with the intent of harming others is forbidden according to Prophet Mohamed's (PBUH) quote 'no harm to oneself or others,' Dar Al-Ifta said, adding that Islam prohibits intimidating others, even for entertainment" (Ahram Online 2022a). This statement reflects the attitude of a reputable Egyptian religious organization towards technologies capable of causing harm to individuals or society, while Islam does not advocate for restricting the development of

information technology, but it does indicate that moral constraints should be paramount. Furthermore, "Dar Al-Ifta also pointed out that the dissemination of misleading information is criminalized by Law 175/2018 covering information technology crimes" (Ahram Online 2022a).

According to Mina Henein, a Ph.D. holder, researcher and lecturer at the Australian National University's School of Cybernetics, the main aspects hindering the integration of chatbots into the daily life of Egyptians are: language barrier, digital illiteracy of the population, legislative framework and cultural norms of the country (Salah A 2023). However, Egypt is now actively conducting educational activities to familiarize the population with the positive and negative consequences of innovative technologies. For example, in March 2023, Egypt's Information and Decision Support Center (IDSC), in collaboration with UNESCO, held a workshop on Generative AI (MENA 2023a). The main objective of such workshops is to help young people form sustainable views on innovations to address future technological challenges.

Conclusion

Based on the analysis, it can be concluded that the problem of malicious use of AI encompasses all three levels in Egypt. The country ranks among the leading countries in the MENA region in the field of AI development and has no plans to stop there. The first level of threats relates to the possibility of manipulation of fears of unemployment growth and personal security risks due to the implementation of AI technologies. At the second level, the most problematic areas remain virtual fraud, including the use of chatbots, and hacker attacks on critical infrastructure objects of the state. Analysis of the third level of threats has shown that there is a growing concern in Egypt's society about the ability of AI to create fake information and actively promote it in virtual space. Citizens themselves are frequently subjected to the use of such technologies due to excessive mistrust and lack of caution. Thus, there is a need to develop mechanisms to detect and prevent such cases in the future, as well as to raise citizens' awareness of the risks on the Internet, in order to ensure a positive momentum in countering AI malware threats in the country. However, considering the rapidly changing technological landscape and geopolitical environment, Egypt's government should also be prepared for new threats that may arise.

References

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

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

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

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

Blunt P (2023) Google's Former Egyptian AI Expert Warns of Impending Disaster as AI Develops Negative Perception of Humanity. In: Asume Tech. https://asumetech.com/googles-former-egyptian-ai-

expert-warns-of-impending-disaster-as-ai-develops-negative-perception-of-humanity/ . Accessed 29 Jan 2024

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

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

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

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

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

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

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

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

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

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

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

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

Salah A (2023) INTERVIEW: How Egyptians can benefit from ChatGPT, avoid potential negative consequences. In: Ahram Online.

https://english.ahram.org.eg/NewsContent/1/2/498717/Egypt/Society/INTERVIEW-How-Egyptians-can-benefit-from-ChatGPT,-.aspx. Accessed 6 Feb 2024

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

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

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

The Malicious Use of AI: Challenges to Psychological Security in the Islamic Republic of Iran

Evgeny PASHENTSEV, Pavel KUZNETSOV

Introduction

Iran has adopted and is implementing a national AI development roadmap, which is designed to raise the country to the top ten leading countries in the field of AI from its current 13th place according to the Nature Index. For these purposes, it is planned to invest US$8 billion on AI research and development (Tehran Times 2022a). Artificial intelligence-powered medical technologies show impressive results in diagnosing diseases - for example, the accuracy of diagnosing breast cancer using the system developed by the Iranian University of Medical Sciences (IUMS) reached 94% according to the latest data. However, as in any other area, the development of AI technologies can result in the risk of their malicious use. The situation in Iran is influenced by: the presence of internal ethnic and political contradictions (Ziya 2021), serious manifestations and consequences of public sector corruption (Iran International, 2023), and, what's even more important, tense outer pressure from Israel and the USA. In the public part of the annual threat assessment report of the US intelligence community, Iran was ranked among the four countries that allegedly pose the greatest threat to the US national security and international security, along with Russia, China and North Korea. It is noteworthy that the corresponding chapter of the report specifically mentions Israel, which is also allegedly threatened by Iran (Office of the Director of National Intelligence 2023). At the same time the United States, are a leader in the development and implementation of AI technologies, and Israel sees Iran as its main threat in the Middle East (Berman 2023). According to Eyal Zamir, Defence Ministry director-general of Israel "our mission is to turn the State of Israel into an AI superpower and to be at the head of a very limited number of world powers that are in this club" (Williams and Maclean, 2023). Thus, the leadership of the two countries has the motive and the capabilities to use AI technologies against Iran to a much greater extent than Iran's own capabilities permits this country to use AI against the USA and Israel. This set of problems and hostile actors towards this country create an extremely tense environment and fertile ground for MUAI.

The First Level of Threats of MUAI Against Psychological Security

The public in Iran, as well as in many other countries where take place active research and implementation of AI technologies in various fields, is worried about the impact that AI can have on the job market (Iran Talent 2023). However, there is no unambiguous point of view on this issue. Some representatives of the country's scientific community focus the risk of job loss on specialists with "average" level qualifications, regardless of the creative component of the work process - generative AI models are already quite capable of solving problems for which nonlinear thinking of the human brain was previously considered necessary. According to Hamidreza Keshavarz, a professor at the University of Tehran and the Tehran University of Medical Sciences, low-skilled (territory cleaning, rough physical work, etc.) and truly highly qualified (requiring serious academic education and a highly developed mental apparatus, or based on the qualities of a person, consistently demonstrating outstanding results) jobs are in relative safety. At the same time, the labor market for specialists with "average" qualifications is most vulnerable from the point of view of replacing humans with AI-based development (Khabar Online 2023). Mass layoffs in the largest international corporations, where AI

technologies are being widely introduced and the risks of even larger layoffs cannot but cause concern for a significant part of the population in Iran. Such natural concerns can be purposefully amplified by malicious internal and external actors, especially if proper measures for social protection and human development are not taken during further larger processes of robotization and the introduction of AI technologies.

Hot controversy and discussion, usually in Western countries, is being inspired by the Iran's potential use of AI technologies to automate the control of public order. A real media explosion in the Western information space was caused by news about the Iranian authorities' use of AI to automatically identify, with the help of city video surveillance systems, women who do not comply with the rules of wearing closed clothing (hijab) (Alkhaldi and Ebrahim 2023). The hype was also supported on the site of *Council on Foreign Relations* (CFR) in December 2023, with a post stating that restrictions on Internet access and a bill on identifying violators of moral standards using AI «reportedly led to the arrests of more than twenty thousand people and the killing of more than five hundred young protesters» (George 2023), probably implying that due to the protests. However, it is hardly accidental that it does not being taken into account that Al Qaeda and other terrorist organizations are actively operating in the country, with some of which at the same time relying on the support of external forces. Terrorist attacks occur quite frequently in the country, the latest and largest of which occurred on January 3, 2024 in Kerman, when thousands of people came to the burial site of Lieutenant General Qassem Soleimani, who became a symbol of the fight against terrorism in the country, to honor his memory on the fourth anniversary of his death. As a result of the explosions carried out by suicide bombers, at least 93 people were killed and several dozen more were injured (Tehran Times 2024).

Publications (Jahan News 2023) in media show videos in which an AI system automatically establishes the identities of the captured women. A number of Iranian police officials and representatives of the Iranian parliament, at the same time, confirmed, according to media, the government's intentions to widely implement AI technologies to identify various offenses, including violations of an "ethical nature" (lack of closed clothing for women) and automated prosecution of facts established by this method. Concerns, however, regarding not the moral aspect of such a practice, but the accuracy of the operation of such technological solutions, were also expressed by the country's former Minister of Communications, Azari Jahromi. However, representatives of the police department are confident, as reported, that possible problems with the accuracy of the system will be eliminated over time, when the AI receives enough data for training (Ensaf News 2023). Despite the obviously politicized nature of attacks on Iranian authorities in Western media, the issues raised in this discussion are indeed relevant. At the same time, these problems are purposefully aggravated by interested malicious actors, primarily external ones, who, using considerable financial, organizational, technological, and military resources, do not pursue the interests of the Iranian people, but their own imperialist goals in the region.

The Second Level of Threats of MUAI Against Psychological Security

Threats at the second level are also real and they are growing. In December 2023, General Gholam Jalali, commander of Iran's civil defense, stated (Tehran Times 2023), that the recent disruption of gas stations in the country was caused by malicious software (hereinafter referred to as "malware") during targeted cyber-attacks. Without declaring a direct connection with MUAI, in the same speech the general noted that up to 50% of cyber-attacks on Iranian critical information infrastructure involve AI technologies in one way or another. In the context of the mentioned cyber-attack, it can be assumed that, as in "global practice," attackers could use AI to prepare phishing messages, through which the malware was delivered to the target infrastructure. Earlier, in August 2023, General Jalali stated that AI

was used by external actors in preparation for mass protests, and that Iran should learn to use AI to counter such use (Mohammadzadegan 2023).

It is also pertinent to remember that the aforementioned General Qasem Soleimani was killed in a U.S. drone strike on January 3, 2020 near Baghdad International Airport while traveling to meet with Iraqi Prime Minister Adil Abdul-Mahdi. The video of the event was transmitted live to the US White House, CIA headquarters in Langley and at least one other location for the benefit of Department of Defense officials. The operation was overseen by Gina Haspel and Mark Esper, who at the time served as CIA Director and Secretary of Defense, respectively (Dilanian and Cube 2020).

The White House sent a notice to Congress outlining the legal and political justification for the airstrike that killed the Iranian general. In the notice, the Trump administration cited Article II and the 2002 authorization for the use of military force against Iraq to justify the US strike. The administration said the purpose of the action was to "deter Iran from conducting or supporting further attacks against United States forces and interests" and to "degrade Iran's and Qods Force-backed militias' ability to conduct attacks." (Setzer 2020) Such arguments provoked an angry reaction not only from Iran, but also condemnation from many legislators, mainly Democrats (Choi 2020, Pengelly and Helmore 2020). According to Peter Singer, an expert on future wars at the New America Foundation, "In less than a generation, we went from something that was abnormal and maybe even science fiction, to the point where it's the new normal." (Dilanian and Cube 2020) Of course, AI technologies were highly involved in the preparation and implementation of this operation, as well as the presentation of this information to the media.

It is also noticeable to remember the tragic incident on November 29, 2020, when renowned Iranian nuclear scientist Mohsen Fakhrizadeh was killed in an attack on a highway near the capital Tehran. According to Iranian authorities, the scientist was the victim of an attack by the Mujahideen-e-Khalq terrorist organization, which prepared the activation of an electronic device, allegedly in the interests of Israel. The equipment installed in the van was targeted Fakhrizadeh, using artificial intelligence to identify him before going off, injuring those who accompanied the scientist during the incident. (Motamedi 2020)

So, the AI technologies are obviously actively used at the second level of threats posed to Iran's IPS.

The Third Level of Threats of MUAI Against Psychological Security

Iranian media are currently in a phase of active transformation, associated, as throughout the world, with the rapid digitalization of society. In addition to quick access to information, the number of links in the chain "content creator – content consumer" that controls published content is also reduced. Due to these changes, information injections that are much more effective and destructive in their consequences, including those created using generative AI models, become possible. As an example, a recent, very offensive, publication can be cited, where in a video the famous Iranian cleric Hossein Ansarian allegedly argues that power in the country has been seized by donkeys and shows on the screen a "proof", where a human head is crudely drawn onto the donkey's body (Iran NTV 2023). Such manifestations reinforce the need to intensify the work of the Iranian National Commission on Artificial Intelligence, because Iran, like all other countries, is in dire need of introducing technologies to detect such disinformation materials, otherwise the country risks losing the masses' trust in information and the psychological stability of society.

The media have also repeatedly reported on information and psychological operations carried out on the social networks Facebook (Meta) and Twitter (X) by the Central Command (CENTCOM) of the US Department of Defense. CENTCOM's information operations have been and are ongoing over

long periods of time, including the dissemination of anti-Iranian propaganda. When carrying out such operations, AI was used both in generating texts and to give more weight to publications by generating photorealistic images (deepfake) of supposedly real social network users on whose behalf the texts were published (Tehran Times 2022b).

Despite the risks associated with the implementation of AI, which are actively discussed by the international community, research is already underway in various, rather "sensitive" industries in Iran regarding the use of AI-based chatbots. For example, researchers from the Tehran University of Medical Sciences in their article describe in detail the advantages of using AI in medicine for processing large amounts of data and using chatbots for medical consultations (Hajialiasgari Khanahmadi and Atashi 2023). Among the aspects of potential MUAI and psychological risks, illegitimate access to confidential information (violation of medical confidentiality), possible psychological reactions of patients to the replacement of a living doctor with a "soulless machine" are mentioned. And external interference or errors in the operation of chatbots can lead to medical errors in prescribing treatment. At the same time, the researchers come to a rather positive conclusion about the need to introduce AI into the work of the Iranian healthcare system.

Representatives of the Iranian leadership up to the Grand Ayatollah have repeatedly and rightly emphasized, since 2020, for obvious and compelling reasons, that Iran should be in harmony with technological progress and become one of the leading countries in AI technologies. However, the introduction of AI into religious practice can have ambiguous consequences. Thus, in 2023, Mohammad Ghotbi, the head of the Eshragh Creativity and Innovation House, literally stated that "Robots can't replace senior clerics, but they can be a trusted assistant that can help them issue a fatwa in five hours instead of 50 days." (Bozorgmehr 2023). The introduction of AI into the decision-making process in such a sensitive area as religion, especially in a country like Iran, where religion is directly linked to government, can lead to the most devastating consequences if the aforementioned "robot assistants" are compromised by malicious actors. Particularly serious risks will arise with the development of emotional AI and in the event of a possible further aggravation of the situation in Iran and around it.

Conclusion

Technological progress always requires balanced decisions, and measures to develop AI as a key technology in the transition period to a new social and international order require a particularly responsible approach. And along with the development of trusted AI systems, it is necessary to comprehensively analyze the threat model of each such system in advance, including taking into account information, psychological security and associated risks. The introduction of AI into public life without taking precautions and introducing countermeasures against MUAI can cause significant damage to any society and state, and Iran, being at the forefront of resistance to the imperialist pressure, is even more exposed to the associated risks. In this case, malicious influence can be exerted both by external actors, in order to influence the internal stability in the country or its foreign policy course, and by internal ones, for example, in an attempt to illegally enrich themselves or raise their own political rating.

Since the external pressure on the country, however, significantly exceeds the internal pressure and external malicious actors have sophisticated AI technologies, first of all the MUAI threats at the second and third levels against psychological security are most relevant and dangerous for Iran. Member countries of the Five Eyes intelligence alliance have repeatedly demonstrated their capabilities in conducting information and psychological operations, and the effectiveness of NATO countries' use of the latest weapons and military equipment using AI technologies can be observed in the Ukrainian

theater of military operations. Therefore, such close attention to Iran by the US intelligence community cannot but cause attention among independent observers.

It should be noted that the Iranian authorities approached the task of achieving substantial progress in the field of AI systematically and from the position of strategic planning. It can be assumed that if the Iranian government pays close attention to the threats to psychological security through MUAI, the approach to countering them will become equally structural and systemic.

References

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

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

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

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

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

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

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

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

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

Iran NTV (2023) Happy courier - Deepfake Hossein Ansarian [دیپ فیک حسین انصاریان . پیک شادی]. https://iranntv.com/908619/دیپ-فیک-حسین-انصاریان-پیک-شادی. Accessed 03 Feb 2024

Iran Talent (2023) Will artificial intelligence really make us all unemployed? [واقعا هوش مصنوعی همه ما را بیکار خواهد کرد؟]. https://www.irantalent.com/blog/impact-of-artificial-intelligence-job-losses/. Accessed 03 Feb 2024

Jahan News (2023) Identification of veiled women with artificial intelligence [شناسایی زنان بی‌حجاب با هوش مصنوعی]. https://www.jahannews.com/news/840663/فیلم-شناسایی-زنان-بی-حجاب-هوش-مصنوعی. Accessed 03 Feb 2024

Khabar Online (2023) Who will be made unemployed by AI? [هوش-مصنوعی-چه-کسانی-را-بیکار-می-کند]. https://www.khabaronline.ir/news/1724621/هوش-مصنوعی-چه-کسانی-را-بیکار-می-کند. Accessed 03 Feb 2024

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

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

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

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

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

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

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

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

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

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

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

The Malicious Use of AI: Challenges to Psychological Security in the Federal Democratic Republic of Ethiopia

Sergey A. SEBEKIN

Introduction

Ethiopia is one of the African countries where AI technologies are being successfully implemented to solve various tasks and where extensive institutional conditions for AI development are being created despite the turbulent socio-political situation (Ade-Ibijola & Okonkwo, 2023, pp. 102, 104; Gadzala, 2018, pp. 1, 2, 5, 8). The priorities for the introduction of AI systems are the agricultural sector (agriculture is the basis of Ethiopia's economy) (Federal Democratic Republic of Ethiopia, 2020, p. 26; Girmay, 2019, pp. 161-162, 166-167), public health, financial sector and public administration. Ethiopia has created and is developing the so-called Sheba Valley – the country's technological centre (similar to the Silicon Valley in the USA) (Eke, Wakunuma, & Akintoye, 2023a, p. 4). The Artificial Intelligence Institute of Ethiopia is engaged in systemic multi-vector progress in the field of AI. There are also private companies carrying out AI research, including iCog Labs, a private AI research lab established in 2013 in Addis Ababa that provides a broad range of AI research and development services for domestic and international clients. In doing so, one of the announced prioritised goals is to create an AI tailored to Ethiopia's specificities and values. To date, the level of people's access to digital services is still not high enough.

Ethiopia has a population of about 120 million. It is estimated that only about 16-20 per cent of them have access to the Internet. The number of social media users is even less – about 5% (Kemp, 2023). The fact that Ethiopia is characterised by quite a low level of digitalisation among the population may be an obstacle to influencing the mass consciousness through AI. Therefore, in the future, as the digital infrastructure in Ethiopia develops, the overall level of digitalisation increases and access to digital services becomes more extensive, the technological and institutional conditions for MUAI will extend significantly.

On the other hand, to date, the social, socio-political and socio-economic situation in Ethiopia remains extremely unstable. The key contradictions here centre around the conflict between the "traditionally rebellious" Ethiopian province of Tigrayan and the party known as Tigrayan People's Liberation Front (TPLF), on the one hand, and the Federal Government (Prime Minister – Abiy Ahmed Ali), on the other hand (Afriyie, Ayangbah, & Effah, 2023; Center for Preventive Action, 2023).

The conflict potential in Ethiopia persists for the following reasons.

Firstly, the country suffers numerous inter-ethnic contradictions and centres of ethnic tension. Clashes at the local (grassroots) and inter-regional level between different ethnic groups take place regularly on the grounds of land claims, religious differences, etc.

Secondly, in addition to the official national armed forces, each province had (or has) its own "ethnic" paramilitary units not controlled by the federal government, which engage in regular ethnic cleansing operations. Although the federal government is trying to eliminate these units and effectuate disarmament, not all provinces agree with this policy.

Thirdly, the above is explained by the actual socio-economic situation repeatedly aggravated by purely natural factors – famine due to drought and locust invasion – and humanitarian crises being a consequence of armed conflicts, e.g. forced migration, lack of food, etc. The issue of access to the Red

Sea remains existentially important for Ethiopia today. These aspirations are being challenged by Eritrea which seceded from Ethiopia in 1993, depriving Ethiopia of access to the sea.

In addition to the internal contradictions, Ethiopia has strong disagreements with neighbouring countries as well. Ethiopia also has strong contradictions with Egypt and Sudan over Nile water allocation. In 2023, Ethiopia started filling the Renaissance Dam reservoir, which is regarded by Egypt as an expressly escalatory action that complicates the dialogue.

Thus, while the technical and infrastructural conditions for using AI are still maturing, the socio-political and economic conditions for achieving disruptive effects through MUAI in Ethiopia have long been in place. If the level of digitalisation in Ethiopia increases in the future while the ingrained problems remain unresolved, these institutional factors together will create an amazing synergistic potential for the realisation of MUAI in Ethiopia through influencing mass consciousness with a view to achieve, explicitly and tacitly, specific effects conceived by interested actors, both internal and external.

The First Level of Threats of MUAI Against Psychological Security

The first and most obvious problem that may arise in connection with the widespread introduction of foreign AI systems in Ethiopia lies in the fact that the algorithm based on machine learning, trained on foreign data, imbued with Western (or other) values, ethics and ideas about the ways of solving the set goal may just prove to be inefficient in the realities of Ethiopia with its completely different ethnocultural, ethical, political and economic traditions, or even lead to negative effects and threaten social, political and psychological stability (Birhane, 2023, p. 250; Blackwell, Damena, & Tegegne, 2021; Eke, Wakunuma, & Akintoye, 2023a, p. 2-3; Eke, Wakunuma, & Akintoye, 2023b, p. VI; Okolo, Aruleba, & Obaido, 2023, p. 54). African experts are concerned about the fact that “non-African” AI technologies for solving African problems do not take into account their specificity (Birhane, 2023, pp. 250, 254-255; Eke, Wakunuma, & Akintoye, 2023a, p. 1-2).

Another significant challenge lies on the flip side of the problem when the personal data of Ethiopians will be collected en masse and used to train specific algorithms for solving problems in the Ethiopian context and to respectively personalise due services and content (Birhane, 2023, pp. 249; 251-252). The foreign actors’ access to these data opens up a broad range of opportunities to manipulate them and influence target audiences in Ethiopia (to be discussed later).

Finally, many experts from Africa fear that the mass introduction of “non-African” AI systems for solving certain problems as well as foreign digital infrastructure in general will make African countries overly dependent on imported technologies and plunge them into “AI-neocolonialism” – the so-called Algorithmic Colonisation of Africa (Adams, 2021; Birhane, 2023; Eke, Wakunuma, & Akintoye, 2023b, p. VI) (or “Digital Colonisation of Africa”), where interested actors will use AI technologies not only to solve Africa’s acute problems but also to tacitly influence the economic, political and social processes in the region for the realisation of their interests. It is also assumed that the main interests of companies and corporations working in the AI sphere will not focus on ethics and ethno-culturally sensitive AI training but on profit-making through the export of AI technologies to African countries (Birhane, 2023, pp. 251-252; Okolo, Aruleba, & Obaido, 2023, p. 41, 54; Eke, Wakunuma, & Akintoye, 2023b, p. VI).

Another obvious challenge to psychological safety as a consequence of the first-level MUAI for Ethiopia is automation and the resulting job losses which could take on a mass-scale character in the context of the considered country (Girmay, 2019, p. 170). The fact is that the consequences of job automation, caused by the introduction of AI, may be much more profound in developing countries (Ethiopia is one of them) than in developed countries (The Conversation, 2023). Firstly, the structure of the economy of the developed countries is much more “multidirectional” and complex in terms of available sectors, which supposes a multitude of highly qualified jobs and thus reduces the risks of

sweeping and all-encompassing AI automation. Secondly, the developed countries respectively have more developed economies and have a multitude of resources for flexible automation policies and the creation of new highly qualified jobs to replace the old ones through strong educational and retraining/upskilling programmes. At the same time, the developed countries can afford to implement complementary direct support measures such as unconditional basic income and various benefits, which Ethiopia is unlikely to be able to realise. Despite various forecasts stating that AI is able, as a minimum, not to reduce, or even create, more jobs, this issue may have far-reaching implications for the developing countries in Africa with their less developed economic structure. For instance, the backbone of Ethiopia's economy is agriculture and services – the sectors that are among the most promising for modernisation owing to the widespread adoption of AI systems (Federal Democratic Republic of Ethiopia, 2020, p. 9, 26; Girmay, 2019, pp. 161-162; United Nations, 2023). Despite the successful development of professional education in the sphere of AI, the problem is exacerbated by the so-far insufficient level of general higher education in Ethiopia and its mainstream accessibility to the general population, which further complicates the ability for the realisation of the labour potential in the new conditions. Ethiopia has a much higher proportion of young people than in the developed countries, and these people are more likely to lose jobs involving AI and AIrobots technologies, having low chances of finding equivalent employment – if no retraining and technical adaptation programmes are promptly introduced and no more technologically advanced jobs are created. The first-level psychological effect of the assumption that AI will "deprive" them of their jobs, especially in case of malicious use of the transition economy problems and in the conditions of high property polarisation, can be extremely destructive and lead to destabilisation of the socio-political situation in the factual Ethiopian conditions, promoting the development of the shadow sector in the economy and aggravating the criminogenic situation as well as activating search for illegal sources of livelihood. The creation of new highly qualified jobs in Ethiopia requires the government to develop due educational programmes in the field of AI (including effective retraining programmes), make this education more accessible, and, most importantly, distribute jobs evenly among different ethnic groups. The underestimation of the coming problems may lead to Neo-Luddism in its various forms – protests against the replacement of jobs with algorithms, strikes and pickets and even new armed actions against the federal government.

Shortly, the "hysteria" about mass automation and job cuts may have a clearly manipulative nature and can be used with an obvious purpose to destabilise social stability. Meanwhile, the slowdown in the introduction of AI technologies will perpetuate the country's backwardness and cannot be the basis for resolving its problems.

The Second Level of Threats of MUAI Against Psychological Security in Ethiopia

Ethiopia, with its traditionally unstable socio-political and politico-military environment, is an excellent target for cyber attacks against it, ranging from cyber attacks on critical infrastructure systems to phishing.

Ethiopia is currently facing a gradually increasing number of cyber attacks on its infrastructural systems. According to the official statistics from Ethiopia's Information Network Security Administration – the main governmental agency responsible for cyber security, the number of cyber attacks on Ethiopia's systems reached almost 7,000 in the last fiscal year (2022-2023) (Ena, 2023b; Ethiopian Monitor, 2023; Reqiq Staff, 2023). Although the number of cyber attacks on Ethiopia's systems is not so high as compared to other countries (for instance, about 106,000 attacks via backdoors and spyware were recorded in South Africa, another BRICS member state, and this figure is even higher in larger BRICS member states), the scale and professionalism of cyber attacks is growing, according to Kaspersky Global Research and Analysis Team (Teshome, 2023). Moreover, Kaspersky Global Research and

Analysis has recorded slightly different data – 18,000 cyber attacks and 30,000 attacks by ransomware (Teshome, 2023).

Cyber attacks are mainly directed at financial institutions, health, education, security, media and government sectors (Ena, 2023a; Ethiopian Monitor, 2023; Reqiq Staff, 2023; Teshome, 2023). The main types of cyber attacks and tools used against Ethiopia's systems and population are DDoS-attacks (Denial-of-service), system scanning and infiltration, and malware (including ransomware) attacks on websites (Ena, 2023a; Ethiopian Monitor, 2023; Reqiq Staff, 2023; Teshome, 2023).

According to Kaspersky Global Research and Analysis data and the Microsoft Security Intelligence Report, Ethiopia has been most heavily targeted by ransomware phishing attacks in recent years (Microsoft, 2023; Tessema, 2023b). One of the most high-profile (and most humorous) удалить cases took place in Ethiopia in 2023, when the Ethiopian Ministry of Finance transferred about $ 5 million, intended for the African Development Bank, to fraudsters as a result of a phishing attack (Tessema, 2023a). The fraudsters used the credentials of the African Development Bank to organise the attack. Most importantly, this incident had real consequences and provoked a diplomatic scandal: after it, two employees of the representative office of the African Development Bank in Adidas Ababa were detained with the use of violence on suspicion of fraud – because one of the employees, Abdul Kamar, had not confirmed the bank transfer (Horn Observer Contributor, 2023). This, in turn, may also point to another cybersecurity problem in Ethiopia – a low level of digital literacy and lack of cyber hygiene awareness – since the Ethiopian Ministry of Finance had not verified the remittee's account numbers.

At the same time, the development of digitalisation and the digital economy has the potential to expand the scope of cyber attacks and generate enhanced cyber exposure against Ethiopia in terms of quantity and quality. In particular, AI may be used for targeted cyber attacks and mass dispatch of targeted phishing scam messages. Big data analysis makes it possible to tailor cyber attacks to a specific target – e.g. a particular organisation, production, system, etc. Using AI-based phishing, specific machine learning algorithms and data analytics, the intruders will be able to generate personalised text messages for "especially important" individuals, for instance, top officials of a company, a governmental institution, etc. (Bahnsen et al., 2018; Goldman, 2022; Guembe et al., 2022, pp. 84-85, 89, 96-97, 102; Seymour & Tully, 2016; Zouave et al., 2020, p. 22-23). Such cyber attacks in the Ethiopian context could easily foment an inter-ethnic conflict and provoke new waves of social tensions. For instance, there exists a real prospect of mass phishing mailing on behalf of an alleged leader of the opposition-minded ethnic group calling for a war against another ethnic community or even the federal government, or phishing scam emails with information on raising funds for organising a militia.

The Third Level of Threats of MUAI Against Psychological Security in Ethiopia

The third level of threat relates directly to specific MUAI technologies which could lead to severe consequences and destabilisation of the military, political, social and public environment in the Ethiopian context. A real-life scenario was demonstrated when a deepfake using AI was widely circulated, stating that the head of the Prosperity Party, Girma Yeshitila, had been killed in the Amhara region by the ethno-nationalist Amhara militia (FANO) (Addis Insight, 2023). In 2023, fraudsters attempted to impersonate the African Union Commission chairman (the organisation is headquartered in Addis Ababa), Moussa Faki, using a deepfake to make video calls to several European leaders (Adjetey, 2023).

A specific destructive socio-political effect can be achieved by chatbots trained on certain information and programmed according to certain ideological, political and other value paradigms (Mihajlenok & Malysheva, 2020). For instance, chatbots, using special digital platforms, can spread information among the Amhara people that the disputed territories in fact historically belong to the

Tigrayan people or vice versa – this can provoke the expression of dissent through localised clashes and lead to destabilisation of the social order.

A multitude of negative scenarios in the Ethiopian context can effectuate the use of predictive capabilities of AI as a prognostic weapon.

By analysing certain data – the level of social stability, political preferences or extent of loyalty to the federal government – the AI that is trained to predict the level of social stability in a particular province can foresee that Somalia, for instance, will face a social explosion. If the artificial intelligence assumes that the number of males in Amhara province is growing it can infer that these people will try to reclaim the disputed territories in a few years. All of this can further destabilise the situation in the provinces.

The AI-based targeted automated profiling technology can be widely used in Ethiopia. It draws psychological portraits and enables the classification of target Internet users on the basis of the analysis of (preferably) open data from social networks, Internet resources, search queries, etc. in order to identify their psychological characteristics, and emotional background and even to predict their future psychological state, with the purpose to influence and motivate them to take certain actions (Bilal et al., 2019; Guembe et al., 2022, p. 95; Zouave et al., 2020, p. 19). For instance, AI can be applied to analyse data on large target groups which can be represented in this case by different peoples of Ethiopia – Amhara, Oromo, Tigrayan, Ometa, Irobs, etc., so as to make a kind of "socio-psychological maps" for them with the help of a certain algorithm: to analyse behavioural patterns and peculiarities, to identify mass political preferences of certain peoples, to assess the extent of loyalty to the federal government and to highlight the factors of concern – in order to influence the mass consciousness, set the desired political agenda, push people towards certain actions, etc.

In addition, a number of moments negative for Ethiopia's psychological, political and social stability can be attributed to the actual implementation of specialised algorithms intended for a wide range of specific tasks.

The interested actors – some states and large corporations – that are implementing various AI platforms and digital ecosystems may be interested in the current unstable situation in Ethiopia to realise their own interests that are not always compatible with Ethiopia's interests. The most important thing to remember is that special platforms and applications with embedded AI algorithms, aimed at the realisation of specific social objectives as well as entertainment purposes, will collect a huge amount of information about Ethiopian citizens. External actors who have access to these data arrays will be able not just to manipulate this information but to use it for their own purposes applying a wide range of specialised AI technologies for the achievement of various goals.

Further, this will allow the use of specific AI technologies to have a targeted impact on the collective consciousness of the Ethiopian population (or specific target audiences) in order to achieve a multitude of sought-after effects. For instance, specially designed AI technologies can be used to create and disseminate a fake among the population asserting that the disputed territories claimed simultaneously by the Amhara and the Tigrayan are currently being conceded (or are going to be conceded) to one side or the other, or that the federal government is planning to reduce the number of representatives of some province (respectively, of some nationality) in EPRDF. This can be achieved with the help of specially trained chatbots, assistants, deepfake, etc.

This way, using specific AI technologies and thus influencing the mass consciousness, any interested actors can shape the desired political agenda and even generate necessary socio-political processes, create hotbeds of tension and provoke new conflicts – which is very easy to achieve in the conditions of Ethiopia, namely, permanent instability and distrust of the authorities on the part of some peoples. For instance, on 31 May 2021, the online media outlet Kello Media published a fake audio

recording in which Prime Minister Abiy Ahmed, during a meeting of the Prosperity Party, allegedly stated that they had won the election and no one else would be able to form an alternative government in the next 10 years (Addis Insight, 2023).

Corporations that own platforms and relevant AI technologies, being affiliated with their government, can use their "digital power" to "inculcate" the need for certain transformations in the minds of Ethiopian residents – promotion of commercial, infrastructural, investment-related and other projects, even if this is contrary to the actual interests of Ethiopians, which may lead to a new wave of discontent and conflicts.

It is important to note that the algorithms integrated into digital platforms analyse the users' data for personalisation of services and content, and thus tailor the content to the preferences of a specific individual (audience). However, such personalisation of content in the Ethiopian context can lead to negative consequences such as further polarisation of peoples in the key issues of politics, territorial affiliation, religious choice, etc. If the users prefer some specific content AI starts showing similar information increasingly often, placing people in a kind of "information bubble", creating an impression that their beliefs are the only true ones and thus tacitly shaping their worldview. For instance, these digital platforms will present the Oromo people with some targeted content – say, on the need to establish their own synod, while the Amhara people will be shown the opposite content on the inadmissibility of church dissent[3]. This can take place due to natural platform-specific algorithm training (say, the Oromo select certain content as a priority, and the Amhara – their preferable content), following which AI simply begins to recommend similar content. And alternatively, this may be a consequence of devised reasons, when the algorithm was set up in a certain way, intentionally. Such mechanisms can be used to further polarise the Ethiopian society, already disintegrated and torn apart by ethnic contradictions, as well as to censor information on a national scale.

A real-life scenario of such destructive algorithmic impact was demonstrated in the armed conflict between the Tigrayan and the Federal Government (2020-2022). According to Amnesty International, Facebook's algorithms tacitly facilitated the spread of destructive content in Ethiopia that called for violence against the Tigrayan people (Amnesty International, 2023). Facebook's content moderation and censorship systems failed to recognise these destructive posts in virtue of the fact that Amhara and Oromo languages were not "priority" languages for moderation (Allen, 2022).

Finally, the issue of data representation and objectivity as well as the extent to which AI will be correctly and appropriately trained towards solving particular tasks becomes particularly thorny in the conditions of the ongoing conflict potential in Ethiopia – acute inter-ethnic conflicts and territorial disputes on this ground, the extremely unstable political and socio-economic situation in Ethiopian provinces. The use of "biased" AI, trained on unrepresentative data, can be fraught with very serious consequences in certain areas and further destabilise the situation, leading to a regular "explosion". Specific AI technologies based on machine learning, inculcated for the solution of certain social, public and political problems by foreign companies and Ethiopian state structures may discriminate against certain groups of people, including on ethnic grounds. For instance, any social algorithm used for particular public objectives – analysis of final examination results, assessing CVs of candidates applying for employment, etc. – can empower the people of Oromo to a pronounced extent, while implicitly discriminating against the Tigrayan: if a particular company employs more Oromos, then the AI trained on such data will continue to prioritise them. Or, an algorithm used in political processes can grant more political rights to some people while marginalising others. For instance, the AI applied to analyse the composition of EPRDF, if trained on unrepresentative data, can theoretically provide more seats to

[3] One should recall the clashes on this ground that broke out in the Oromo-populated areas in February 2023.

Amharas people than to Tigrayans. Representatives of small-numbered peoples will be particularly vulnerable to discrimination in such scenarios.

It is not difficult to imagine that the use of AI to influence the mass consciousness of Ethiopians can catalyse another wave of violence and lead to bloody conflicts grounded on multiple ethnic tensions. And this can be exploited by interested actors, both external and internal, who disagree with the policies of the federal government.

Conclusion

Today, Ethiopia is facing an extensive range of internal contradictions, at the same time experiencing pressure from external actors. Such conditions make Ethiopia an excellent potential target for MUAI in terms of destabilisation of the public and socio-political situation, which can be achieved through influencing popular mentality by interested actors. Moreover, the very fact of applying socially and politically oriented algorithms in the Ethiopian context can lead to the discrimination of certain groups of the population – if AI is trained on unrepresentative data.

At the same time, the existential challenges in Ethiopia may be provoked by the proliferation of imported AI systems and digital platforms with embedded algorithms. Potentially, this can lead to "AI neocolonialism" and make Ethiopia overly dependent on foreign digital infrastructure and introduction of imported AI systems that do not take into account the local interests, ethnic specificity, culture and mentality. It is expected that the ethical, legal and socio-cultural consequences of using such systems will be ignored. Moreover, using digital platforms, corporations can gain access to a huge amount of personal data of Ethiopian residents, which will subsequently allow them to manipulate these data in order to realise their own interests.

References

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

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

Afriyie FA, Ayangbah S, Effah KO (2023) Diagnosing Ethiopia's Tigray War: Reverberations in the Horn of Africa. Insight on Africa, Volume 15 (2), pp. 139-151. https://doi.org/10.1177/09750878231170177

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

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

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

Center for Preventive Action (2023) Conflict in Ethiopia. In: Council on Foreign Relations. https://www.cfr.org/global-conflict-tracker/conflict/conflict-ethiopia. Accessed 29 Jan 2024

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

Eke DO, Wakunuma K, Akintoye S (eds) (2023b) Responsible AI in Africa: Challenges and Opportunities. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-08215-3

Ena (2023a) Ethiopia Fends Off Exponential Spike in Cyber Attacks, Thwarts over 96 Percent of Attack. https://www.ena.et/web/eng/w/eng_3120234. Accessed 03 Feb 2024

Ena (2023b) INSA Foils 6768 Cyber-attacks in Concluded Fiscal Year. https://www.ena.et/web/eng/w/eng_3120234. Accessed 03 Feb 2024

Ethiopian Monitor (2023) In 12 months, INSA Foils Over 6,700 Cyberattacks on Ethiopia. https://ethiopianmonitor.com/2023/07/24/insa-foils-over-6700-cyberattack-attempts/. Accessed 03 Feb 2024

Federal Democratic Republic of Ethiopia (2020) Digital Ethiopia 2025 – A Strategy for Ethiopia Inclusive Prosperity. In: Ethiopian Legal Information Portal. https://www.lawethiopia.com/images/Policy_documents/Digital-Ethiopia-2025-Strategy-english.pdf. Accessed 03 Feb 2024

Gadzala A (2018) Coming to Life: Artificial Intelligence in Africa. In: The Atlantic Council. https://www.atlanticcouncil.org/wp-content/uploads/2019/09/Coming-to-Life-Artificial-Intelligence-in-Africa.pdf. Accessed 03 Feb 2024

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

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

Kemp S (2023) Digital 2023: Ethiopia. In: Datareportal. https://datareportal.com/reports/digital-2023-ethiopia. Accessed 03 Feb 2024

Microsoft (2023) Microsoft Security Intelligence Report. https://info.microsoft.com/SIRv24Report.html. Accessed 03 Feb 2024

Mihajlenok OM, Malysheva GA (2020) Robotizaciya social'nyh setej i ee politicheskie posledstviya [Robotization of Social Media and Its Political Implications]. Vlast' (Power), Volume 1 (28), pp. 85-92.

Okolo CT, Aruleba K, Obaido G (2023) Responsible AI in Africa – Challenges and Opportunities. In: Eke DO, Wakunuma K, Akintoye S (eds) Responsible AI in Africa: Challenges and Opportunities. Palgrave Macmillan, Cham, pp. 35-64. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-08215-3

Reqiq Staff (2023) A Daunting Digital Frontier: The State Of Cybersecurity In Ethiopia. In: Reqiq Insights. https://reqiq.co/a-daunting-digital-frontier-the-state-of-cybersecurity-in-ethiopia/. Accessed 03 Feb 2024

Teshome M (2023) Cyber attacks bombard Ethiopia. In: Capital Ethiopia. https://www.capitalethiopia.com/2023/06/12/cyber-attacks-bombard-ethiopia/. Accessed 03 Feb 2024

Tessema B (2023) Increasing cyber-attacks target Ethiopia. In: Abren. https://abren.org/increasing-cyber-attacks-target-ethiopia/. Accessed 03 Feb 2024

The Conversation (2023) Whose job will AI replace? Here's why a clerk in Ethiopia has more to fear than one in California. https://theconversation.com/whose-job-will-ai-replace-heres-why-a-clerk-in-ethiopia-has-more-to-fear-than-one-in-california-216735. Accessed 03 Feb 2024

The Economist (2023) The world's deadliest war last year wasn't in Ukraine. https://www.economist.com/international/2023/04/17/the-worlds-deadliest-war-last-year-wasnt-in-ukraine. Accessed 03 Feb 2024

United Nations (2023) With AI, jobs are changing but no mass unemployment expected - UN labour experts. In: Department of Economic and Social Affairs. https://www.un.org/ru/desa/ai-jobs-are-changing-no-mass-unemployment-expected-un-labour-experts. Accessed 03 Feb 2024

Zouave E, Bruce M, Colde K, Jaitner M, Rodhe I, Gustafsson T (2020) Artificially intelligent cyberattacks. In: Totalförsvarets forskningsinstitut FOI. https://www.statsvet.uu.se/digitalAssets/769/c_769530-l_3-k_rapport-foi-vt20.pdf. Accessed 03 Feb 2024

The Malicious Use of AI: Challenges to Psychological Security in the Federative Republic of Brazil

Darya BAZARKINA, Evgeny PASHENTSEV

Introduction

Brazil's president, Luiz Inacio Lula da Silva, in January 2024 unveiled a development plan for the next ten years aimed at boosting industrial growth with state credits and subsidies. Digital transformation is among the goals set out in the plan: the aim is to digitalize 90% of all businesses operating in the industrial sector in Brazil (the current percentage of companies that operate digitally in the sector is 23.5%). This will involve investments in the Industry 4.0 approach, with the integration of intelligent digital technologies into manufacturing and industrial processes, as well as boosting national semiconductor production (Mari 2024). The market size in the AI market of Brazil is projected to reach US$4.37bn in 2024. In global comparison, the largest market size will be in the United States (US$106.50bn in 2024). The market size of Brazil is expected to show an annual growth rate of 17.65%, resulting in a market volume of US$11.59bn by 2030 (Statista 2023a).

Produced by visual content creation and marketing firm Getty Images, the study VisualGPS surveyed over 7,000 adults from over 25 countries. It found that four out of six Brazilians believe AI can have a positive impact on their lives, higher than the global average, where only half of those polled agreed with that prospect. Brazilian consumers are 15% more interested in AI compared to the rest of the world, according to the study. In contrast to countries like the United States, Canada, France, the United Kingdom, and Australia, less than 34% of Brazilians feel threatened by the advancement of this technology (Mari 2023a). At the same time, MUAI already poses real threats to psychological security in Brazil, which are largely determined by the severity of socio-political contradictions in the country. The immediate future is also alarming for Brazilians.

Lula’s foremost priority remains lifting a portion of the 71 million Brazilians (33% of the population) grappling with poverty. But the IMF, possibly adopting a pessimistic outlook, projects Brazil’s growth at only 2% annually until 2028. This modest growth rate may offer limited contributions to poverty reduction (Martin 2024, p. 2). Nearly half of Brazilians (49%) fear in January 2024 their incomes will decline in the next six months oa 2024 rather than increase (36%), though they expect the job market to improve (Reuters 2024). In the first six months of 2023, 1,790 homicides were recorded in Brazil, compared with 1,526 between January and June 2022 (Instituto Sou da Paz 2023). The municipal elections in October 2024 will pose the greatest political challenge for President Lula da Silva in 2024, a year that brings a new clash with the far-right led by Jair Bolsonaro. Further improvement and dissemination of AI technologies against the background of increasing socio-political problems and contradictions will naturally result in the growth of MUAI.

The First Level of Threats of MUAI Against Psychological Security

The already mentioned prevailing optimistic vision of the AI future by Brazilians has both its pros and cons. A positive attitude towards new technologies facilitates their development and dissemination, generally increasing the level of well-being of society. However, where and because people underestimate the risks to their security, a collision with reality at a certain stage can provoke an exaggerated negative public reaction. In today's complex internal situation and the increasingly

acute situation in the international arena, any difficulties, imbalances, or errors in the development of AI can be amplified by the actions of internal and external malicious actors. Much depends on the government's ability not only to accept the legal framework for the development of AI in the country (work is ongoing in this direction), but also to convey to the general public both the great opportunities and the equally significant risks that the development of AI technologies brings with it.

The Second Level of Threats of MUAI Against Psychological Security

According to the report released by cybersecurity firm Trend Micro, Brazil is the second country in the world most vulnerable to cyberattacks (Mari 2023b).

In 2021, Brazil and Ecuador were the two Latin American countries with the highest proportion of users exposed to phishing attacks, accounting for 12.39% and 10.73%, respectively (Bianchi 2021). In 2020, Brazil set a global record for number of phishing attacks: every fifth internet user in the country was exposed to a phishing attached at least once (Mari 2022b). According to Kaspersky in 2021 there were 25 million attempted attacks in Brazil, and 134 million phishing attempts were recorded in 2022. (Folha Vitória 2023).

Moreover, e-commerce and social networks have also been gaining ground in this type of crime because they use e-mail addresses to authenticate the identity of their users and, generally, people standardize the e-mail address and password for various online accesses. In this way, often when fraudsters discover a victim's e-mail password, it gives them access to their bank account.

In February 2022, Brazil was included on a Spamhaus list of countries where the largest number of spambots were detected. Most of these bots are used for spamming, phishing, distributed denial-of-service (DDoS) attacks and other malicious activities. Analysts associate the abundance of bots in Brazil's digital space with technical, political and socio-economic factors (The Spamhaus Project 2022). As in all other countries, the growth of cyberattacks in Brazil is accompanied by a general increase in AI technologies in their provision.

The sophistication of cyberattack techniques poses an existential danger to enterprises, and organization infrastructures, with the power to interrupt corporate operations, wipe away critical data, and create reputational damage. Cybercriminals will be able to direct targeted attacks at unprecedented speed and scale while avoiding traditional, rule-based detection measures (Guembe et al 2022). Due to the general lag in cybersecurity, Brazil's vulnerability to these types of attacks can be assumed to be high.

The Third Level of Threats of MUAI Against Psychological Security

The third-level threat MUAI, related to the use of *social media bots* during election campaigns, has been relevant in Brazil for several years. In the case of the presidential elections in 2022 the number of fake accounts grew significantly on Twitter in the months after the vote was counted, with most of them attacking leftist president Lula da Silva. The data obtained indicate a significant surge in the alleged activity of bots on the Internet (Lima 2023).

The 2022 national elections in Brazil were marked by the malicious use of deepfakes created by AI, where fabricated images and videos were used to spread disinformation, showing leading candidates involved in various scandalous and compromising situations. The use of advanced AI and sophisticated editing tools to accurately mimic the voices and facial expressions of candidates, creating a convincing result, often leads to destabilization of public trust and distortion of the electoral environment (Ünver 2023).

In 2022, a political deepfake was directed at the current president, Luiz Inacio Lula da Silva. On August 5, a fake video appeared on social networks where the popular Brazilian TV presenter Renata Vasconcellos allegedly gives false information about the results of voting in the presidential election. A small change was made in Renata's voice so that people who were watching understood that Jair Bolsonaro was ahead of Luiz Inácio da Silva in a voting intention survey that had its result released on August 15, 2022. But in real video it can be seen that Lula was with 44% of voting intentions and Jair Bolsonaro reached 32% among respondents. The big problem in this case is that despite the video having been on YouTube for a short time, there was also a circulation in WhatsApp groups and social networks (Pacheco 2023).

According to Javier Rincón, Regional Director for Latin America at Avast, “a high level of awareness about fake news is essential in society to combat them, as those who start consuming news sites with false information can increasingly enter a whirlpool of fake news. Research from Avast AI team shows that more than 17% of websites that spread disinformation have links to other fake news sites. This can quickly create a chain of fake news consumption” (Bento 2022).

Manipulating the electorate is always a destructive weapon, and it can be amplified exponentially using AI. MUAI in Brazil's elections can frighteningly manipulate and undermine the will of voters (Resende 2023).

Undermining trust in the media can have deep repercussions, particularly in fragile political environments. Sam Gregory, the program director of Witness, a nonprofit that helps people document human rights abuses, offers an example. In Brazil, which has a history of police violence, citizens and activists worry that any video they film of an officer killing a civilian will no longer be sufficient grounds for investigation. Gregory says that the fear real evidence can plausibly be dismissed as fake has become a recurring theme in his workshops (Hao 2019). The problem is not only exposing the practice of deepfakes, but also effectively detecting deepfakes. This is compounded by authorities attempting to declare any fair criticism based on photos, videos, audio or written evidence as deepfakes. Professor Hao Li at the University of Southern California agrees. The risk that arises from the misuse of deepfakes is that people are using their existence to discredit genuine video evidence: “Even though there’s footage of you doing or saying something you can say it was a deepfake and it’s very hard to prove otherwise”. Politicians around the world have already been accused of using this ploy, including João Doria, mayor of Sao Paulo. In 2018, the married politician claimed that a video showing him participating in an orgy was a deepfake that no one could convincingly prove otherwise (Thomas 2020). Thus, the malicious use of deepfakes in Brazil can be both a first- and third-level threat.

Deepfakes can be used as powerful and threatening tool of misinformation, undermining public consciousness (Pinheiro de Resende 2021). Using deepfakes, it's quite easy to fool people because viewers believe they're watching something that actually happened. The sound can also be generated synthetically. This is what journalist Magali Prado says in her book "Fake News and Artificial Intelligence: about applying algorithms to combat disinformation." It is emphasized that deepfake audio files can be easily distributed on platforms such as WhatsApp, which is widely used in Brazil. With the help of available software, which is constantly being improved, it is possible to simulate a human voice. The victims are mostly public figures whose voices can easily be found in the public domain. This method can also be used for financial scams. “In one case, an employee of a technology company received a voice message from a top manager asking him to transfer some money to him. He became suspicious, and the message was analyzed by a security company, which confirmed that it was created using artificial intelligence” (Schmidt 2022).

The more sophisticated deepfakes and the more anti-democratic the state; the more politically apathetic and digitally illiterate the population, the more socially dangerous deepfakes will be. To neutralize deepfakes—and fakes in general—authorities must have a stable and non-selective desire to

uncover socially significant cases of deepfakes and to label deepfakes with facts that represent the objective truth. Without the conscious attention and participation of citizens, authorities are unlikely to be willing or able to effectively counter the malicious use of deepfakes. Proving what is real is real, and what is fake is fake, will be extremely difficult without tangible political prerequisites.

In Brazil (as in a number of other countries), virtual influencers are becoming popular. Although they look like a real person, the virtual influencers are 100% digital but they do real stuff. For example, they can talk, dance, play, in one word they can have real people's behaviour. Besides that, they have a strong opinion, and this makes them famous on social media and maybe that's why they have so many followers and are so popular (Little Black Book 2022). Lil Miquela, otherwise known as Miquela Sousa. The 19-year old Brazilian-American social media star has amassed over 2.5 million followers on the platform and regularly posts sponsored content in partnership with brands like BMW and Pacsun. But she's not a young adult who moved to Los Angeles, she's virtual and was made using computer-generated imagery. Lil Miquela first appeared back in 2016. After her debut on Instagram, she went viral. Once Miquela was forced to admit to her fans that she had been 'hacked' by her nemesis, a pro-Trump troll Bermuda (Petrarca 2018). Fascinated by Miquela's human-like appearance, many wondered whether she was a marketing stunt, a real person, or something else entirely. Finally, in 2018, the truth was revealed when her creators, Trevor McFedries and Sara DeCou of robotics and AI firm Brud, announced they were behind Lil Miquela (as well as Bermuda) (Sheena 2023).

With 31,2 million followers on social media, Lu Do Magalu, a Brazilian social media star, is the biggest virtual influencer on internet (Petrarca 2018). Lu is the brainchild of Frederico Trajano, the CEO of Magazine Luiza, a diversified Brazilian conglomerate of consumer-facing brands including Magalu, one of the country's largest big-box retailers with over 1,300 physical stores spread across the country. Lu came to life back in 2003 just as the e-commerce space was beginning to show signs that it might eventually become a viable option to traditional brick-and-mortar sales. "We're responsible for the humanization of Lu," remarked Aline Izo in an interview with Observer (Wierson 2021). Izo, who guides a team of 3D designers, programmers, and marketers that meticulously cares for every aspect of this in-demand influencer claims that "Lu has millions of fans and when she takes a stand on something – for example on bringing awareness to domestic abuse or standing up and advocating for LGBT rights – people pay attention. Here in Brazil, Lu is not a sales gimmick; Lu is an influencer in the true sense of the word; Lu can move the needle on important issues in society" (Wierson 2021). A virtual influencer is a digital personality that posts on social media to build an audience of passionate fans, just like a human influencer; at least, that's how it seems. A certain political content of such AI influencers is obvious, and the extent to which it is a manipulation not only of adults but numerous numbers of children and teenagers remains an open question. According to Yakov Bart, professor of marketing at Northeastern University and a leadership committee member at Northeastern's Institute for Experiential AI "in some contexts using virtual or synthetic influencers is more efficient when you take into account the return in terms of changes to consumer mindset after interacting with the influencer versus the costs" (Contreras 2024). Michael Gerlich from SBS Swiss Business School comes to the conclusion that virtual influencers can create more profound and long-lasting connections with customers, virtual influencers' flexibility in programming and training AI engines allows them to adapt to changing customer behaviors, and virtual influencers, with higher credibility and trustworthiness than human influencers, are the future of influencer marketing and can increase purchase intention and overall brand awareness for companies (Gerlich 2023, p. 19). Mona Mrad, a marketing professor at the American University of Sharjah in the United Arab Emirates admits that to some, the concept of real humans following virtual influencers might seem ludicrous, but not to certain generations. Generation Z, those born in the late 1990s to the early 2010s, are embracing virtual influencers. "This generation seems to be well-connected to these influencers, to the extent that they are forming emotional and mental relationships with them…They reveal feelings of love and attachment to them."

In some cases, they even perceive virtual influencers to be *more* reliable than human ones (Kugler 2023). With nearly 9.2 million Instagram influencers as of May 2020, Brazil was the Latin American country with the highest amount of digital influencers. Although far behind, Argentina ranked second, registering over 1.1 million influencers (Statista 2023b). This fact makes Brazil especially vulnerable to manipulation through virtual influencers.

It is quite remarkable how predictive analytics based on AI selectively works in Brazil. On 8 January 2023, exactly one week after Luis Inacio Lula da Silva's third inauguration as the President of Brazil, supporters of the far-right ex-president Jair Bolsonaro stormed governmental buildings in Brasilia. "Everyone saw Brazil violence coming. Except social media giants," POLITICO's Mark Scott wrote the day after the insurrection, positing that "Silicon Valley's biggest names have again been caught asleep at the wheel…" (Digital Action 2023). For months before the attack experts had warned that the far-right had been using encrypted platforms like WhatsApp and Telegram to organize and spread falsehoods and provoke insurrection and urged them to keep the election-related safety measures in place until the first weeks of the new government. Many made these requests – which the companies ignored (Digital Action 2023). Not only were the voices of experts not taken into account, but also the resources of predictive analytics based on AI, were not used. On the other side, according to Emerson Saraiva, head of the association of political marketers of Brazil, AI will allow people to know the results of the 2024 municipal elections "well in advance" (Silva 2023). Thus, announced the possibility of using AI to obtain early electoral results that could influence voters' decisions on whether to go to the polls and how to vote. It can be concluded that conditions are already emerging in Brazil (and not only) where predictive analytics based on AI can be used as a prognostic weapon.

Deepfakes, virtual influencers, chatbots, predictive analytics and other means of influencing public opinion using AI could have the most destructive impact on the political process in the country. The Superior Electoral Court (TSE) will hold during the first quarter of 2024 debates to regulate the use of AI in the next elections. Alexandre de Moraes, the President of TSE, believes that with AI it is possible, for example, to modify videos of opposing candidates, making them give statements they have never given. "Imagine how many people might be bombarded with fraudulent news, misinformation, but misinformation from a speech video with almost certainty of truthfulness. The aggression is very great. This aggression, especially with the use of artificial intelligence, can really change the electoral result, can distort the electoral result in polarized elections" (Tocarnia 2023).

It is noteworthy that measures aimed at suppressing propaganda and manipulation on the Internet, proposed by the government of Lula, are met with sharp resistance from leading US multinational technology companies. The pending Brazilian Congressional Bill No. 2630 (Senado Federal 2020), officially Brazilian Law on Freedom, Responsibility and Transparency on the Internet and dubbed by the Brazilian media the Fake News Bill and Censorship Bill by its opponents, which is intended to fight the spread of disinformation, more and more based on support of AI technologies. Regulation in this area is long overdue. According to the study of Avast, four in five (79%) Brazilians found fake news about the elections 2022 on social networks, and the majority (57%) do not believe (or are not sure) if social media is a reliable source of information. In addition, 86% of Brazilians think that the media should take responsibility for the withdrawal of fake news on their networks (Bento 2022). Taking urgent action to correct the situation is obvious to the government and the vast majority of Brazilians, but not to multinational US HT giants.

In early May 2023, when the bill was about to be approved, Google and Telegram used their own platforms to express their opposition to the bill to their Brazilian users. May 1, Brazilians were a bit surprised to see at the familiar search field of the Google homepage, a link said: "The fake news bill can make your internet worse." Whoever clicked on the link was taken to a Google blog that criticized draft law 2630, which was to be voted on Brazilian Congress the next day. The search homepage, used

by more than 90% of 160 million internet users in Brazil, also claimed in another link that “the fake news bill can create confusion about what is true and what is a lie in Brazil”. The Google’s strategy included sending emails to YouTubers saying there would be less money to invest in their channels and asking them to talk to their Congress. The tech giant also fumbled with search results, prominently showing its own blog post and other articles that were critical of the bill, according to a study by the Federal University of Rio de Janeiro (Viana 2023). "Brazil is about to pass a law that will end free speech," the Telegram said in a message sent in May to users on Bill 2630, which has passed the Senate and was awaiting a vote in the lower house of Congress (France 24 2023).

Thus, there is every reason to believe that leading US tech companies are slowing down the adoption of effective measures against the spread of disinformation, the creation and distribution of which is increasingly based on AI technologies. Behind this lies both the selfish financial interests of companies (reluctance to take legal obligations to combat unwitting or conscious violators of public norms on the Internet), and the desire to put pressure on their opponents, even to the point of provoking riots and overthrowing an undesirable government that intends to encroach on excess profits. An additional but important factor is the pressure of the US government, which seeks to use HT US companies both to obtain sensitive information and to put pressure on undesirable governments, which in general rather unites than separates Washington and leading information companies, but does not exclude certain sharp disagreements between them.

Conclusion

The threats to the psychological security through MUAI are currently clearly expressed at the second and third levels, which does not exclude them at the first level as AI technologies develop, the scale of their capitalization, and the persistence of acute socio-political contradictions in Brazil. MUAI threats are growing in quantity and quality, and are becoming more diverse. Brazil's entry into BRICS opens up both new opportunities for the country and promises considerable risks, primarily from those forces that do not accept the country's independent course. In the future, a surge in MUAI is possible, since the level of cybersecurity in Brazil leaves much to be desired, legislative initiatives do not keep up with the actual practices of malicious actors, and there is a lack of a systemic approach to countering MUAI in the field of psychological security. At the same time, the clear focus of the Lula government on countering psychological aggression as part of attempts to socio-politically destabilize the country, taking into account the growing role of AI in such attempts may be the initial basis for the development of such a systemic approach in the future.

References

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

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

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

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

municipal-elections-have-big-tech-companies-learnt-anything-from-the-january-8th-attacks/. Accessed 02 Feb 2024

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

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

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

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

Hao K (2019) The biggest threat of deepfakes isn't the deepfakes themselves. In: MIT Technology Review. https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/. Accessed 29 Jan 2024

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

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

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

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

Mari A (2022) Brazil stagnant in tech investments and innovation. In: ZDNet. https://www.zdnet.com/article/brazil-stagnant-in-tech-investments-and-innovation/. Accessed 29 Jan 2024

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The Spamhaus Project (2022) The Top 10 Worst Botnet Countries. In: Spamhaus.org. https://www.spamhaus.org/statistics/botnet-cc/. Accessed 29 Jan 2024

Thomas D (2020) Deepfakes: A threat to democracy or just a bit of fun? In: BBC News. https://www.bbc.com/news/business-51204954. Accessed 29 Jan 2024

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

Ünver H (2023) The role of technology: new methods of information manipulation and disinformation. In: ResearchGate.

https://www.researchgate.net/publication/373445537_THE_ROLE_OF_TECHNOLOGY_NEW_METHODS_OF_INFORMATION_MANIPULATION_AND_DISINFORMATION. Accessed 12 Dec 2023

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

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

The Malicious Use of AI: Challenges to Psychological Security in the Kingdom of Saudi Arabia

Vitali ROMANOVSKI

Introduction

Crown Prince Mohammed bin Salman's strategic priority is the structural overhaul of the Kingdom's economy. For these purposes, priority is given to stimulating innovation, concentrating resources on R&D aimed at producing high-technology products, introducing advanced production technologies, developing high-tech and knowledge-intensive sectors with high added value, and speeding up integration of the Kingdom into the global economic space.

Indeed, Saudi leadership focus on AI development has allowed to create a distinctive AI landscape. Established in 2019 The Saudi Data and Artificial Intelligence Authority (SDAIA) is owner of the Kingdom's AI agenda "mandated with unlocking the value of data and AI to elevate Saudi Arabia as a pioneering nation among the elite league of data-driven economies" (OECD 2024). A branch of the SDAIA, the National Center for Artificial Intelligence (NCAI) takes responsibility for AI innovation, capacity building and promoting the National Strategy for Data and AI. Moreover, Riyadh is pioneering the promotion of ethics in advanced technologies with its newly established International Center for Artificial Intelligence Research and Ethics.

At the same time, the Kingdom does not stay away from the growing global tendency of contemplating AI-associated risks and challenges. There is growing concern among the Saudi decision-makers, industry leaders, experts and academia that the AI immense transformative potential needs careful and nuanced approach to avoid unforeseen consequences of its use in various spheres (Saudi Broadcasting Authority 2023). This chapter attempts to provide a brief on the potential malicious use of AI and challenges to psychological security in Saudi Arabia through the optics of the 'three-level of MUAI threats to psychological security' model.

The First Level of MUAI Threats to Psychological Security

Researches dealing with the psychological security aspect of MUAI identify "deliberately distorted interpretations of the circumstances and consequences of AI development for the benefit of antisocial groups" as the first level of MUAI threats to psychological security (Pashentsev 2023).

Around 70% of the Kingdom's Vision 2030 reform goals are directly linked with the data and AI development strategy, that aims at bringing the Kingdom to the top 15 AI nations by 2030. Recent evidence of Riyadh's continuing focus on developing its AI capacity is the June "2023 Global AI" index that ranked Saudi Arabia first in the Government Strategy parameter and thirty-first globally (Alarabiya 2023). Moreover, Saudi Arabia is among the world's leaders in adopting AI in financial services, according to November 2023 Finastra's annual global survey "Financial Services: State of the Nation Survey 2023", revealing that 55% of the Saudi Arabia respondents (the highest globally) say they have deployed or improved AI in the last 12 months (Asharq Al-Awsat 2023).

Given the level of attention given by the Saudi authorities and Mohammed bin Salman personally to the AI development in the country and the KSA mediasphere strict regulation policies, it seems natural to witness little statistically significant evidence in the Saudi open sources that could back the

assumption on the planned or ongoing information campaigns to discredit Riyadh's efforts to develop its AI capacities.

However, the possibility of distorted interpretations of AI-related issues in the KSA is on the table, particularly because of the intensifying competition between the US and China in cutting-edge technologies sphere and Riyadh's expected efforts to balance between Washington and Beijing. For example, a recent Financial Times publication warned that Saudi-Chinese collaboration in AI, and technologies transfers in particular, could threaten the access of the King Abdullah University of Science and Technology to advanced US-made chips (Kerr et al., 2023). That is an important signal to Riyadh given the Kingdom's ambition to become the regional leader in AI development able to "build large supercomputers and roll out LLMs, the technology that underpins generative AI systems as chatbots" (Ibid.).

In general, given the Saudi leadership's goal of becoming a regional AI development leader, the possibility of public disapproval of Riyadh's efforts in this sphere persists and could potentially rise.

The Second Level of MUAI Threats to Psychological Security

AI used as a specific targeting technology, for example, through the cyberattacks against critical infrastructure, stands for the second level of MUAI threats to psychological security (Pashentsev 2023).

Indeed, Saudi recent advancements in cybersecurity reveal a growing Riyadh's attention to the matter. Riyadh "has prioritised cybersecurity as a pillar of its economic development, implementing major initiatives to raise the levels of cybersecurity readiness" (Arabian Business 2022). A notable example is the latest at the time of writing ITU Global Cybersecurity Index 2020 ranking that put the Kingdom first in the Arab States region and second globally, with the same score as the United Kingdom (ITU, 2020). It is noteworthy that the Saudis achieved such remarkable results in less than 10 years (Tsukanov 2024).

Research conducted in December 2023, which surveyed 50 Saudi-based cybersecurity and IT leaders, revealed that 40% of cyberattacks on organizations in the Kingdom were successful in the past two years (Clewlow 2023). Another research highlights that the most affected sectors of targeted cyberattacks are retail trade, e-commerce, information services, telecommunications, finance, insurance, commercial banking, public administration (SOCRadar 2023), i.e. key public infrastructure sectors that the country residents deal with daily. The same report underlines that the Kingdom, being a regional political and global economic powerhouse with one of the largest oil reserves in the world, is "especially at risk of cyber-attacks targeting critical infrastructure such as oil and gas fields, power plants, and transportation hubs, given the region's crucial role in energy production" (Ibid.). It seems logical to assume that simultaneous, potentially AI-supported targeted cyberattacks, even if partially successful, could pose a threat to the psychological security of the residents of a particular area, if public infrastructures are affected, or even the entire country, if critical infrastructures are disrupted.

AI-associated risks to the country's critical infrastructure are already on the agenda at the senior leadership level. In particular, in an address at the Global Cybersecurity Forum in Riyadh in November 2023, Aramco CEO Amin Hassan Nasser highlighted the need to identify the risks and vulnerabilities associated with the generative AI, "being a game-changer for many industries, including energy", and warned that the energy sector is susceptible to cybersecurity attacks by new technologies (Barakati 2023). However, is yet to be seen whether in the nearest future the KSA will be able or 'allowed' to develop its own AI-supported cybersecurity shields to protect its infrastructure from targeted AI-supported cyberattacks: despite the outstanding efforts to improve the Kingdom's cybersecurity standing, Riyadh still prefers to purchase foreign ready-made cybersecurity packages rather than focus on developing its own solutions (Tsukanov 2024),

The Third Level of MUAI Threats to Psychological Security

The third level of MUAI threats to psychological security are those that are primarily aimed at causing damage in the psychological sphere or establishing control over the public consciousness (Pashentsev 2023).

With the signature of the Bletchley Declaration on 1 November 2023, KSA's attention to this level of MUAI threats has been elevated. That policy paper, which was signed by 28 states and the EU, noted unforeseen risks stemming from the AI capability to manipulate content or generate deceptive content (Gov.uk 2023). Moreover, the represented countries pointed out that "frontier AI systems may amplify risks such as disinformation" and underlined "potential for serious, even catastrophic, harm, either deliberate or unintentional, stemming from the most significant capabilities of these AI models" (Ibid.).

In fact, Saudi focus on this type of threats is natural since AI-generated fake news has become a global phenomenon, and the Kingdom doesn't stay away from this trend (Jones 2019; Fusco 2022; Sumsub 2023). Fusco (2022) highlights that "the prevalence and impact of AI-generated fake news in Saudi Arabia requires attention as it has the potential to create significant harm to individuals and society as a whole, including inciting violence, spreading hatred, and eroding trust in institutions".

Another issue lies in the increasing demand of using Arabic chatbots in the Middle East and Saudi Arabia in particular. Even though the Arabic language morphological features, its complexity, and a possibility of confusing the chatbot with a slight change to a single Arabic word prevents a wide use of chatbots, for example, in Saudi research institutions (Almurayh 2021), the problem of purposeful data-poisoning of Arabic language chatbots will likely manifest itself as a high-feasibility risk in information and cognitive warfare sphere in the near future provided the use of Arabic chatbots becomes a general trend in the regional media.

The rapid dissemination of AI-generated fake news and disinformation pose a serious challenge to the KSA government that is struggling with the rampant spread of such content and the time pressure to work out respective policies to ensure the reliability of official or approved news sources for the domestics and foreign audiences. For now, the government seems to opt for a balancing policy between the prohibitive norms and the legal framework to support the development strategy course. Nevertheless, in view of the growing attention to the AI-associated risks and challenges, it seems natural to expect from the government a stricter regulation aimed at the AI-generated media content.

Conclusion

Despite promoting the AI agenda, Saudi leaders seem to paying a growing attention to the risks and challenges associated with the AI development. Unlike the UAE, whose reputational assets are considered by the government to be extremely important for maintaining the investment attractiveness of the UAE high-tech sector, for Saudi Arabia the most important priority in the psychological security aspect of the national security is to preserve the reputation of a state capable of providing the safety of its technological assets. This approach focuses on preserving the efficiency of the Kingdom's leadership course in implementing Vision 2030.

Technological security and safety of its technological assets, particularly AI, are essential for Riyadh's state development priorities in BRICS, which involve intensifying bilateral trade and using new mechanisms of mutual settlements. The latter takes on particular importance given the Kingdom's interest in getting leadership positions in contributing to the regional FinTech market (Al-Baity 2023).

The initial findings indicate that the second and third levels of MUAI threats pose the most significant risk to the psychological security of the KSA, with potential for the emergence of the first

level of threats. A more comprehensive and nuanced model of AI-associated psychological security threats in Saudi Arabia could be achieved through additional research.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tsukanov L (2024) «Po obe storony Persidskogo zaliva»: razvitiye vysokotekhnologichnogo biznesa v regione i interesy Rossii [“On both sides of the Persian Gulf”: development of high-tech business in the region and Russian interests]. In: PIR Center. https://shorturl.at/ckFJO. Accessed 02 Feb 2024

The Malicious Use of AI: Challenges to Psychological Security in the People's Republic of China[4]

Evgeny PASHENTSEV, Darya BAZARKINA, Ekaterina MIKHALEVICH, Nelson WONG

Introduction

The Chinese case is of interest for several reasons. First, the country's growing leadership in the field of AI has not only generated more malicious use of artificial intelligence (MUAI) threats for China, but also greater opportunities to counter them. Second, the planned nature of the Chinese economy—a developed system of public–private partnerships, including the field of cybersecurity—and the largest population in the world supply colossal amounts of big data for AI training make China's experience in the fight against MUAI truly unique (Bazarkina et al 2023). China has to date made enormous contributions to AI research and development. The AI Index Report 2022 shows that China has generated 27.6% of the world's AI conference publications, while the United States produced only 16.9% in 2021. Similarly, in terms of AI-related patents, China is also currently leading the world, filing 51,69% of the world's AI patents and being granted about 6%, while the shares of the EU and UK and the USA count 3,89% and 16,92 respectively (Zhang et al 2022). China leads in AI adoption, with 58% of companies already deploying AI while 30% are in the process of considering its integration. In comparison, the United States has a lower adoption rate, with 25% of companies using AI and 43% exploring its potential applications (Haan and Watts 2023).

General public consensus in China believes that the country should still embrace the technological advancements offered by AI. At the same time, they also believe that the malicious use of which must be properly regulated in order to protect the country's security and make sure that people's privacy is respected. Fortunately, a large majority of Chinese have faith in their government to take timely and necessary actions to prevent the malicious use of AI. Arguably, such public trust will-put China ahead of many countries in reaching the fine balance of encouraging the fast development of AI while safeguarding the country's public safety (Xu 2022). However, given the deterioration of the international political situation, growing cybersecurity challenges and expanding threats of malicious use of AI by adversarial states, criminal organizations and rogue actors, greater vigilance of China's national security is imperative.

The First Level of Threats of MUAI Against Psychological Security

The gradual emergence of China to a leading position in the world, including and relying on the sphere of high technology, has become a dominant reason by the West to discredit the development of AI in China, and the country as a whole. This situation was personified by a statement by, George Soros at the World Economic Forum in 2019. The wealthy neoliberal warned that open societies face

[4] The contributors of the current chapter express their gratitude for the collection and processing of materials on the topic of this study to Bo Peng, Director of Programmes and Researcher, Shanghai Centre for RimPac Strategic and International Studies; Serene Chen, Assistant Researcher, Shanghai Centre for RimPac Strategic and International Studies; and Nikita Tarasov, PhD student at the Applied Mathematics and Control Processes Faculty Faculty of Saint Petersburg State University. In the current publication are used the materials of the article: Pashentsev, E., Blekanov, I., Mikhalevich, E., Wong, N. (2024). Malicious Use of Artificial Intelligence and Challenges to Psychological Security in China. Vestnik St. Petersburg University. International Relations, No. 2.

"mortal danger" from high-tech authoritarian regimes, and then went on to say "China isn't the only authoritarian regime in the world, but it's undoubtedly the wealthiest, strongest and most developed in machine learning and artificial intelligence. This makes Xi Jinping the most dangerous opponent of those who believe in the concept of open society" (Watts 2019). This accusation can well be considered a realization of the psychological security (PS) risks and threats of the first-level (in MUAI the state itself is accused against citizens).

This general line to discredit the development of AI in China has wide variability in specific issues. For example, in the United States and other Western countries, there are many publications accusing China of total surveillance and persecution of the ethnic minority, the Uyghurs, using AI (Taddonio 2019; Bhuiyan 2022). Meanwhile, AI technologies in China are widely used as part of predictive policing (something widely practiced in the USA), for solving past crimes or preventing future crimes, regardless of the nationality of the criminal (Mantello, 2017). However, there are also more "objective" publications that do not focus on the Uighurs, but talk about the total threat of AI in the hands of an "authoritarian" ("communist", etc.) dictatorship" (Singman 2023; Kasperowicz 2023; Lanum 2023; Raasch and Sahakian 2023; Hauf 2023). In the USA, total surveillance based on AI technologies are developed no less than in China, and their use raises big questions. According to Virginia Doellgast, a professor of employment relations at Cornell University, "[w]orkers are being constantly under automated forms of surveillance, and AI-based monitoring tools can make mistakes that can translate into unfair pay cuts or firings. Workers in developed countries such as the US and Japan often don't know what monitoring tools are being used, what data the tools are collecting or how that data is used to evaluate their performance" (Greenhouse 2023). However, instead of professionally discussing the real benefits and risks of implementing automated surveillance systems (as they exist all over the world), the US prefers a line of propaganda confrontation with the aim of damaging both the AI industry in China (along with numerous sanctions against Chinese AI companies), and in general the interests of the socio-economic development of this country.

MUAI threats at the first level also include widespread attempts by Western mainstream media to sow doubts about China's ability to develop AI technologies under sanctions, to convince Chinese AI developers that it is impossible to work successfully under the dominance of the Chinese Communist Party (CCP), and to sow doubts among foreign buyers about the quality/safety of AI products from China, etc.

The Second Level of Threats of MUAI Against Psychological Security

At the second level of threats for PS by MUAI in China concerns rapid adoption of AI in management systems (Wang 2023). Numerous infrastructure facilities, such as robotic self-learning transport systems with centralized control based on AI, can become convenient targets for high-tech terrorist attacks. If, for example, antisocial actors seize control of the transport management system of a large city (or other critical infrastructure such as a power plant, railway line, television tower, etc.), this can lead to numerous accidents and casualties, cause panic and confusion, thereby creating a precarious psychological climate facilitating a dangerous situation and possibly hostile reactions (Bazarkina and Pashentsev 2019).

Examples of the implementation of the borderline (between the second and third) level of PS threat associated with MUAI are provided by the practice of phishing and social engineering. According to the joint report 2023, private cyber intelligence companies such as Group-IB, Bridewell and the SideWinder hacker group are using a new attack infrastructure to launch targeted cyber strikes against targets in Pakistan and China. According to researchers, hackers registered 55 domains imitating various organizations in the fields of news, government, telecommunications and finance. The above-

mentioned domains created by attackers imitate government organizations in Pakistan, China and India. Many of them contained “trap documents” on government activities. They are designed to download the next stage payload to the target device (SecurityLab 2023).

In January 2020, a Hong Kong bank manager was the victim of a highly advanced heist, in which he was directed to transfer 35 million US dollars to various bank accounts for a company acquisition. The voice on the other end of the line sounded exactly like a familiar business associate, but it was actually an AI-generated clone doing the talking. This incident involved as many as 17 attackers working together, using fake emails to verify the purchase (Veldkamp 2022).

Another example in Hong Kong occurred in January 2024 when a finance worker at a multinational firm was tricked into paying out $25 million to fraudsters using deepfake technology to pose as the company’s chief financial officer in a video conference call. According to the Hong Kong police, the elaborate scam saw the worker duped into attending a video call with what he thought were several other members of staff, but all of whom were in fact deepfake recreations (Chen and Magramo 2024). Acting superintendent of the Cyber Security and Technology and Crime Bureau in Hong Kong, told the city’s public broadcaster, Baron Chan Shun-ching told the city’s public broadcaster RTHK, "I believe the fraudster downloaded videos in advance and then used artificial intelligence to add fake voices to use in the video conference" (Sharma 2024). It all started when the clerk received a seemingly legitimate message from the company's financial officer. The message invited them to a confidential video call to discuss crucial transactions. At first, the financial worker seemed suspicious as the message said that a secret transaction needed to be carried out. But, when he got on the call with his colleagues, who seemed very real, his suspicions were allayed. The people on the call looked and sounded exactly like his colleagues. Only after the clerk spoke with the company's head office later did they realize the entire affair was an intricate scam. We can see from this case that fraudsters that the barriers to entry for AI technology is increasingly lower. The case is one of several recent episodes in which fraudsters are believed to have used deepfake technology to modify publicly available video and other footage to cheat people out of money. The police of Hong Kong made six arrests in connection with such scams. On at least 20 occasions, AI deepfakes had been used to trick facial recognition programs by imitating the people pictured on the identity cards, according to police (Chen and Magramo 2024). The above example shows how fishing and pretexting, two common but distinct social engineering techniques overlap and create a more convincing and dangerous manipulation product based on the use of AI technologies. Moreover, the open-source nature of many AI applications means that malicious use of the technology is not simply a concern related national security. Rather it has quickly broadened to include ordinary citizens who must be ever more vigilant in their business and financial dealings.

The Third Level of Threats of MUAI Against Psychological Security

At the third level of threats to PS in China, the malicious use of *deepfakes* is recognized as one of the significant threats. Deepfake refers to a set of AI technologies for creating or changing audio, video, photo content and machine-generated texts.

An illegal and profitable underground deepfake production chain has quickly emerged in China. In May 2023, the Cybersecurity Bureau of the Ministry of Public Security in northwestern China's Gansu province uncovered a case of AI technology being used to create and spread false information and detained a criminal suspect. According to the police, he modified and edited the collected news items using the popular AI software ChatGPT (which required him to bypass the “Great Firewall”), and then used the software “Seal Technology” to upload his “news” to the Baijia account, which he acquired for illegal profit. The information in the fabricated article, “A train hit a road construction worker in Gansu

this morning, killing 9 people," was patently false and untrue. The Internet Security Police found that a total of 21 Baidu accounts published the article, which received 15,000 views in a short period of time (Gansu Public Security Bureau 2023). This is one of the first enforcement actions under China's recently passed laws regulating the use of deepfakes.

Concurringly, AI-produced "spoof" videos of celebrities and social media influencers have become commonplace. For example, a fake video of Indonesian President Joko giving a speech in fluent Chinese recently circulated on Indonesian social media attracted a great deal of attention. When the Indonesian president talks, the fake video convincingly mimics his mouth shape, facial expressions, voice tone and speech patterns. The artificially generated doppelganger looks very much like him, and the background sound of the video is accompanied by laughter from the audience. The Indonesian government department recently clarified this, pointing out that the video was produced using deepfake technology (Zhang 2023).

Deepfake is an ideal tool for disinformation campaigns because it can generate credible fake news that takes time to debunk. At the same time, the damage caused by fake news, especially those that affect people's reputations, is often long-term and irreversible. All this leads to a default of public trust and suspicion – a situation where society is so accustomed to constant deception that it tries to filter all information received and not trust even official sources, which, in the context of serious internal and external problems, can pose a growing threat to socio-political stability in China.

Another emerging threat is conversational AI. Otherwise known as social robots, bots or simply, chatbots, conversational AI is becoming another threat to the PS through the MUAI. Chinese consumers have long been accustomed to communicating with chatbots and, when, November 30, 2022 OpenAI (one of its founders is Elon Musk), with the support of Microsoft, launched ChatGPT with more advanced capabilities, this caused a strong response in China. In March 2023, an even more advanced multimodal GPT4[5] model from the GPT (Generative Pre-trained Transformer) family of language models came to market.

The first of similar Chinese services was presented by the Chinese search engine Baidu in March 2023. The service was called ERNIE (Enhanced Representation through Knowledge Integration) and has 550 billion different facts. In terms of its capabilities, ERNIE is close to the GPT4 neural network, presented by OpenAI a few days earlier, and in some respect it surpasses it. Tencent and Alibaba have focused more on AI products for business partners, but both offer chatbots to the public in China (Cheng 2023). The GPT4 System Card recognizes that the model can compete with human propagandists in many areas, especially when paired with a human editor (OpenAI 2023). The China Daily, a news outlet owned by the Chinese government, warned that ChatGPT could "strengthen the propaganda campaigns launched by the US" (Schuman 2023).

As agent provocateurs whose speed, scale, and ease of use excel far beyond the scope of traditional forms of propaganda and human troll farms, Chatbots pose a real threat to psychological security and political stability. These facts explain the fast operational response taken by the Chinese leadership to regulate their use. By blocking ChatGPT and other bots supported by Microsoft, Chinese regulators are providing domestic companies with support in developing relevant technologies. The White Paper on the Development of Beijing Artificial Intelligence Industry, published by the Beijing Municipal Bureau of Economics and Information Technology on February 13, 2023, declares the goal of fully strengthening the foundation for the development of the AI industry by 2023, including by supporting leading enterprises in creating technology like ChatGPT (The People's Government of

[5] If previously users could only interact with the neural network using text messages, GPT-4 opens up the horizons of interaction through images, audio and video.

Beijing Municipality 2022). Thus, China is not afraid of technological progress, but strives to prevent the antisocial use of its achievements.

NewsGuard, a North American company that monitors and studies online misinformation, has found that AI tools are being actively used to create so-called *"content farms*," referring to low-quality websites around the world that produce massive amounts of clickbait articles to optimize revenue from advertising. In April 2023, NewsGuard identified 49 websites in seven languages – Chinese, Czech, English, French, Portuguese, Tagalog, and Thai – that appear to be entirely or primarily created by AI language models designed to mimic human communication – in this case, language patterns are disguised as typical news sites (Sadeghi and Arvanitis 2023). Arguably, NewsGuard's discovery is likely just the tip of the iceberg. The power of increasingly advanced generative AI and large language models can turn such sites into effective and relatively low-cost conduits of propaganda.

China is now grappling with a growing wave of fake news accounts and AI-generated posts. In mid-May of 2023, the National regulator Cyberspace Administration of China (CAC) said it had already "cleared" more than 107,000 fake news accounts and 835,000 pieces of false information and urged citizens to report fake news accounts and stories (Frank 2023). Importantly, they were not just talking about text messages, but also virtual hosts, fake studio scenes that allow you to imitate existing registered sites, and, using a variety of methods aimed at the emotional response of network users, increase traffic. Such activities can also have a clearly malicious nature (Dobberstein 2023). Spontaneous use of generative AI by individual users and small firms can mask large-scale campaigns of psychological influence by individual large state and non-state actors. Due to acute geopolitical and economic contradictions, China may objectively be at the epicenter of such campaigns.

It is important to note that many MUAI threats are still nascent. For example, the concept of a *"metaverse"* may open up many new opportunities for the economic and social development of China in the near future. Chinese tech companies have begun testing the waters by developing their own metaverse applications and investing in metaverse-related technologies. The metaverse is a virtual world that exists parallel to the physical world. In the metaverse, greater overlap between our digital and physical realities (in the realms of work, socialization, and entertainment) is possible – enabled by certain advanced technologies, including AI technologies, that may shape the next generation of the Internet. Six of China's leading technology companies, including Baidu Inc, Alibaba Group Holding Ltd and Tencent Holdings Ltd (collectively known as BAT), were among the top 10 firms worldwide that filed the most patents related to the development of critical metaverse technologies (Interesse 2022). Technologies such as augmented reality and the metaverse will present the scene of events in a more holographic and visual way. The highly immersive and interactive character of the metaverse will, in all likelihood, make its audience more susceptible to AI generated mis/disinformation. Thus, it will pose a higher threat of malicious impact on the public (Zhang 2022).

Leading the way in technology, China has a lot to do in terms of providing PS in the metaverses, given the anti-social actors who will undoubtedly try to exploit this new space for their own interests. It should be emphasized that the development of AI-based technologies, and the possibility of their use for malicious purposes by antisocial actors, dictates the need for a more careful study of the potential of these technologies and the development of a systematic approach to the tasks of neutralizing them.

Conclusion

In China, there is a gradual transition from a strategy of catching up to a strategy of advanced development (not least in the field of AI) with great opportunities, but also great risks for making innovative solutions. In a difficult and unstable environment, PS threats are extremely dangerous,

especially if the possession of new technologies provides new opportunities for malicious actors, from criminal organizations and corrupt elements of the governing apparatus to unfriendly aggressive states.

Currently, China is mainly faced with cases of MUAI of the first (numerous attempts to discredit Chinese AI) and third (malicious use of deepfakes, chatbots, news farms, etc.) PS threat levels. Border threats (between the second and third levels) are expressed mainly in phishing and social engineering. Threats to infrastructure facilities and control systems at the second level have not yet led to major incidents as a result of MUAI with corresponding negative psychological and social consequences, but this cannot be ruled out in the future due to the activities of high-tech internal and external malicious actors. The development of metaverses, improvements to emotional AI, and advances in general AI, as well as progress of human sciences and their applied applications will pose even more complex problems in the field of PS.

While MUAI represents a significant threat to China's PS, traditional forms propaganda are renovating their potency through AI technologies. However, in the near future, due to the quantitative and qualitative improvement of AI capabilities and its further implementation in various spheres of public life, MUAI may come to the fore.

All of the above requires a dialectical approach to assessing the role of AI. The adequate use of AI technologies gives a powerful impetus to all social development, as modern China clearly demonstrates. However, with the development of these technologies, the threats of MUAI also grow (sometimes outpacing them), the neutralization of which requires not so much technological or administrative measures as social ones. Unique technologies that are aimed at partially or completely replacing a person in various forms of cognitive emotional and psychological activity (for example, emotional AI in working with the disabled), monotonous or harmful physical work (AI robots), will increase threats to humanity. It will happen if the latter will direct its growing capabilities due to AI not to build a more harmonious social system and the development of an individual with a qualitatively new level of intellectual and psychophysical capabilities, a higher level of social responsibility, but to strengthen social, national or racial polarization. China has enormous scientific, technical, economic, and, most importantly, human potential to prevent negative scenarios in the interests of its national development and any contributions to mankind. The question is how effectively China will use this potential. This question remains open, the main opportunities and risks are still ahead.

References

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

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

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

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

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

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

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

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

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

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

Haan K, Watts R (2023) 24 Top AI Statistics and Trends in 2024. In: Forbes. https://www.forbes.com/advisor/business/ai-statistics. Accessed 08 Apr 2024

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

Haan K (2023) 24 Top AI Statistics and Trends in 2024. Forbes Adviser. https://www.forbes.com/advisor/business/ai-statistics/#sources_section Accessed 08 Feb 2024

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Veldkamp D (2022) Cyber Awareness 2022: Consider Deepfakes, NFTs, and More. In: InfoSystems. https://infosystems.biz/cybersecurity/cyber-awareness-2022-consider-deepfakes-nfts-and-more/. Accessed 29 Mar 2024

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

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

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

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

The Malicious Use of AI: Challenges to Psychological Security in the Republic of India

Darya Bazarkina, Evgeny Pashentsev

Introduction

India's AI market size is projected to expand significantly, reaching an estimated $6,358.8 million by 2025 (Srivastava R 2023). The Indian government announced in February 2018 that the National Institution for Transforming India (NITI Aayog), a government think tank, will lead a national AI research program. In 2017, the Department of Trade and Industry formed the Task Force on Artificial Intelligence for India's Economic Transformation (Faggella 2019). AI is driving massive change in the Indian economy. Indian startups are developing solutions using AI in education, healthcare and financial services. AI development in India also includes work on digital assistants that allow organizations to communicate with customers, AI-based decision-making systems and the use of AI and blockchain in trade (Chakraborty 2022). AI is expected to add up to $500 billion to India's gross domestic product by 2025 and $967 billion by 2035. India's investments in AI are growing by 30.8 per cent annually (Bundhun 2023). "The global trends of pragmatism, open-source models, regulatory focus, job evolution, and efficiency recalibration are setting the stage for transformative changes in AI. At the same time, India's specific growth in the IT and AI sectors adds a unique dimension to this narrative" (Srivastava 2023). Of course, such a rapid development of the field of AI is associated with both benefits and disadvantages, including in the psychological sphere.

The First Level of Threats of MUAI Against Psychological Security

Currently, one of the pressing first-level threats in India remains the manipulation of fears of job loss due to automation. The Microsoft Work Trend Index 2023 reveals 74% of Indian employees fears that AI will replace their jobs. Meanwhile, 83% of employees are willing to delegate a major portion of their workload to AI in a process to mitigate any threat of job loss that puts forward by the technology. The new report is themed around "Will AI fix work," which is gradually becoming a concern following the rollout of tools like ChatGPT, Google Bard, and Microsoft Bing Chat. (Sengupta 2023). This opens up space for attackers to manipulate public opinion, allowing them to present in an extremely unfavorable light those politicians who will actively promote the implementation of AI in various spheres of public life.

There have already been cases of active controversy on social media around job cuts due to AI. An Indian CEO is being criticized after he said that his firm had replaced 90% of its support staff with an AI chatbot. Suumit Shah, founder of Dukaan, said on Twitter that the chatbot had drastically improved first response and resolution time of customers' queries. The tweet sparked outrage online. It comes at a time when there has been a lot of conversation and apprehension about AI taking away people's jobs, specially in the services industry (BBC 2023). Such examples clearly illustrate the existing concerns in Indian society. Increasing advertising expectations for AI is also an issue, but less so than just a few years ago as concerns and fears about AI begin to become more significant. The future will tell whether a balanced attitude towards AI will be achieved.

The Second Level of Threats of MUAI Against Psychological Security

Among the second-level threats that are relevant in India, it is worth noting the increased cases of fraud in the virtual space using AI technologies. The main aspect of such fraud is social engineering, which helps the attacker gain the trust of the victim and force the victim to act recklessly. For example, the authors of phishing messages use the following psychological techniques:

- Urgency: a phishing email usually wants something done right now, as the longer you have to think, the more you may question whether it's legit;
- Plausibility: phishing attempts will be based on real-life scenarios. An invoice needs paying, or a file needs sharing;
- Familiarity: there's been a marked rise in spear phishing, where the attack is at least partially tailored to an individual – often claiming to be from an authority figure such as their CEO or head of security;
- Confidentiality: the action required is specific to you and needs to be done by you alone, as getting someone else involved increases the chances of the scam being spotted (Egress 2021).

AI can significantly increase the "effectiveness" of such attacks, which inevitably affects crime statistics. For instance, a recent report by cybersecurity firm Group-IB exposed that about 100,000 ChatGPT accounts have been compromised, and their data are illegally sold through the dark web, with India alone reporting 12,632 stolen credentials (Stanly 2023). As AI tools become cheaper and more accessible, the risks concerning cybersecurity associated with the technology might increase. The Data from cybersecurity firm CheckPoint Research shows during the first quarter of 2023, India's average weekly attacks increased by 18% when compared to the corresponding period in 2022, with every organization facing a minimum of 2,108 weekly attacks in each organization. However, the number of bot attacks has risen by 48% in the second quarter as against the first one and nine out of 10 websites encountered a bot attack, according to the State of Application Security Q2 report by Indusface (an application security software-as-a-service company) (Stanly 2023). The industry is striving to defend against such threats by developing various solutions to test software quality that can identify and exploit anomalies. However, the attackers have their ways of misusing such technologies. For example, Metasploit is a computer security project that offers information on security vulnerabilities and is one such tool which became popular among hackers (Stanly 2023).

Scammers are increasingly riding the technological sophistication of AI to target millions of Indians who are spending as many as 105 minutes per week reviewing, verifying or deciding whether a message sent through text, email, social media is real or fake. As many as 82% Indians have clicked on or fallen for fake messages, according to a study by Mcafee (Doval 2023). Amongst the sophisticated trickery, the most common forms people fall for include fake job notifications or offers (64%) and bank alert messages (52%). "AI is a scammer's favorite tool, helping cybercriminals increase the scale and sophistication of scam messages. The speed of phishing and text message scams is on the rise - a new phishing site is created every 11 seconds. This spotlights the increasing need for solutions that turn the tables on AI scammers; there has never been a more critical time for the country's 900 million internet users to protect themselves online" (Doval 2023). Online scams have its psychological consequences. It highlights the increased stress people are facing due to the AI driven increase in the number and sophistication of scam messages. This stress could set the stage for social upheaval as trust in security structures declines amid the rise of AI-enhanced crime.

Cybercriminals are now venturing into sophisticated generative AI chatbots like FraudGPT and WormGPT, touted as "bots without limitations, rules, and boundaries", to carry out their nefarious activities. These Chatbots, which recently emerged on the dark web, are being sold to anyone who

wants to create phishing emails, malware or cracking tools. The problem of using such chatbots is also relevant in India. Manish Thakar, a cyber evangelist and vice president IT at Hitachi Hi-Rel, warned people, “These chatbots are based on the popular ChatGPT-3 technology, which can generate realistic and coherent texts from user prompts. Hackers with this tool can create dubious mails to trick unsuspecting victims into believing that they have received an official business email or an SMS or a bank notice” (The Times of India 2023a). He revealed that FraudGPT can write malicious code, create undetectable virus or malware, find non-VBV bins, and create phishing pages and hacking tools to barge into groups, sites and markets. It can even write scam pages or letters, find leaks, vulnerabilities and even access active cards. Gujarat CID officials said that the FraudGPT vendors are quite known in the underground dark web marketplaces like Empire, WHM, Torrez, Alphabay and Versus (The Times of India 2023a). In the future, the number and improvement of such malicious chatbots is likely to increase, which could make social engineering techniques much more dangerous.

Along with the benefits of the digitalization of the Indian economy, there is also an increase in the risk of MUAI. The 2022 budget introduced the central bank digital currency (CBDC), which will be based on blockchain. Although a number of politicians and financiers are in favor of strengthening legislative regulation of cryptocurrencies, Reserve Bank of India Governor Shaktikanta Das believes that CBDC can be subjected to digital frauds as well.

Several fake investment portals have already been on the radar of Indian law enforcement for over a year. Many residents of Kerala were deceived by the creation of an imaginary cryptocurrency called Morris Coin, with Indian law enforcement assessing the amount of the fraudulent assets to be worth 200 million dollars. A similar incident occurred in Karnataka. Experts believe that the combination of fraud and malicious use of deepfakes in the field of cryptocurrencies will take phishing to a new level (Biswas 2022), making it easier to convince consumers of the authenticity of a fake website accepting payments in cryptocurrency.

The Third Level of Threats of MUAI Against Psychological Security

The level of MUAI in the realm of voice control continues to grow, with Pindrop indicating that fraud in this area has grown by more than 350 percent from 2013 to 2017 (Pindrop 2020). At least 39 percent of India’s internet-connected population is likely to own some sort of digital voice assistant. Experts predict that voice will become the preferred transaction method for e-commerce, banking and payments in the near future (Bhatt 2018). In the context of future threats, it is advisable, in our opinion, for interdisciplinary studies to consider the risks of not only voice interface or chat bot hacking, but also the hacking of electronic translation programs. For example, the distortion of official documents by MUAI can be just as provocative as sending a deepfake capable of igniting conflict between countries.

In India, bot activity was recorded during the 2019 election campaigns, with bot accounts on Twitter deployed on a massive scale from February 9–10, spinning hashtags both for and against incumbent Prime Minister Narendra Modi. At the same time, small groups of accounts posted thousands of tweets per hour. The main political parties increased digital communication compared to the previous elections in 2014 but in 2019, the impact of such campaigns was rather modest (Thaker 2019) due to the relatively small number of voters on Twitter. However, the massive scale of bots that were involved in the campaigns reduced the quality of online debates. On December 27, 2021, prominent opposition leader, sent a letter to the CEO of Twitter complaining that his number of followers had not increased since July, unlike the numbers for other political leaders. On this, Twitter replied that they remove millions of bots and malicious accounts using machine learning tools (OpIndia 2022). In this manner, bot activity became the subject of political controversy, which could be used by antisocial actors, creating a minimum psychological security threat of the first level.

Chatbots, along with political campaigns, can be involved in religious activities, where they can also pose a level three threat if used maliciously. India's experience shows a rise in chatbots commenting on religious texts, and their comments (the origins of which are unclear) can be quite provocative. Many people in India are foregoing that in-person contact with a spiritual guru interpreting the Bhagavad Gita and turning to online chatbots, which imitate the voice of the Hindu god Krishna and give answers to probing questions about the meaning of life. It's new technology with the tendency to veer off script and condone violence, according to experts, who warn that AI chatbots playing god can be dangerous. Several of the bots consistently provide the answer that it's OK to kill someone if it's your dharma, or duty. "It's miscommunication, misinformation based on religious text," said Lubna Yusuf, a Mumbai-based lawyer and a co-author of The AI Book. "A text gives a lot of philosophical value to what they are trying to say and what does a bot do? It gives you a literal answer and that's the danger here" (Shivji 2023). At least five Gita chatbots appeared online in early 2023, powered by the language model Generative Pre-trained Transformer 3 (GPT-3). They're using artificial intelligence, which simulates a conversation and creates answers based on statistical probability models. The sites say they have millions of users. "Yusuf said the potential danger of answers that condone violence is more acute in a country like India, where religion is so emotionally charged" (Shivji 2023).

On January 7, 2023, the account of activist Mahesh Vikram Hegde tweeted a screenshot from ChatGPT to its more than 185,000 followers; the tweet appeared to show the AI-powered chatbot making a joke about the Hindu deity Krishna. On questions of faith, ChatGPT is mostly trained to be circumspect, responding "I'm sorry, but I'm not programmed to make jokes about any religion or deity," when prompted to quip about Jesus Christ or Mohammed. That limitation appears not to include Hindu religious figures. When WIRED journalist gave ChatGPT the prompt in Hegde's screenshot, the chatbot returned a similar response to the one he had posted. OpenAI, which owns ChatGPT, did not respond to a request for comment. The tweet was viewed more than 400,000 times as the furor spread across Indian social media. Within days, it had spun into a conspiracy theory on social and broadcast media channels This case, in which ChatGPT's developers' indiscretion led to conflict, can be considered both a manifestation of level one threats (discrediting AI products using opinion leaders) and level three (using the chatbot itself to incite conflict).

However, the most extensive examples of MUAI in India can be found in deepfakes. According to a report by Deeptrace, 3 percent of websites that contained deepfake porn videos in 2019 were Indian (Ajder et al. 2019, p. 2). Deeptrace also notices "a significant contribution to the creation and use of synthetic media tools from web users in China" (Ajder et al. 2019, foreword). There is already a documented case of deepfakes being used in India to damage a reputation. Photos and videos of a controversial Indian journalist were made into deepfake porn videos (Ajder 2019), pointing to the gradual appropriation of AI technologies by modern criminals and the danger of widespread discrimination campaigns being unleashed against various interest groups in the future.

In India, the use case (perhaps the first in the world) of deepfakes specifically for campaign purposes has become infamous. In February 2020, the Bharatiya Janata Party (the Indian People's Party) used this technology to create two videos in which party leader Manoj Tiwari addresses voters in two languages ahead of the Delhi Legislative Assembly elections. The candidate's goal was to send a message to two groups of voters speaking different languages—Haryanvi and English. The videos, according to party representatives, were sent to about 5,800 WhatsApp groups (Alavi and Achom 2020) and in them, Tiwari also congratulated his supporters on the passage of an amendment to the Citizenship Act by the parliament in India. In the original video, Tiwari's message is delivered in Hindi, with his facial expressions and lip movements mimicking the language simulated using AI for the Haryanvi version. The party's media officer, said that the video in Haryanvi received positive feedback, after which it was decided to create an English version. However, it became clear that events could get out of control: "Someone used a Facebook video of our Delhi BJP president…Manoj Tiwari and sent us

his video with changed content in Haryanvi dialect," Delhi BJP Media officer said. "It was shocking for us, as it may have been used by the opposition in bad taste…We strongly condemn the use of this technology, which is available in open arena and has been used without our consent" (Mihindukulasuriya 2020).

Although deepfakes were used in this case solely to overcome a language barrier, the further modification of the deepfakes for malicious purposes—and then the reaction of the party—caused a lively discussion in the Indian media. This understandably fuels concerns about the possible use of technology to spread disinformation, such as politicians using deepfakes to put controversial and inadmissible statements into the mouths of their opponents (Alavi and Achom 2020). The practice of using deepfakes was publicly abandoned by India's major political parties (Mihindukulasuriya 2020). For India, as well as for many other countries, in the current environment of rapid digitalization and increasing disinformation campaigns, the challenge of combining MUAI technologies with different psychological security threat levels may become relevant.

In November 2023, the Indian government has issued a warning about the "dangerous and damaging" implications of AI technology after a deepfake video of Bollywood actress Rashmika Mandanna went viral. A video purportedly showing Bollywood star Mandanna — who has 39 million Instagram followers — wearing black activewear and exiting an elevator went viral on social. However, the footage was actually an AI-generated deepfake. The woman in the six-second video was actually a British-Indian influencer named Zara Patel who had posted the original clip on her Instagram account. According to the BBC, Abhishek Kumar, a journalist who works with the fact-checking platform Alt News, first reported that the viral video seemingly showing Mandanna was a deepfake (Bandara 2023). Bollywood star Mandanna described her distress at the deepfaked video being circulated online and said that the issue of deepfake technology must be addressed with "urgency." Veteran Bollywood actor Amitabh Bachchan argues that his co-star Mandanna had "a strong case for legal action" over the identity theft in the AI-generated video (Bandara 2023). Fake images of actresses Katrina Kaif and Kajol also emerged in quick succession—a pointer to the disturbing fact of deepfakes proliferating.

As Rajeev Chandrasekhar, the Union minister of state for electronics and information technology (MeITY), said, "The phenomenon of deepfakes, which is really a marriage of AI and the misinformation industry, is certainly something that all of us should be worried about because it really is very potent for individuals, for societies, for communities and countries" (Sukumaran 2023). On November 17, 2023, Prime Minister Narendra Modi, warning that deepfakes could lead to a crisis and 'stoke the fire of discontent' in society, revealed how he recently saw a deepfake of himself doing the Gujarati 'garba', a dance he hadn't performed since school. Doctored videos as a psychological tactic, for instance, were in circulation following the Galwan skirmish between Indian and Chinese soldiers in Ladakh in June 2020. India saw an example of facial expression manipulation in May 2023, bang in the middle of a pivotal moment in the five-month-long protest by female Olympic wrestlers pressing for investigation into complaints of sexual harassment. When the Delhi Police detained the protesters, wrestlers Vinesh and Sangeeta Phogat posted a selfie of themselves sitting grimly in a police van with the cops. But it was another version—of both women smiling at the camera—that circulated virally until it was debunked as a fake (Sukumaran 2023). All of these cases clearly illustrate the threat to third-level psychological security posed by the malicious use of deepfakes.

There is already growing concern in Indian society about the psychological consequences of this threat. Alka Kapur, student counselor, principal, Modern Public School, Shalimar Bagh says, "For students, social media is more than just a platform for connection; it's a virtual realm where their identities are shaped, friendships are forged, and self-esteem is cultivated. As deepfake scandals infiltrate social media, the consequences on the mental health of students become increasingly alarming. The digital alteration of content, once perceived as harmless fun, now poses a severe threat

to the psychological well-being of young individuals. The pressure to conform to digitally altered standards of beauty and behavior, coupled with the fear of becoming victims of malicious deepfake manipulation, creates a toxic environment for students" (The Times of India 2023b). Journalists offer us to imagine a scenario where a child finds itself obnoxiously featured in a morphed video, engaging in activities in which they never participated. "The possibility of harm to one's reputation and the subsequent experience of social exclusion can be profoundly damaging. The fear of becoming the next target leaves children in a perpetual state of anxiety, eroding their trust in the digital realm and exacerbating the already challenging journey through adolescence" (The Times of India 2023b).

A political controversy rocked the southern Indian state of Tamil Nadu in April when K. (Kuppusamy) Annamalai, state head of the Bharatiya Janata Party (BJP) — India's ruling party — released a controversial audio recording of Palanivel Thiagarajan, a lawmaker from the Dravida Munnetra Kazhagam (DMK) that is currently in power in the state. In the 26-second low-quality audio tape, Thiagarajan, who was the finance minister of Tamil Nadu at the time, could allegedly be heard accusing his own party members of illegally amassing $3.6 billion. Thiagarajan vehemently denied the veracity of the recording, calling it "fabricated" and "machine-generated." "Never trust an Audio clip without an attributable source," Thiagarajan tweeted on April 22, 2023. He argued that it's now easy to fabricate voices. On April 25, Annamalai released a second clip — 56 seconds long, and with much clearer audio — where Thiagarajan allegedly spoke disparagingly of his own party and praised the BJP. This time, Thiagarajan said no one had claimed ownership of the source of the clips. The analysts were divided on the first clip, either finding it too poor in quality to come to a conclusion, or judging that the clip was "very likely fake." However, they all agreed on the second clip, deeming it authentic (Christopher 2023). The inconsistency and ambiguity of this case has become the subject of speculation in India that in the era of total malicious use of deepfakes, unscrupulous politicians can declare any incriminating evidence, even genuine, fabricated. One way or another, such cases reveal the other side of the problem of deepfakes – a decline in trust in information sources in general.

Conclusion

Currently, in India, threats to psychological security exist at all three levels, but the threats of the second and third levels are most pronounced. Second-level threats can lead to major man-made disasters. At the same time, the direct physical impact of these disasters on economic, military, and political structures can be supplemented by information and psychological impact (fear, panic and other mass effects), which can cause a secondary blow to these structures through weakening the will, determination, and ability of people to resist the consequences of man-made disasters. catastrophes, and are accompanied by a kind of residual "psychological radiation", long-term traumatic psychological consequences, which can also negatively affect the life of society for a long time. It should be taken into account that many second-level threats are spreading as AI technologies become more widespread and cheaper, on the basis of which attackers are already creating their AI products to produce fraudulent content. Psychological techniques of social engineering are greatly enhanced by technology, which increases the danger of spear phishing. Third-level threats include dangerous chatbots and the rapid spread of malicious use of deepfakes. The systemic nature of threats requires a systemic response, a demand for which has already matured in Indian society.

References

Ajder H (2019) Social Engineering and Sabotage: Why Deepfakes Pose An Unprecedented Threat To Businesses. In: Deeptrace. https://deeptracelabs.com/social-engineering-and-sabotage-why-deepfakes-pose-an-unprecedented-threat-to-businesses/. Accessed 29 Mar 2024

Ajder H, Patrini G, Cavalli F, Cullen L (2019) The State of Deepfakes: Landscape, Threats, and Impact. In: The Register. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf. Accessed 29 Mar 2024

Alavi M, Achom D (2020) BJP Shared Deepfake Video on WhatsApp During Delhi Campaign. In: NDTV. https://www.ndtv.com/india-news/in-bjps-deepfake-video-shared-on-whatsapp-manoj-tiwari-speaks-in-2-languages-2182923. Accessed 29 Mar 2024

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

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

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

Bhatt S (2018) How Indian startups gear up to take on the voice assistants of Apple, Amazon and Google. In: The Economic Times. https://economictimes.indiatimes.com/small-biz/startups/features/how-indian-startups-gear-up-to-take-on-the-voice-assistants-of-apple-amazon-and-google/articleshow/64044409.cms. Accessed 29 Mar 2024

Biswas P (2022) Deepfakes, Crypto Scams on the Rise in India 2022. In: Digit. https://www.digit.in/features/crypto/deepfakes-crypto-scams-on-the-rise-in-india-2022-63740.html. Accessed 29 Mar 2024

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

Chakraborty M (2022) Artificial Intelligence: Growth and Development in India. In: Analytics Insight. https://www.analyticsinsight.net/artificial-intelligence-growth-and-development-in-india/. Accessed 29 Mar 2024

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

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

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

Faggella D (2019) Artificial Intelligence in India—Opportunities, Risks, and Future Potential. In: Emerj Artificial Intelligence Research. https://emerj.com/ai-market-research/artificial-intelligence-in-india/. Accessed 29 Mar 2024

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

Mihindukulasuriya R (2020) Why the Manoj Tiwari deepfakes should have India deeply worried. In: The Print. https://theprint.in/tech/why-the-manoj-tiwari-deepfakes-should-have-india-deeply-worried/372389/. Accessed 29 Mar 2024

OpIndia (2022) “Removed accounts for manipulation and spam,” Twitter says after Rahul Gandhi loses bot followers. In: OpIndia. https://www.opindia.com/2022/01/removed-accounts-for-manipulation-and-spam-twitter-says-after-rahul-gandhi-loses-followers/. Accessed 29 Mar 2024

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

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

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

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

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

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

Thaker A (2019) Automated bots manipulated Twitter traffic before Narendra Modi’s visit to Tamil Nadu: US think-tank. In: Scroll.in. https://scroll.in/article/919445/automated-bots-manipulated-twitter-traffic-before-narendra-modis-visit-to-tamil-nadu-us-think-tank. Accessed 29 Mar 2024

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

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

The Malicious Use of AI: Challenges to Psychological Security in the Republic of South Africa

Darya BAZARKINA, Evgeny PASHENTSEV

Introduction

In South Africa, a great deal of attention is paid at the state level to the benefits that society and the economy can derive from the full implementation of AI. In 2019, President Cyril Ramaphosa appointed the Presidential Commission on the Fourth Industrial Revolution (PC4IR). The PC4IR assists the government with taking advantage of the opportunities presented by the digital industrial revolution, including AI and machine learning (The Presidency, Republic of South Africa 2019). "The commission is composed of representatives of tech startups, academia, cybersecurity specialists, researchers, social scientists, trade unionists, and other representatives from key economic sectors" (Ramafosa 2020), indicating a deep understanding of the impact AI has on all areas of society, studied jointly by specialists in the field of technical and social sciences. According to President Ramafosa, South Africa aims to fully harness the potential of technological innovation by 2030, growing the economy and lifting the people (Ramafosa 2020). In October 2020, the PC4IR recommended establishing an AI institute to foster a generation of new knowledge and AI applications in sectors such as health, agriculture, education, energy, manufacturing, tourism and information, and communications technology, as well as training to ensure positive social impact (PC4IR 2020). On November 30, 2022, The Department of Communication and Digital Technology (DCDT) launched the Artificial Intelligence Institute of South Africa (AIISA) (AIISA 2023). South Africa's AI development has led to the formation of a number of notable companies in the field (GoodFirms 2022), and government initiatives are helping build the infrastructure for further AI adoption.

The First Level of Threats of MUAI Against Psychological Security

For South Africa, first-level psychological security threats from MUAI are relevant. Among these threats is the possibility of social destabilization due to the displacement of an increasing number of workers by AI technologies. Ian Goldin, professor of globalization and development at Oxford University, says that AI could displace millions of jobs in the future, damaging growth in developing regions, such as Africa (BBC News 2022). Many workers in South Africa have already been replaced by robots, according to the McKinsey Institute, especially from mechanization in the historically labor-intensive mining sector. Trade and service jobs are also at risk. For example, the Pick n Pay supermarket chain introduced an AI-enabled automated checkout system that eliminates the need for cashiers (Van den Berg 2018). Against the background of a high unemployment rate, general understanding of the specifics of AI in South Africa is extremely low. Consequently, AI-related anxiety about job loss is rapidly increasing among working citizens (Business Tech 2019).

According to Kaspersky research, employees in South Africa believe that the better robots become at different tasks, the fewer jobs will remain for humans. The majority of local employees surveyed (74%) believe that robots should be more widely used across different industries, however, many fear the hacking of robots. Employees reported an increase in robotization level in their companies over the last two years. 33% of employees from South Africa said their organizations already use robots, and 39% of local organizations plan to use them in the near future (IT-Online 2022). The majority of employees surveyed in South Africa (92%) believe robots will eventually replace humans in

their industry. As robots are advancing in all market sectors, humans need to receive new knowledge and skills to not lose their job to robots. Moreover, they are ready to do so: among those who think that their jobs could be replaced by robots, the majority (75%) are willing to learn new skills or improve their existing skills and expertise. Another important finding was that cybersecurity risks increase because of robotization. The majority of local respondents (89%) believe that robots can get hacked, and 53% know of such incidents within their company or other local companies. Respondents are split in their assessment of how protected robots are: almost half of employees surveyed in South Africa (42%) believe that not enough cybersecurity measures are in place to protect the robots in different industries (IT-Online 2022).

Additionally, first-level threats appeared when the United States attempted to oust the Chinese telecommunications company Huawei from the South African market, which supplies the country with the 5G internet technologies that are necessary for AI development. Huawei is at the forefront of 5G development, and in 2018, the United States launched a campaign to stop other countries from buying Huawei equipment. The president of South Africa, Cyril Ramaphosa, claimed in his opening speech at a digital economy conference in Johannesburg in July 2019 that the United States was "clearly jealous that a Chinese company called Huawei has outstripped them and because they have been outstripped they must now punish that one company." In 2019, Huawei signed a contract with South Africa for the first 5G commercial network on the African continent. According to Cobus van Staden, a Chinese–African researcher at the South African Institute of International Affairs, Huawei has already built around 70 percent of the continent's 4G networks (EFE–EPA 2019).

On February 26, 2020, South African mobile communications company Rain announced plans to build a 5G-transport network using Huawei's optical cross-connect (OXC) and 200G solution, leveraging Huawei's latest all-optical switching product—OXC (P32)—to build a metro-optical transport network. Rain is focused on bringing mobile broadband networks to South Africa and becoming the first operator to deploy 5G networks in the country (Huawei 2020).

On March 12, 2020, at the Health and Human Rights Summit in Tucson, Arizona, Dr. Thomas Cowan hypothesized that COVID-19 may have been caused by 5G (Huawei 2020). Cowan started his career while teaching gardening as a Peace Corps volunteer in Swaziland and South Africa, and later served as vice president of the Physicians Association for Anthroposophical Medicine. A founding member of the Weston A. Price Foundation, Cowan's claims have largely been debunked, but discussions about his hypothesizes have attracted many supporters on online platforms. This is understandable in an environment of growing panic that is actively perpetuated by media and social networks.

Tom Cowan's original hypothesis provoked heated discussion in South African media and social networks that extended beyond the country's borders. More than 4,000 people signed a petition on Change.org to stop the rollout of 5G in Cape Town (Independent 2020), and similar petitions are circulating in South Africa. Access to these sites became restricted by government in April, although this is unlikely to stop supporters of Cowan's ideas.

This will not be an examination of the essence of Cowan's argument, but without a doubt, his assessment of the relationship between COVID-19 and 5G technologies has become a tool in a competitive struggle. In South Africa, Huawei is the undisputed leader in 5G, and active campaigning against 5G in the country is directed almost exclusively at this company. It is unlikely that such an information campaign will be an immediate success, but if COVID-19 cases increase in the country and, most importantly, inevitably affect the socio-economic situation, then other scenarios cannot be excluded.

This case illustrates the path trade wars can take. In this instance, AI and related technologies have become the subject of psychological warfare, which, along with business structures, involves political elites who are not facilitating the resolution of the contradictions of the modern world order. Such incidents are worrisome because this type of warfare involves many non-state actors who, guided by false messages and conspiracy theories, can cause damage physical infrastructure, cause additional panic in society, or discredit technologies aimed at overcoming economic problems.

The Second Level of Threats of MUAI Against Psychological Security

South Africa, like many other countries, has not been immune to the threat of phishing attacks, which have been on the rise in South Africa, targeting both individuals and businesses. “One of the reasons behind the success of phishing attacks in the country is the level of trust people place in familiar institutions. Cybercriminals exploit this trust by using local branding and relevant issues to increase the likelihood of their scam being successful. Common phishing lures in South Africa may include fake banking emails, SARS (South African Revenue Service) tax refund scams, and messages related to popular events or news” (Apex 2023). For instance, according to a message that's been spreading around on WhatsApp, and has been sent to our WhatsApp line several times, the South African government is going to give “all South African parents” a child support grant of R1,100 (about US$59) for six months. The message has also been shared on Facebook—on public groups with thousands of members—with some users asking if the claim is true. There is no “minister of humanitarian affairs and poverty alleviation” in South Africa. Child support grants are administered by the South African Social Security Agency (Sassa), a national agency of the South African government. *Africa Check* searched Sassa’s social media accounts for any mention of the claim and came across a Facebook post published by Sassa on January 12, 2024, with a screenshot of the message stamped “Fake” (Khourie 2024). Steven Powell, head of forensics at ENSafrica, says cyber fraud is costing the South African economy an estimated 2 billion rands a year, which is why companies are spending millions on strengthening their cyber security (Moodley 2023). “Generative AI has new implications for crafting remarkably convincing phishing emails, messages, and websites that mirror legitimate entities to deceive individuals into divulging payment data and sensitive information” (Africa Business 2023), *says Ryan Mer, CEO of eftsure Africa, providing an overview of current threats to South Africa.*

Lawyers arguing a case at the Johannesburg regional court have been called out in a judgement for using fake references generated by ChatGPT. According to the *Sunday Times*, the judgement also delivered the lawyers’ client with a punitive-costs order (Prior 2023). The case in question involved a woman who was suing her body corporate for defamation. The counsel for the trustees of the body corporate argued that a body corporate could not be sued for defamation, and the counsel for the plaintiff, Michelle Parker, said there were earlier judgements that answered the question—they had just not had the time to access them. Magistrate Arvin Chaitram postponed the case to late May to give both parties ample time to source the information they needed to prove their cases. In the two months that followed, the various lawyers involved in the case tried to track down the information the lawyers had referred to. Instead, they found that, although ChatGPT had referred to actual cases and given real citations, the citations related to different cases than the ones named. Additionally, these cases and citations were not at all appropriate for defamation suits between body corporates and individuals. It was then admitted by lawyers that the judgements had been sourced “through the medium of ChatGPT.” Chaitram ruled that the lawyers had not intended to mislead the court—they were “simultaneously simply overzealous and careless.” This meant no further action was taken against the lawyers beyond the punitive costs order. “The embarrassment associated with this incident is probably sufficient punishment for the plaintiff’s attorneys,” said Chaitram (Prior 2023). The case occupies a

borderline position between threats of the second and third levels, since the lawyers did not intend to mislead the court, but misinformation was unintentionally used during the trial.

"In recent months, multiple employers, including global tech giants, have been the subject of "conversational AI leaks". "Conversational AI leaks" is a phrase used to describe a loss of data when a chatbot is involved. These leaks involve incidents where sensitive data/information, which is fed into chatbots like ChatGPT, is unintentionally exposed. When information is disclosed to chatbots, the information is sent to a third-party server, and is used to train the chatbot further. What that means, in simple terms, is that the information input into the chatbot may be used by the chatbot in the future generation of responses" (Boda et al. 2023). This becomes particularly problematic when the chatbot has access to, and is using, confidential information. Between March 2022 and March 2023, the global average cost of data breaches reached an all-time high of 4.45 million dollars, and for South Africa specifically, it exceeded 50 million rands (Boda et al. 2023). This case describes a technical error, but its existence opens the door for attackers.

The Third Level of Threats of MUAI Against Psychological Security

A pressing social problem in South Africa, in the context of third-level psychological security threats, is bullying. When moved to social networks or mobile applications, bullying can be amplified by built-in AI algorithms. According to a 2020 Independent Polling System of Society report, South Africa had the highest prevalence of cyberbullying, with 54 percent of parents indicating that they are aware of a child in their community who has been a victim of bullying (Kahla 2020). Due to the ability of social media algorithms to rank content that attracts a lot of attention, controversial content can be quickly shared by real users, generating a higher rating—and a larger audience—for messages of hate speech. It is known that terrorist organizations have used AI-enabled bots in mobile applications to spread propaganda, potentially increasing the destructive effects of cyberbullying.

In South Africa, manipulated audio recordings have been leaked from meetings of the governing African National Congress (ANC) leadership ahead of the party's elective conference in 2022. "Whoever is doing it, it was very professionally done," said ANC deputy secretary general Jesse Duarte in a radio interview with Eyewitness News on April 15, 2021. "It was cut and spliced very carefully to give a particular impression of a very sober conversation in the meeting of the officials of the ANC" (Maree 2021). "One of the leaks contained a purported recording of Duarte's input at a meeting between the ANC's top six officials and former president Jacob Zuma at the end of last month. The party is trying to persuade Zuma to testify in front of the commission of inquiry into the large-scale corruption during his tenure, dubbed 'state capture,' but he continues to refuse" (Ibidem).

This was not the first time seemingly-manipulated information was featured from the party's internal battles. However, this case of disinformation, probably intended to show voters a split in the ANC's ranks, was called a deepfake in the media. New and easily accessible technologies make the manipulation of information easier and more potent, warns University of Johannesburg vice-chancellor Tshilidzi Marwala. He also notes that the expertise of sophisticated cyber manipulation also exists in South Africa (Maree 2021), and that the malicious use of technologies in political rivalries, such as deepfakes and voice-changing in particular, can threaten a country's democracy.

Deepfakes have skyrocketed across the globe, including in South Africa. Amongst African countries, South Africa (19.7%) and Nigeria (11.5%) have seen a higher number of deepfake attacks compared to other African nations. "Of concern to South Africans is the astounding rise in deepfake frauds by 1,200%," said Hannes Bezuidenhout, Sumsub's VP of Sales for Africa (Fraser 2023). Seen against a rise of 450% in identity fraud for the MEA Region, this poses a significant threat and cause for concern. Bezuidenhout said that creating deepfakes is becoming easier as fraudsters use a person's

genuine document, and extract a photo to create a 3D persona. “Providers lacking continuous efforts to update deepfake detection technologies are jeopardizing businesses and users. Updating these technologies is crucial for modern, effective verification and anti-fraud systems” (Fraser 2023).

Current examples include the malicious use of deepfakes of South African television TV anchors. The South African Broadcasting Corporation (SABC) was compelled, on November 14, 2023, to clarify that their anchors, Bongiwe Zwane and Francis Herd, were impersonated in deepfake videos circulating online. These videos, promoting a fraudulent investment scheme, amassed significant attention, with one featuring Herd garnering over 123,000 views on YouTube since its appearance on November 3 (Women Press Freedom 2023). The deepfakes, which included the SABC News logo and depicted a counterfeit Elon Musk promoting the scam, have raised serious concerns about the impact of such technology on the credibility of news organizations, and the freedom of the press. In response to the scam, Francis Herd and Bongiwe Zwane have publicly denied any involvement in the AI-generated videos. Moshoeshoe Monare, SABC’s Group Executive for News and Current Affairs, denounced the scam, stressing the need to safeguard the reputation of the public broadcaster and its journalists. The incident not only damages the trust in individual journalists, but also poses a broader threat to media integrity, complicating the public's ability to distinguish between real and manipulated content (Women Press Freedom 2023).

While just under half of employees surveyed in South Africa (42%) said they could tell a deepfake from a real image, only 21% could actually distinguish a real image from an AI-generated one in a test, says Kaspersky. This means that organizations are vulnerable to such scams, with cybercriminals using generative AI imagery in several ways for illegal activities. They can use deepfakes to create fake videos, or images, that can be used to defraud individuals or organizations. For instance, cybercriminals can create a fake video of a CEO requesting a wire transfer, or authorizing a payment which can be used to steal corporate funds. Compromising videos or images of individuals can be created to be used to extort money or information from them. Cybercriminals can also use deepfakes to spread false information or manipulate public opinion—55% of employees surveyed in South Africa believe their company can lose money because of deepfakes (IT-Online 2024). Thus, level three threats are constantly growing in South Africa.

Conclusion

The analysis shows that, in South Africa, primarily level one psychological security threats caused by MUAI are manifested, however, as on a global scale, the number of cases of digital fraud in which AI can be used is growing. South Africa represents a promising market for technology companies from other countries, which sometimes results in aggravation of psychological warfare (as in the case of attempts to oust Huawei from the country’s market). The country has already faced the political use of deepfakes, which, against the backdrop of South Africa’s admittedly weak ability to recognize fake media content, exacerbates third-level threats.

References

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

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

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

BBC News (2022) Will AI kill developing world growth? In: BBC News. https://www.bbc.com/news/business-47852589. Accessed 29 Mar 2024

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

Business Tech (2019) How AI is being used in South Africa. In: Business Tech. https://businesstech.co.za/news/enterprise/322505/how-ai-is-being-used-in-south-africa/. Accessed 29 Mar 2024

EFE-EPA (2019) South African president says USA jealous of Huawei. In: www.efe.com. https://www.efe.com/efe/english/business/south-african-president-says-usa-jealous-of-huawei/50000265-4016943. Accessed 29 Mar 2024

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

GoodFirms (2022) Top Artificial Intelligence Companies in South Africa 2022. In: GoodFirms. https://www.goodfirms.co/artificial-intelligence/south-africa. Accessed 29 Mar 2024

Huawei (2020) South Africa's Rain and Huawei Build the First 5G Transport Networks Using OXC+200G Solution. In: Huawei. https://www.huawei.com/en/press-events/news/2020/2/5g-transport-networks-oxc-200g-solution. Accessed 29 Mar 2024

Independent (2020) Watch: Debate raging on link between 5G technology, coronavirus pandemic. In: Independent Online (IOL). https://www.iol.co.za/capetimes/news/watch-debate-raging-on-link-between-5g-technology-coronavirus-pandemic-45124913. Accessed 29 Mar 2024

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

Kahla C (2020) Social media platforms need to take a stand against cyberbullying. In: The South African. https://www.thesouthafrican.com/technology/social-media-stand-against-cyberbullying/. Accessed 29 Mar 2024

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

Maree A (2021) South Africa: Leaks and deepfakes shaping the race for ANC presidency. In: The Africa Report. https://www.theafricareport.com/80648/south-africa-leaks-and-deepfakes-shaping-the-race-for-anc-presidency/. Accessed 29 Mar 2024

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

PC4IR (2020) Report of the Presidential Commission on the 4th Industrial Revolution. In: South African Government. https://www.gov.za/documents/report-presidential-commission-4th-industrial-revolution-23-oct-2020-0000. Accessed 29 Mar 2024

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

Ramaphosa C (2020) A national strategy for harnessing the Fourth Industrial Revolution: The case of South Africa. In: Brookings. https://www.brookings.edu/blog/africa-in-focus/2020/01/10/a-national-strategy-for-harnessing-the-fourth-industrial-revolution-the-case-of-south-africa/. Accessed 29 Mar 2024

The Presidency, Republic of South Africa (2019) President appoints Commission on Fourth Industrial Revolution. In: The Presidency. https://thepresidency.gov.za/press-statements/president-appoints-commission-fourth-industrial-revolution. Accessed 29 Mar 2024

Van den Berg A (2018) Hoe kunsmatige intelligensie die werkwêreld gaan verander (How artificial intelligence will change the world of work). In: Jou Werk / Solidariteit Wêreld. https://jouwerk.solidariteit.co.za/hoe-kunsmatige-intelligensie-die-werkwereld-gaan-verander/. Accessed 29 Mar 2024

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

The Malicious Use of AI: Challenges to Psychological Security in the Russian Federation

Darya BAZARKINA, Evgeny PASHENTSEV

Introduction

The main areas of digital technology research and development in Russia include machine learning, human-machine interfaces, industrial internet technologies, the use of spatial data (transport networks) and much more. On October 10, 2019, the 2030 National Strategy for the Development of Artificial Intelligence was adopted (President of the Russian Federation 2019a). It is significant that the strategy covers a period of ten years, and the principles of implementation include security: "…the impermissibility of using artificial intelligence for the purpose of intentionally causing harm to citizens and legal entities, as well as preventing and minimizing the risks of negative consequences of using artificial intelligence technologies" (President of the Russian Federation 2020). In Russia, priority is given to psychological security threats through second-level MUAI, although the strategy leaves room for maneuver in regulating the other two levels. In January 2024, President Vladimir Putin ordered the relevant departments to analyze the practice of using artificial intelligence (AI) technologies in the investigation of crimes. The Supreme Court, the Prosecutor General's Office, the Investigative Committee, the Ministry of Internal Affairs and the Ministry of Justice must deal with the issue before July 1. If necessary, they must provide proposals for improving the technology (Uvarchev 2024).

Among other state structures, the Foundation for Advanced Research (FAS) supports research on countering MUAI. The National Center for the Development of Technologies and Basic Elements of Robotics was established at the FAS. FAS supported a competition to develop technology to convert hard-to-recognize (due to background noise or accents) Russian speech into text, resulting in the creation of a unique Russian speech recognition technology. FAS also supports projects for interpreting images obtained from satellites and unmanned aerial vehicles, with MIPT as the main contractor. As part of this project, MIPT is creating technologies aimed at combating terrorism by identifying weapons caches and camouflaged terrorist bases from drone images. Work is also underway on a project to identify threats in social networks, developed by the State Research Institute of Aviation Systems (GosNIIAS) as part of the FAS project. GosNIIAS created a technology that makes it possible to identify wanted persons in a crowd, on public transport and in other difficult conditions.

The First Level of Threats of MUAI Against Psychological Security

From February to April 2020, the State Duma of the Russian Federation considered a draft law to establish an experimental legal regime for the implementation of AI technologies in Moscow. This bill elicited somewhat ambiguous reactions from the media, with both neutral (Interfax 2020) and negative (RIA Katyusha 2020) comments referring to the dangers of infringing on citizens' right to privacy. On April 24, 2020, the bill was signed into federal law (President of the Russian Federation 2019b). In this situation, it is especially important to pay attention to first- and second-level psychological security threats. AI abuse by business entities can occur during the process of collecting personal data; that data can not only be transferred to government bodies (as required by law) but also, for example, be used to send aggressively targeted advertising. Reactions to the law with comparisons made to a "concentration camp" and even the "Chinese model" (RIA Katyusha 2020), suggesting that negative media campaigns against AI practices in both Russia and China could follow a similar pattern. This

indicates that it is all the more important for Russian government agencies to ensure citizens are adequately informed about the uses of AI, especially since services that already use AI elements are beginning to actively influence the lives of Russians. This became widely known when a Tele2 mobile operator robot called a subscriber, but a bot named Oleg from Tinkoff Bank picked up the phone instead. The Tele2 robot offered the Oleg bot a new rate; Oleg agreed without the smartphone owner's consent (Gavrilyuk and Korolyov 2022). Despite the desire of businesses to limit AI regulation through the AI Code of Ethics, the need for adequate legislation matching the level of technology development is obvious.

Among first-level threats, manipulation of fears of job loss due to the introduction of AI is relevant. For instance, in the Russian Federation, AI tools are beginning to be actively implemented, for example, in the book market. Thus, Stroki and LitRes services are already using AI to dub audio books. This has already caused dissatisfaction among professional speakers: their union proposed to the State Duma to establish regulation of voice synthesis using AI (Yurasova et al 2023). According to a study by the educational platform GeekBrains (the end of 2022), more than half of Russians – 60% – know what AI is, but only 14% of them fully trust the technology. More than 2,000 respondents aged 18 to 55 from various regions of Russia took part in the survey. Analysts found that most of the concerns of respondents about AI are related to the fear of job cuts due to the development of technology. Thus, 58% of respondents are afraid that AI may take their jobs. 11% of respondents are confident that their professions may completely disappear due to technology, and 46% of them have no doubt that AI will be able to partially take over their functionality (Mamikonyan 2022). This is a relatively new phenomenon in Russia, where people generally do not view AI as a competitor. A poll conducted by the Sberbank Life Insurance company for Sputnik ahead of Sberbank's AI Journey 2023 international conference revealed a sharp break along gender lines, with 61% of men saying AI would improve lives, compared with just 39% of women (Sputnik Africa (2023).

The Second Level of Threats of MUAI Against Psychological Security

The MUAI threats (existing and future) that have manifested in other BRICS countries are also relevant for Russia. This was demonstrated during the COVID-19 pandemic, with phishing cases becoming more frequent in Russia. Against the backdrop of news about benefits payments to families with children, fake sites began to appear asking people to apply for benefits. In the .ru zone, about thirty fake domains were found in 2020. According to Alexei Drozd, head of the SearchInform security department, many sites were not yet complete, probably preparing to "mirror" their designs on the original, official site (Stepanova 2020). According to a survey conducted by Avast, 45 percent of Russians experienced phishing attacks in 2021, an increase of 4 percent compared to the company's 2020 results. Additionally, 72 percent of respondents received phishing calls compared to 56 percent in 2020, 60 percent received malicious emails, and 52 percent encountered smishing (SMS phishing) (TASS 2021), an indication of the attackers' ability to quickly master AI technologies and transition to the digital environment. According to Roskomnadzor in the first quarter of 2023, more than 7.2 thousand phishing resources were removed and blocked in Russia; during the same period of 2022, their number did not exceed 2 thousand. (Isakova 2023). Kaspersky Lab is warning that cybercriminals in 2024 will use more advanced technologies, including artificial intelligence (AI), to launch phishing campaigns (Lenta 2023a).

Pavel Korostelev, head of the product promotion department at the Security Code company, warns that with the help of language models, scammers have already increased the literacy of phishing links — addresses that threaten to lose data or money. Users are more likely to click on a link leading to a page with perfect text and malicious software, the expert explains (Yuriev 2023). Kaspersky Lab experts have revealed details of a new scam scheme in the popular Telegram messenger. The attackers

are luring people into a chatbot that supposedly works based on the ChatGPT 4.0 code. The authors of the bot claim that with its help you can search for merged pictures of a person, having a link to his profile in social networks or a phone number. If you launch a chatbot, a message will appear with an offer to send a link to the profile of the person you are interested in on one of several popular social networks. After that, the service will begin to simulate the process of work — first it will display the message "a search is underway", and then that "the page has been found in the database" and "the material is being sent". The owners of the bot indicate the estimated date of the material drain, as well as how many intimate shots with a person were found. At this stage, a person will see some screenshots, but it is impossible to make out what is depicted on them (the pictures are hidden). To get all the photos and videos, you need to pay 399 rubles for one—time access to the database or 990 rubles for unlimited access. If you transfer money, the deceived user does not receive any materials, and his money goes to the attackers (Kaspersky 2023). According to Kaspersky Lab, attackers are luring money under the guise of currency exchange in a Telegram bot. Now, with the help of chatbots, hackers are already creating encryption viruses and browser plug-ins that can steal passwords and card data (Yuriev 2023).

In the beginning of 2024, Russians were warned about a new type of fraud. Using neural networks, attackers began to fake voice messages on social networks. “Today this is a task that can be solved almost with the snap of a finger. Because there are AI models that require a human voice from 3 to 20 seconds to generate any text, in any emotional tone, regardless of whether it was in your conversation. This voice will be practically indistinguishable from the original even for those close to this person,” says director of the company developing artificial intelligence Roman Dushkin (MIR24TV 2024). “There is another widespread type of fraud when the head of an organization allegedly begins to write to all employees. This is also common at our university; the rector of MEPhI allegedly writes to all employees, down to the smallest assistant of the department, and says that the so-called curator will now contact him. I assume that soon the rector will not just write, but speak in voice messages and the voice of the real rector will be used, because he is a public person, his voice is available in the public space. This will put even more pressure on people who receive such messages” (ibidem). Previously it was reported about phishing using a fake image of a bank card created using a neural network (VTB 2023). Thus, the threat of AI-enhanced spear phishing is growing in Russia.

In September 2021, fraudsters made a deepfake ad using the image of Tinkoff Bank founder Oleg Tinkov. In the video, the fake billionaire encourages people to invest and receive bonuses by clicking on a link below. The fake ad was published on a fake Tinkoff Bonus Facebook page. Its profile picture resembled the logo of the bank. According to Fakecheck, when users clicked on a link below the video they were redirected to a landing page with the bank logo where people were supposed to answer a few questions about investing and fill in a form with their name, email and phone number (Dulneva and Milukova 2021). Obviously, scams like this can easily provoke stress and panic of deceived people, especially in a critical situation. As deepfake technologies continue to improve and more effective schemes of manipulative influence emerge, their psychological impact will only grow.

According to the 2021 research (Statista 2021), more than 10% of Russian citizens regularly use smart voice assistants in their everyday life. For comparison, in the USA that is vying with China for leadership in the field of AI, this indicator reached 30% the same year (Edison Research 2022). Thereby, the authors suggest that there is a real threat of voice assistant’s malicious use in Russia. Hacking voice assistants may result in the same circumstances that are typical for cyberattacks on chat-bots. In addition, hijacking a smart home system or even simply connecting to a smart speaker through this technology would allow attackers to violate people’s privacy and affect their psychological state by intercepting control of devices in their homes.

Just like in other countries where AI-powered robots are being developed, there is a risk of the technology falling into the hands of intruders, making it necessary for Russia to deal with second-level MUAI threats. For example, robot dogs are gaining popularity in Russia. “Intellekt Mashin” (“Intellect of Machines”) company began producing the M-81 model of robot dog at the end of 2021 (TV BRICS 2022) (based on Chinese technologies (IXBT.com 2022)). The question being asked by Russian media, is what will happen if the robot is used by malicious actors? Even the mention of such a possibility reduces public confidence in AI and robotics, but public assessments of such developments are generally positive.

The Third Level of Threats of MUAI Against Psychological Security

In Russia, large banking companies are among the leaders in the development and implementation of AI technologies. In particular, the Russian government has recently signed an agreement on development of AI with one of the country's largest banks called “Sberbank” (The Russian Government 2023). More than that, Russian digital banking is recognized as one of the most dynamic in the world (Wodzicki et al 2020, p. 8). Given these circumstances in Russia the malicious use of banking chat-bots may become a rather dangerous type of MUAI aimed at obtaining users' personal data. The issue of malicious and even terrorist use of bots, created not for communication, has long been discussed in academia and professional circles. For instance, such bots are found to be used to manipulate public opinion and cause reputational damage including during election campaigns (Bazarkina and Pashentsev 2019, p. 155), attract new members to criminal organizations and coordinate their activities (Mihalevich 2022). Meanwhile, intruders use popular in Russia chat-bots for other purposes: logical vulnerabilities allow them to be used to steal bank customer data (Ilyina, 2021). Obviously, chat-bots can simply be hacked in order to obtain information directly from users. It is worth mentioning that this technology is also used in the Russian unified online system for providing public services to citizens called "Gosuslugi". Despite the fact that data leaks are impossible through chat-bot of “Gosuslugi”, at the highest peak of the COVID-19 pandemic it was still subjected to a cyberattack: criminals used it to misinform people about the existence of the Coronavirus and threatened vaccinated citizens with death (Ushkov and Balashova 2021). This example vividly illustrates that chat-bots are a vulnerable technology and its use by intruders can both cause psychological harm to an individual and affect psychological security of an entire country.

Internationally, Russia faces deranking actions by US internet companies in the area of psychological security. In 2017, Google announced its intention to downgrade reports from Russian state-owned publications Russia Today (RT) and Sputnik. The chairperson of Alphabet (the holding company that owns Google), Eric Schmidt, said that the search giant needed to fight the spread of disinformation; meanwhile, US intelligence agencies were calling RT “Russia’s state propaganda machine.” Relevant media publications identified this move as a form of censorship. Speaking at the Halifax International Security Forum in Washington, DC, Schmidt said, “I am strongly not in favor of censorship. I am very strongly in favor of ranking. It’s what we do.” Deranking occurs when Google changes its algorithms to detect information “weapons,” which Schmidt considered publications of the Russian state media (BBC News 2017). These comments prompted a legitimate protest from RT and Sputnik, based on a recording of Google’s statement to the US Congress. Margarita Simonyan, the editor-in-chief of the two Russian publications, said Google’s statement confirmed that it found no manipulation of its platform or other violations by RT (RT 2017). US intelligence agencies accused Russia of trying to influence the 2016 US presidential election in favor of Donald Trump by spreading fake news and hacking the Democratic Party resources to undermine his opponent Hillary Clinton (BBC News 2017). This accusation prompted Twitter to ban RT and Sputnik ads on its platform in October 2017. In November 2017, the US Department of Justice forced RT to register as a “foreign agent.”

The spread of negative imagery about Russia in the Western information environment was a form of action that, in relation to the Russian leadership and the target audience (US citizens), can be assessed as a third-level MUAI threat, since it reinforces a stereotypical perception of Russia's leader. In 2020, two political advertisements using deepfakes of President Putin and North Korean leader Kim Jong-un were posted to social media. The message in both videos was the same: there is no need for Russia or North Korea to interfere in US elections; the United States will destroy its own democracy. The videos were produced and distributed by the human rights group RepresentUs to raise awareness about the need to protect voter rights in the upcoming US presidential election. The videos were released amid harsh public criticism of mail-in voting by then-President Donald Trump, and there was speculation that Trump might refuse to cede the election if he lost. The purpose of the campaign, according to media reports (Hao 2020), was "to shock Americans into understanding the fragility of democracy as well as provoke them to take various actions, including checking their voter registration and volunteering for the polls." RepresentUs collaborated with the creative agency Mischief at No Fixed Address, which came up with the idea "of using dictators to deliver the message." The statement at the end of the video said, "The footage is not real, but the threat is" (Hao 2020). The stereotypical image of the deepfake "dictator" reflects, on the one hand, the realities of the psychological warfare of elites on the world stage; on the other hand, it can also strengthen that image, instilling a sense of anxiety in that country's leader by an economic competitor and political opponent. Significantly, US media networks did not dare take on such a responsibility—the ad was supposed to air on Fox News, CNN and MSNBC but was pulled at the last minute.

The cases of deranking Russian information resources and using a deepfake of the Russian president in the psychological warfare show that third level of MUAI can be used openly, and not only by criminal entities. Currently, the ability to share the Russian state's point of view with an international audience is hampered by the fact that the companies that develop the most important English-language social networks are located in the United States, under the influence of an anti-Russian elites. This challenges Russia to develop alternative social media outlets that can reach an audience beyond national borders. However, if this situation is evaluated not only as a danger but also as an opportunity, the audience expansion for Russian social networks will significantly increase the volume of big data to train domestic AI.

Assessing the growth of threats of the malicious use of AI technologies in Russia is impossible without taking into account external risks in this area. The US Big Tech sector proved itself as a powerful tool for confronting Russia in cyberspace. Brad Smith, president and vice chair of Microsoft, writes in no uncertain terms about the role of his company in Ukraine. "Ukraine's government has successfully sustained its civil and military operations by acting quickly to disburse its digital infrastructure into the public cloud, where it has been hosted in data centers across Europe. This has involved urgent and extraordinary steps from across the tech sector, including by Microsoft. While the tech sector's work has been vital, it's also important to think about the longer-lasting lessons that come from these efforts" (Microsoft, 2022).

General Paul Nakasone, director of the National Security Agency, confirmed in his interview to Sky News as of June, 2022 that the United States had conducted offensive hacking operations in support of Ukraine: "We've conducted a series of operations across the full spectrum; offensive, defensive, [and] information operations" (Martin 2022). Such operations are impossible without the involvement of the Big Tech sphere. Thus, high-tech agenda setting in the United States, that today is unthinkable without the full use of AI technologies, has turned out to be openly subordinated to military and political interests and the needs of psychological warfare.

The use of Western-produced AI technologies in the ongoing military conflict in Ukraine is significant. US facial recognition start-up "Clearview AI" has provided technical support to Ukraine.

Clearview AI's tools can identify faces in videos, comparing them to a company's database of 20 billion images from public networks and identifying potential spies and killed people. AI tools also play an important role in Ukraine's propaganda war and in processing critical information about the conflict. A program from the US company "Primer" can perform speech recognition, transcription and translation. It intercepts and analyzes Russian data, including conversations between Russian soldiers in Ukraine. A Swiss encrypted chat service called "Threema" allows Ukrainian users to send this data to the military without revealing their identities (Global Times, 2022). 5 June 2023, in Russian regions, some radio stations and television broadcasted a fake appeal on behalf of President Vladimir Putin about the introduction of martial law in three regions, as well as about the announcement of mobilization. These claims turned out to be false and the video was a deepfake (Lenta 2023b), which, of course, opens a new stage of psychological confrontation within the framework of the Ukrainian conflict.

Conclusion

Russia meets with internal and external threats of MUAI in the sphere of psychological security. Moreover, the latter are clearly increasing with the growth of international tensions, active hybrid warfare against Russia waged by the United States and its allies. Obviously, with the increasing development of AI in various states the probability of using practically any type of AI for unlawful purposes is becoming higher. Thus, it seems advisable to establish regional and international cooperation in order to jointly develop measures to counteract MUAI which is using personal data that threatens the security of all countries. Besides, interstate cooperation is also necessary to determine the interrelation between personal data and AI and to establish interdisciplinary standards. What is more, it is crucial not only to determine the cases in which the use of personal data for AI would be considered a breach, but also to work out measures for their protection, up to the restriction of the use and further development of AI under certain conditions.

References

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

BBC News (2017) Google to 'derank' Russia Today and Sputnik. In: BBC News. https://www.bbc.com/news/technology-42065644. Accessed 29 Mar 2024

Dulneva M, Milukova Y (2021) "Hugged everyone!": the image of Oleg Tinkov was used in deepfake advertising. In: Forbes. https://www.forbes.ru/milliardery/439255-vseh-obnal-obraz-olega-tin-kova-ispol-zovali-v-dipfejk-reklame. Accessed 19 Jan 2024

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

Gavrilyuk A, Korolyov N (2022) Grazhdan zashchityat ot robotov (Citizens will be protected from robots). In: Kommersant. https://www.kommersant.ru/doc/5173457. Accessed 29 Mar 2024

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

Hao K (2020) Deepfake Putin is here to warn Americans about their self-inflicted doom. In: MIT Technology Review. https://www.technologyreview.com/2020/09/29/1009098/ai-deepfake-putin-kim-jong-un-us-election/. Accessed 29 Mar 2024

Ilyina N (2021) Cheating by correspondence: vulnerabilities in bank chatbots allow money theft. In: Izvestiya. https://iz.ru/1214668/natalia-ilina/obman-po-perepiske-uiazvimosti-v-bankovskikh-chat-botakh-pozvoliaiut-krast-dengi. Accessed 19 Jan 2024

Interfax (2020) The State Duma approved a special legal regime for the development of artificial intelligence for Moscow. In: Interfax. https://www.interfax.ru/russia/704092. Accessed 29 Mar 2024

Isakova T (2023) Khakery nashchupali tochku rosta (Hackers have found a growth point). In: Kommersant. Accessed 20 Jan 2024

IXBT.com (2022) V Rossii predstavili robosobaku s granatomyotom (In Russia, a robot dog with a grenade launcher was introduced). In: IXBT.com. https://www.ixbt.com/news/2022/08/15/v-rossii-predstavili-robosobaku-s-granatometom.html. Accessed 29 Mar 2024

Kaspersky (2023) Obmenu i vozvratu ne podlezhit: zloumyshlenniki vymanivayut den'gi pod vidom obmena valyuty v Telegram-bote (Cannot be exchanged or returned: attackers lure money under the guise of exchanging currency in a Telegram bot) https://www.kaspersky.ru/about/press-releases/2023_obmenu-i-vozvratu-ne-podlezhit-zloumyshlenniki-vymanivayut-dengi-pod-vidom-obmena-valyuty-v-telegram-bote?ysclid=lserd46yap191790756. Accessed 10 Feb 2024

Lenta (2023a) Rossiyan predupredili o samykh opasnykh khakerskikh atakakh v 2024 godu (Russians warned about the most dangerous hacker attacks in 2024). https://lenta.ru/news/2023/11/14/rossiyan-predupredili-o-samyh-opasnyh-hakerskih-atakah-v-2024-godu/?ysclid=lrm9958719396125920. Accessed 20 Jan 2024

Lenta (2023b) «Obrashcheniye» Putina o mobilizatsii i voyennom polozhenii okazalos' dipfeykom (Putin’s “appeal” about mobilization and martial law turned out to be a deepfake). https://lenta.ru/news/2023/06/05/fake_radio/?ysclid=lrm69q0hln874723542. Accessed 20 Jan 2024

Mamikonyan O (2022) 58% rossiyan boyatsya sokrashcheniya rabochikh mest iz-za razvitiya iskusstvennogo intellekta (58% of Russians are afraid of job cuts due to the development of artificial intelligence). In: Forbes. Aceessed 20 Jan 2024

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

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

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

MIR24TV (2024) Kak iskusstvennyy intellekt ispol'zuyut moshenniki v Rossii? (How is artificial intelligence used by scammers in Russia?). https://mir24.tv/articles/16577222/kak-iskusstvennyi-intellekt-ispolzuyut-moshenniki-v-rossii?ysclid=lrkoid1382919794190. Accessed 20 Jan 2024.

President of the Russian Federation (2019a) Decree of 10.10.2019 No. 490 “On the development of artificial intelligence in the Russian Federation.” In: Official Legal Information Portal. http://publication.pravo.gov.ru/Document/View/0001201910110003. Accessed 29 Mar 2024

President of the Russian Federation (2019b) Federal Law No. 123-FZ of 24.04.2020 “On conducting an experiment to establish special regulation in order to create the necessary conditions for the development and implementation of artificial intelligence technologies in the subject of the Russian Federation—the Federal City of Moscow and amending articles 6 and 10 of the Federal law ‘On personal data.’” In: Official Legal Information Portal. http://publication.pravo.gov.ru/Document/View/0001202004240030?index=0. Accessed 29 Mar 2024

President of the Russian Federation (2020) Decree of 10.10.2019 No. 490 “On the development of artificial intelligence in the Russian Federation.” In: Official Legal Information Portal. http://publication.pravo.gov.ru/Document/View/0001201910110003. Accessed 29 Mar 2024

RIA Katyusha (2020) Hello, Chinese-style electronic concentration camp: Sobyanin wants to put Muscovites under the control of artificial intelligence. In: RIA Katyusha. http://katyusha.org/view?id=14044. Accessed 29 Mar 2024

RT (2017) Google will “de-rank” RT articles to make them harder to find—Eric Schmidt. In: RT International. https://www.rt.com/news/410444-google-alphabet-derank-rt/. Accessed 29 Mar 2024

Sputnik Africa (2023) What Do Most Russians Think About Impact of AI on People's Lives? https://en.sputniknews.africa/20231122/what-do-most-russians-think-about-impact-of-ai-on-peoples-lives-1063715493.html?ysclid=lu2vicgqf4500218098 Accessed 19 Jan 2024

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

Stepanova Y (2020) Parents have fallen like children. Online fraudsters took advantage of the demand for benefits. In: Kommersant. https://www.kommersant.ru/doc/4343398. Accessed 29 Mar 2024

TASS (2021) Eksperty: s fishingovymi atakami v 2021 godu stolknulis' 45% rossiyan (Experts: 45% of Russians faced phishing attacks in 2021). In: TASS. https://tass.ru/ekonomika/13105631. Accessed 29 Mar 2024

The Russian Government (2023) The Government signed the final package of cooperation agreements on the development of high-tech areas. http://government.ru/news/47551/. Accessed 19 Jan 2024

TV BRICS (2022) V Rossii razvivaetsya rynok robosobachestva (Robotic dog market develops in Russia). In: TV BRICS. https://tvbrics.com/news/v-rossii-razvivaetsya-rynok-robosobachestva/. Accessed 29 Mar 2024

Uglova Y (2023) Rossiyane teryayut den'gi iz-za «umnogo» chat-bota: podrobnosti (Russians are losing money because of a “smart” chatbot: details). In: Hi-Tech. https://hi-tech.mail.ru/news/100835-rossiyane-teryayut-dengi-iz-za-umnogo-chat-bota-podrobnosti/. Accessed 10 Feb 2024

Uskov M, Balashova A (2021) The authorities announced an attack on "Gosuslugi" after reports of an anti-vaxxer bot. In: RBC. https://www.rbc.ru/technology_and_media/11/11/2021/618d42109a7947252fe7d448. Accessed 19 Jan 2024

Uvarchev L (2024) Putin poruchil izuchit' primeneniye iskusstvennogo intellekta v rassledovaniyakh (Putin ordered to study the use of artificial intelligence in investigations). In: Kommersant. https://www.kommersant.ru/doc/6454973?ysclid=lrkljlwu1i347691345. Accessed 20 Jan 2024.

VTB (2023) VTB: moshenniki vymanivayut sredstva kliyentov v Telegrame s pomoshch'yu kart s poddel'nym dizaynom (VTB: Fraudsters are swindling clients' funds on Telegram using cards with fake designs). https://www.vtb.ru/about/press/news/?id=198614. Accessed 20 Jan 2024

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

Yurasova Y, Tishina Y, Petrova V (2023) Gore ot intellekta (Woe from intellect). In: Kommersant. https://www.kommersant.ru/doc/5928661. Accessed 20 Jan 2024.

Yuriev D (2023) Ne myt'yem, tak plaginom: kak moshenniki ispol'zuyut chat-boty v Rossii (Not by chance, but by plugin: how scammers use chatbots in Russia). In: Ferra. https://www.ferra.ru/news/v-rossii/ne-mytem-tak-plaginom-kak-moshenniki-ispolzuyut-chat-boty-v-rossii-10-05-2023.htm?ysclid=lser17n5lg203176776. Accessed 10 Feb 2024

The Malicious Use of AI: Challenges to Psychological Security in the United Arab Emirates

Evgeny Pashentsev, Vladilena Chebykina, Ruslan Nikiforov

Introduction

Over the past two decades, the United Arab Emirates has been actively developing its technological capabilities, including the field of artificial intelligence. The rapid growth in the number of startups and the high investment rate in this area indicate the significant commitment of the citizens and government in implementing AI in key sectors of the state's life. According to the report written by Economist Impact and supported by Google "Pushing forward: the future of artificial intelligence in the Middle East and North Africa", the annual growth in economic contribution of AI is expected to reach 20-34% per year across the countries in the Middle East and North Africa (MENA) region, with the highest rates expected in the UAE and Saudi Arabia. The potential economic impact of AI in the region, however, is likely to rise even further, with a more recent country-based study by The Economist Intelligence Unit (EIU) forecasting that Saudi Arabia and the UAE alone will be accruing US$200bn and US$120bn, respectively (The Economist Group 2022).

The UAE is the first country in the Middle East to launch its Ministry of Artificial Intelligence in 2017 and adopt the National AI Strategy 2031. The strategy outlines a plan to equip employees with all the necessary skills to navigate the constantly evolving technological landscape. In 2019, the world's first Artificial Intelligence University (MBZUAI[6]) was launched in the UAE capital city with the aim of developing the required AI ecosystem to harness its potential at all levels (Zaatari S 2019). A massive influx of tech workers since 2021 is helping fuel the Gulf state’s AI ambitions. As of September 2023, there were 120,000 people working on AI or AI-related industries — up from 30,000 two years prior, Al Olama the UAE’s Minister of State for Artificial Intelligence Omar Al Olama told Bloomberg in an interview at the World Government Summit in Dubai in February 2024 (Abu Omar 2024). Over 50 per cent of working people in the UAE apply AI in the course of their work in scope of activity such as media, education, healthcare, banking, etc. The country's government is actively embracing AI technologies across all sectors through strategic public-private partnerships.

However, alongside the positive aspects of AI development, the practice of its malicious use poses growing challenges. Such practice is gaining momentum in a country with a high level of socio-political stability, but located in a high-conflict region, with the potential for local conflicts to escalate into a great Middle East war or even World War III.

The First Level of Threats of MUAI Against Psychological Security

With the development and proliferation of AI technologies and the integration of automated systems into ordinary social life, the main first-level threat is targeted speculation on the growing concerns about the loss of millions of jobs. According to PwC's ‘2023 Workforce Hopes and Fears Survey’ report, over a third of respondents in the UAE note the positive impact of artificial intelligence on labor efficiency (PwC Press Release 2023). However, 52% of survey participants reasonably believe that they will need to undertake advanced training courses due to the nature of their work changing significantly in the next five years or becoming irrelevant altogether (Abbas W 2023a). In the National Strategy of the UAE, one of the key provisions is the goal to provide the population with opportunities

[6] Mohammad Bin Zayed University of Artificial Intelligence

for retraining, such as through specialized training or international internship programs. The UAE government funds STEM education programs to create a talent pool in the field of innovation, including free courses for those willing to improve their AI competencies (Alrahmah B, Ahmed M A 2024). Nevertheless, this is often insufficient to prevent people's growing concerns about remaining competitive in the labor market.

The study, conducted by communication advisers duke+mir, in collaboration with YouGov, asked people how they believed AI would affect their lives. 55% said they were worried that their roles would be replaced by AI or robots by 2033. Approximately 24% of people were unsure, and 21% were not concerned about their roles being replaced by AI technologies. It is noteworthy that 66% of people under 25 were fearful that AI and robots would take their jobs in the next decade, compared to 57% of those in the age group of 25 to 44, while and 43 per cent of those aged 45 and older were concerned (Webster N 2023a). Jonathan Ivan-Duke, co-founder and partner at duke+mir, commented on the survey results: “With such a strong focus from the UAE government on providing and protecting Emirati jobs both now and in the future, it’s quite unexpected to see the youth and Emiratis of the UAE being the most concerned about future technological advancements,” (Webster N 2023a).

Apparently, the main reason for concern about potential job loss due to the introduction of AI technologies is their pervasive nature, which affects (or will affect as AI advances) virtually all occupations. And the better informed people are about the possibilities of AI, the higher they assess the risks of job loss. Experts are making a significant contribution to raising awareness of these risks. According to Shalini Verma, CEO, Pivot Technologies, “AI in its generative version will completely transform certain jobs, taking over almost all entry-level jobs. Anyone entering the job market will start at a much higher level than is expected of interns and junior executives today. This human-machine swap won’t just stop here. If you are not exceptional, you are at grave risk of being replaced by a bot. If you are an expert, then you are likely to be doing your work very differently, as core tasks will be taken over by AI in every job role” (Abbas W 2023b). The very term "exceptional" makes most non-“exceptional” specialists increasingly anxious about the future.

The use of AI at the national level in the UAE and, above all, the possible negative consequences of such use, may become a means of manipulating the consciousness of the broad segments of the population, especially the youth. This could happen in the UAE even earlier than in many other countries precisely due to the socially necessary, but with its own risks, rapid development and rollout of AI technologies. The UAE leadership should be commended for recognizing the complexity of this situation. Omar Al Olama, Minister of Artificial Intelligence in the UAE, speaking at the Dubai Assembly on Generative AI in October 2023, urged citizens not to fear job loss, but to focus on expanding the positive aspects of AI technologies (Awienat D 2023). The use of AI in the UAE could benefit all citizens, including employees and companies, provided a responsible and, in some respects, proactive approach to the training and skills system in the application of AI is developed. Subsequently, this could open up new opportunities for human enhancement and the development of increasingly sophisticated forms of hybrid intelligence.

The Second Level of Threats of MUAI Against Psychological Security

The second-level of threats represents risks aimed at critical infrastructure facilities, physical safety of a person and damage to his property and well-being, using AI technologies. According to the International Telecommunication Union’s Global Cybersecurity Index 2020, the UAE ranked 5th, however, the country still suffers greatly from cyberattacks: many companies have paid more than $ 1.4 million in ransom. 42% of them were forced to close after the incident, and 90% were subjected to repeated attacks. The UAE government, in turn, invests heavily in cybersecurity, but doesn’t always

manage to implement all measures evenly, which makes local companies less resistant to large-scale cyber incidents (Filatov A 2021).

A record high of more than 26,000 vulnerabilities were reported in 2022 identified as per the NIST National Vulnerability Database (NVD). Two-thirds (66%) of UAE respondents reported one or more breaches to their organization from cyberattacks according to Infoblox's 2023 Global State of Cybersecurity Report. Phishing was the most common attack method against organizations that were breached, accounting for 62% of attack methods in the past year, followed by advanced threats (APTs) (53%) and ransomware (51%) (Bandyopadhyay S 2023). On average, UAE organizations detected more issues resulting from email/phishing attacks than any other type.

According to a report by Proofpoint Inc, an information security company, in 2022, approximately 2/3 of enterprises (64%) in the UAE were attacked using so-called "ransomware schemes" (Ryan P 2023). So, in 2023, an employee of a company in Dubai received a message on WhatsApp (owned by Meta (recognized as an extremist organization in the Russian Federation, its activities are prohibited) according to an already well-known scheme — the attacker introduced himself as its director, uploaded his photo to the profile and thereby lured almost $ 4,000 from the employee "to purchase certificates for their client" (Nair D 2023). The employee wasn't confused by someone else's number, she trusted the photos and the attacker's stories that his phone simply ran out of power.

One of the main threats at the second level is a cyberattack on critical infrastructure facilities. These can be power plants, water supply systems, or transportation systems. In 2018, two serious data leaks occurred in the country, as a result of which "14 million records were compromised" (Chandra G R, Sharma B K, Ali I 2019). Two data leaks were recorded, in particular, in the Dubai-based car travel platform Careem. The use of AI allows attackers to automate and optimize their actions, which makes such attacks more effective and dangerous. For example, attackers can use AI to find and exploit vulnerabilities in power plant control systems, which can lead to power outages in large cities.

Human physical security also refers to threats at the second level. For example, autonomous drones equipped with weapons and programmed to attack can pose a threat to mass events or highly visited places. Given that the use of AI-equipped drones is growing, and they are being tested in combat in hot spots of the planet, including the Middle East, risks of their use by malicious actors in the UAE should be expected. According to A. Al Khoori, senior vice president of strategy and excellence at UAE defense conglomerate Edge Group "… at the end of the day, the user will have a system that can operate autonomously, that can decide for the end user» under continually changing landscape of autonomous defense technology amid the onslaught of generative AI developments (Combs C 2024).

In addition to physical security, AI can be used to cause damage to human property and well-being. For example, attackers can use AI to create fraudulent financial schemes, hack online banking systems or steal personal data. New malware contributed to an increase in the number of banking attacks using malicious Trojans in the first quarter of 2023 compared to the first quarter of 2022. In general, in the first quarter of 2023, the Middle East region also saw an increase in the number of Trojan attacks on banking services. The UAE saw an increase of 67% (El-Din M A 2023). Such actions can lead to significant financial losses and violation of a person's privacy.

According to Arif Aljanahi, security engineering department director of the Security Industry Regulatory Agency, which was set up by the Dubai government, "A key point that people do not consider about AI is that it depends on the information that is being fed to the algorithm…The success of AI depends on who is teaching the system. AI can be used in a good way, and also in a bad way" (Webster N 2023a). In the context of acute geopolitical confrontation, the growing influence of organized and increasingly high-tech crime on a global scale, this statement appears particularly significant.

The above-mentioned and other risks to critical infrastructure, life and property of people imply certain psychological consequences, both immediate and delayed, which can have a controlled and uncontrolled negative impact on people's consciousness. For example, an AI-based drone attack can be accompanied as a spontaneous reaction (feelings of fear, hatred, panic states, etc.), which at the third level of threats can be complemented by the formation of a disinformation agenda using AI technologies and appropriate information campaigns.

The Third Level of Threats of MUAI Against Psychological Security

Despite the existence of national laws prohibiting cyberbullying, the creation of deepfakes and the publication of content containing fake news, the UAE, like many other countries, is facing an outbreak of fake news that can disrupt the political system of the state, as well as damage private life of individuals and influence public consciousness.

Sumsub Identity Fraud Report 2023 confirms the new danger and points out that the global volume of deepfakes has grown exponentially (more than 10 times) between 2022 and 2023. In the Middle East and Africa, the increase was 450%. It is also noted that the UAE passport has become the most forged document in the world, so the proliferation of deepfakes is a pressing issue for the country that requires immediate action (Wassi E 2023).

In 2020, the attackers called the manager of a branch office of a Japanese company and pretended to be the director heading the firm in the UAE. The manager, not suspecting anything wrong as he had heard this voice repeatedly, made the requested $35 million money transfer (Brewster T 2021). The fraudsters used "deep voice" technology to mimic the director's speech. The investigation revealed that at least 17 people were involved in this crime – the stolen money was transferred to accounts in different parts of the world. Experts believe that manipulating "deep voice" is much easier in the creation process than fake videos, and therefore the number of such crimes will increase every day, jeopardizing companies and ordinary people. However, some companies, such as Pindrop, recognize the potential for malicious use of AI, and are developing software that can detect synthesized voices and further prevent the spread of deepfakes.

In 2023, another high-profile case in the UAE was the case of a "mate" making phone calls to his mate in Kerala, India. The man used fabricated audio and video calls to obtain thousands of dirhams, citing health problems. The 73-year-old victim did not suspect the fakes as the attacker used details from his mate's personal life, thereby trying to gain trust from the victim (Sankar A 2023). This case also illustrates that abusers use and will continue to use a variety of methods to psychologically influence their victim, whether it is a company or an individual.

Another significant issue is the use of generative AI to create obscene materials and further manipulate children. According to research by WeProtect, the number of grooming messages for monetary gain increased from 139 in 2021 to over 10,000 in 2022 (Webster N 2023b). Experts attribute the rise in offences of this nature to the proliferation of social networks and the advancement of AI technologies. The most frequent victims of virtual sexual abuse are adolescent boys; with perpetrators posing as young girls, sending fake explicit photos and videos of an obscene nature, receiving real materials in return, and demanding money to keep silent in front of their parents.

The development of AI has also led to the creation and spread of religious chatbots. The UAE is among the 10 countries where users turn to the QuranGPT chatbot (Prabhakar A 2023). There are two main issues in this matter: AI bias and the risk of religious chatbots falling into the hands of malicious users. The creators aim to represent their religion faithfully to the world; however, the virtual world still remains Islamophobic, with GPT-3 being accused of 'anti-Muslim bias' in 2021 (Samuel S 2021). Therefore, the AI may misinterpret user queries, leading to misinformation at least and ethnic

problems in society at most. There is a danger that religious chatbots could be exploited by malicious individuals to spread deliberately false information or propaganda, as well as to incite hatred. In order to avoid these problems, developers need to be particularly careful in selecting sources of information, accurately representing religious teachings and ensuring system security.

The use of chatbots in public administration has also been initiated in the UAE: "Rashid" chatbot, an AI-powered government chatbot assistant, built to answer questions regarding the necessary government procedures, documents and requirements for conducting various transactions (The Economist Group 2022). At the same time, the potential for malicious actors to intercept control of chatbots should be taken into account, which, at some point, could pose a serious threat to psychological security.

The proliferation of deepfakes creates a problem of decreasing trust in information sources within society. The UAE government aims to help the public learn how to recognize deepfakes by publishing a guide designed to raise awareness of the harmful and beneficial uses of deepfake technology. It is possible to identify deepfake independently, but experts believe that as technology improves, it will become more challenging to do. Eventually, the only way to determine the authenticity of information will be the AI itself. The guidance states that: "The most accurate approach to detect forged contents is through a systematic screening of the deepfakes using AI-based tools that need to be regularly updated" (The National, 2021). Such tools can analyze text, images, audio and video files for signs of manipulation or distortion. This approach will help improve the accuracy of detecting fakes and protect users from their negative impact.

Conclusion

There are various levels of threats related to AI technologies in the UAE. At the first level, there are risks of intentionally misinterpreting the development of AI in the interests of antisocial groups, which can lead to socio-economic conflicts. Such interpretations haven't become a significant factor in public life in the UAE yet, but we should expect attempts by malicious actors to use negative consequences, and even significant achievements in the development of the AI industry, for their own purposes, especially in the extremely difficult and dangerous situation in the Middle East and in the global dimension. At the second level, threats are associated with the possibility of cyberattacks on critical infrastructure and damage to people and their property with concomitant psychological effects. The number of such attacks is growing rapidly, as is the role of AI technologies in their implementation. At the third level, there is a danger of destructive psychological effects, including the use of deepfakes and chatbots to manipulate public consciousness. Threats of various levels require attention and appropriate measures to ensure psychological security in the UAE.

References

Abbas W (2023 a) UAE: Will automation replace jobs or help employees gain new skills? In: Khaleej Times. https://www.khaleejtimes.com/jobs/uae-is-ai-a-threat-to-jobs-how-employees-can-use-tech-to-boost-hiring-chances-get-shortlisted-by. Accessed 24 Jan 2024

Abbas W (2023 b) Jobs in UAE: Entry level roles set to be wiped out by AI, automation and ChatGPT. In: Khaleej Times. https://www.khaleejtimes.com/jobs/jobs-in-uae-entry-level-roles-set-to-be-wiped-out-by-ai-automation-and-chatgpt. Accessed 24 Jan 2024

Abu Omar A (2024) UAE Backs Sam Altman Idea to Turn Itself into AI Testing Ground. In: Bloomberg. https://www.bloomberg.com/news/articles/2024-02-15/minister-backs-altman-s-idea-to-turn-uae-into-ai-testing-ground. Accessed 15 Mar 2024

Alrahmah B, Ahmed M A (2024) The UAE's harnessing of AI at the national level can benefit everyone. In: The National News. https://www.thenationalnews.com/opinion/comment/2024/01/16/the-uaes-harnessing-of-ai-at-the-national-level-can-benefit-everyone/. Accessed 24 Jan 2024

Awienat D (2023) Generative AI should not be feared despite risks, says UAE minister of artificial intelligence. In: Arab news. https://www.arabnews.com/node/2390716/middle-east. Accessed 24 Jan 2024

Bandyopadhyay S (2023) UAE Cybersecurity: 26,000 vulnerabilities were reported in 2022. In: Khaleejtimes. https://uaetimes.ae/uae-cybersecurity-26000-vulnerabilities-reported-in-2022-news/. Accessed 2 Feb 2024

Brewster T (2021) Fraudsters Cloned Company Director's Voice In $35 Million Heist, Police Find. In: Forbes. https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=442e5dd67559. Accessed 27 Jan 2024

Chandra G R, Sharma B K, Ali I (2019) UAE's Strategy Towards Most Cyber Resilient Nation. In: International Journal of Innovative Technology and Exploring Engineering. https://www.researchgate.net/publication/337146109_UAE%27s_Strategy_Towards_Most_Cyber_Resilient_Nation. Accessed 2 Feb 2024

Combs C (2024) AI has changed defence industry expectations, says Edge executive. In: The National News. https://www.thenationalnews.com/business/future/2024/01/25/uae-ai-edge-umex/. Accessed 2 Feb 2024

Devi A (2021) UAE ranks 5th in UN's 2020 Global Cybersecurity Index. In: Security Middle East&Africa. https://securitymea.com/2021/07/01/uae-ranks-5th-in-uns-2020-global-cybersecurity-index/. Accessed 2 Feb 2024

El-Din M A (2023) 49% increase in phishing attacks in Egypt during 1Q 2023: Kaspersky. In: Daily News Egypt. https://www.dailynewsegypt.com/2023/05/07/49-increase-in-phishing-attacks-in-egypt-during-1q-2023-kaspersky/. Accessed 2 Feb 2024

Filatov A (2021) Russia shared the fifth place in the GCI cybersecurity rating with Malaysia and the UAE. In: Digital Russia. https://d-russia.ru/rossija-razdelila-s-malajziej-i-oaje-pjatoe-mesto-v-rejtinge-kiberbezopasnosti-msje.html. Accessed 2 Feb 2024

Nair D (2023) The Dept Panel: 'I lost $4,000 in a WhatsApp scam'. In: The National News UAE. https://www.thenationalnews.com/business/money/2023/08/17/the-debt-panel-i-lost-4000-in-a-whatsapp-scam/. Accessed 27 Jan 2024

Prabhakar A (2023) Religious GPT: The chatbots and developers fighting bias with AI. The National News UAE. https://www.thenationalnews.com/weekend/2023/07/28/religious-gpt-the-chatbots-and-developers-fighting-bias-with-ai/. Accessed 27 Jan 2024

PwC Press Release (2023) Workforce in the Middle East is ambitious, enthusiastic about change, embracing AI and upskilling, says new PwC report. https://www.pwc.com/m1/en/media-centre/2023/workforce-in-the-middle-east-is-ambitious-enthusiastic-about-change-embracing-ai-upskilling.html. Accessed 24 Jan 2024

Ryan P (2023) How the future of cyber crime could involve fake voice messages from loved ones. In: The National News UAE. https://www.thenationalnews.com/uae/2023/03/17/how-the-future-of-cybercrime-could-involve-fake-voice-messages-from-loved-ones/. Accessed 27 Jan 2024

Samuel S (2021) AI's Islamophobia problem. In: Vox. https://www.vox.com/future-perfect/22672414/ai-artificial-intelligence-gpt-3-bias-muslim. Accessed 27 Jan 2024

Sankar A (2023) Deepfake video call said to be from Dubai used to swindle Kerala man out of thousands. In: The National News UAE. https://www.thenationalnews.com/uae/2023/07/19/deepfake-video-call-pretending-to-be-dubai-friend-used-to-swindle-man-out-of-thousands/. Accessed 27 Jan 2024

The Economist Group (2022) Pushing forward: the future of AI in the Middle East and North Africa. In: Economist Impact. P. 5, 55. https://impact.economist.com/perspectives/sites/default/files/google_ai_mena_report.pdf. Accessed 27 Jan 2024

The National (2021) UAE asks public to help tackle deepfakes. In: The National News UAE. https://www.thenationalnews.com/uae/2021/07/09/uae-asks-public-to-help-tackle-deepfakes/. Accessed 27 Jan 2024

Wassi E (2023) Fraude de identidade y deepfakes: suenan las alarmas. In: La Prensa. https://www.laprensa.com.ar/Fraude-de-identidad-y-deepfakes-suenan-las-alarmas-538397.note.aspx. Accessed 26 Jan 2024

Webster N (2023 a) Rise of AI creates job worries, UAE survey finds. In: The National News UAE. https://www.thenationalnews.com/uae/2023/01/19/rise-of-ai-creates-job-worries-uae-survey-finds/. Accessed 24 Jan 2024

Webster N (2023 b) Online gaming poses alarming threat to children's safety, report finds. In: The National News UAE. https://www.thenationalnews.com/uae/2023/10/19/online-gaming-poses-alarming-threat-to-childrens-safety-report-finds/. Accessed 24 Jan 2024

Zaatari S (2019) University of Artificial Intelligence launched in Abu Dhabi. In: Gulf News. https://gulfnews.com/uae/university-of-artificial-intelligence-launched-in-abu-dhabi-1.67170778. Accessed 24 Jan 2024

Conclusion: Future Risks of the Malicious Use of AI and Challenges to Psychological Security

Evgeny PASHENTSEV

The future is multivariate, therefore, now it is only possible to talk about the approximate parameters of the future risks of the MUAI to psychological security, taking into account existing global trends and forecasts, which are sufficiently contradictory. In the near future, we should expect an increase in such risks due to the rapid development of AI technologies, their relative cost effectiveness and accessibility to an increasingly wide range of users, the growth of crisis phenomena in the modern world, the high level of geopolitical rivalry that is turning into dangerous confrontations, the direct influence of antisocial forces on information flows in individual countries and at the global level—all of these and other factors, apparently, they will make the threats of MUAI to psychological security more widespread and dangerous all over the world, including among the BRICS countries.

New threats to psychological security are emerging from the advantages of both offensive and defensive psychological operations using AI. These advantages—as well as threats—are increasingly associated with quantitative and qualitative differences between traditional mechanisms of producing, delivering and managing information, new possibilities for creating psychological impacts on people, and the waging of psychological warfare. In particular, these advantages may include:

(1) the amount of information that can be generated to destabilize an adversary;

(2) the speed of generation and dissemination of information;

(3) new opportunities for obtaining and processing data;

(4) the application of predictive analytics using AI;

(5) new decision-making process opportunities from big data analysis with the help of AI;

(6) new ways to educate people with intelligent systems;

(7) the perceived credibility of generated (dis-)information;

(8) the strength of the intellectual and emotional impact of generated information on target audiences; and

(9) a qualitatively higher level of thinking in the future through the creation of general and strong AI, as well as through the further development of human cyborgization, development of advanced forms of hybrid intelligence.

Based on the research project titled “Malicious Use of Artificial Intelligence and Challenges to Psychological Security in Northeast Asia” jointly funded by the Russian Foundation for Basic Research (RFBR) and the Vietnam Academy of Social Sciences (VASS) “realized in 2021-2023, it can be concluded that advantages 1–6 have already been achieved and continue to grow in a number of important aspects—though not in all—qualitatively exceeding human capabilities without AI. At the same time, all the possibilities of narrow (weak) AI are still generally under human control. Advantages 7–8 have not yet been practically implemented; this does not exclude recent achievements in the formation of these advantages, such as credibility and emotional persuasiveness (see, for example, Unity 2022), but they can be achieved through the quantitative and qualitative improvement of existing technologies in the foreseeable future. The future benefit of number 9 may be require fundamental scientific

breakthroughs and new technological solutions. This list of benefits from using AI in psychological warfare is not exhaustive and is highly variable. (Pashentsev, 2022, p. 7).

MUAI and Three Levels of Threats to Psychological Security: Prospects for the Future

At all three levels, MUAI threats to psychological security will increase due to the growing danger, variety of methods, bigger audiences, and frequency of malicious impact on people.

The first level. In the coming years, it is possible to strengthen negative attitudes towards AI, up to the formation of stable panic states, phobias, and active rejection of technologies, which can be supported by both mistakes in their implementation and actions of malicious actors. It is impossible to exclude the emergence of ultra-radical movements, both for and against AI. For example, some new, still emerging religious beliefs with faith in artificial superintelligence may eventually, in the context of an increasing global crisis, give rise to sectarian offshoots and give fanatical and militant protagonists the speedy arrival of this superintelligence in the name of saving/eliminating humanity. The emergence of religious faith in AI is already quite acceptable, justified and welcomed in some publications (McArthur 2023).

On the other hand, any socially significant and large-scale negative consequences of the development and introduction of AI technologies can provoke the emergence of "new Luddite" movements, which can also be exploited by malicious actors. A particularly significant threat may be decisions to introduce *more advanced and cheaper AI technologies* (the imminent appearance of which is almost inevitable) not as a *mass human assistant*, but as a *mass replacement tool* of the workforce without creating alternative jobs and appropriate retraining programs.

Many centuries ago, long before the very prerequisites for the emergence of AI technologies, the ancient Greek philosopher Aristotle made this famous quote: "…if every tool could perform its own work when ordered, or by seeing what to do in advance…if thus shuttles wove and quills played harps of themselves, master-craftsmen would have no need of assistants and masters no need of slaves" (Aristotle, Politics 1.1253b). Seeing the prospect of large-scale implementation of AI and smart robots, Big Tech actively supported the theory of the Universal Basic Income (UBI). UBI is an economic theory that stipulates every citizen should have a government-provided income regardless of need.

Tech billionaires, like Sam Altman, say they're big fans of UBI. Musk, the CEO of Tesla and SpaceX, told CNBC that "there's a pretty good chance we end up with a universal basic income, or something like that, due to automation" (Weller 2017). Facebook co-founder Chris Hughes is an active supporter of UBI, and he urges people to consider what systems we'll need to create if millions more follow (Weller 2017). It is very unlikely that in the short term the threat of unemployment due to the introduction of AI will become a reality for the obvious majority of the population, but in the medium term it can become a factor of social and political destabilization.

Proposals on the need to adopt UBI, including due to the implementation of AI, are unlikely to solve the problem. Of course, it will be good and just if a person is freed from monotonous types of work that do not develop intelligence and the emotional sphere, as well as from activities harmful to health, due to AI technologies and robots. But *if the majority of the population does not work all their lives and find happiness in idleness, such a society will deteriorate dangerously* (the signs of that are present in the West, where in many countries there is a high level of long-term youth unemployment in the absence of mass poverty, which is characteristic of poor and technologically backward countries). Let us also recall the fate of Ancient Rome, where the emperors, giving citizens bread and circuses at the expense of the labor of numerous slaves, eventually lost both citizens, slaves, and power.

There are already studies confirming the negative impact of AI technologies on personality.

So the study published in 2023 by a big team of researchers examines the impact of AI on loss in decision-making, laziness, and privacy concerns among university students in Pakistan and China. This study is based on qualitative methodology using PLS-Smart for the data analysis. Primary data was collected from 285 students from different universities in Pakistan and China. “The findings show that 68.9% of laziness in humans, 68.6% in personal privacy and security issues, and 27.7% in the loss of decision-making are due to the impact of AI in Pakistani and Chinese society. From this, it was observed that human laziness is the most affected area due to AI. However, this study argues that significant preventive measures are necessary before implementing AI technology in education. Accepting AI without addressing

the major human concerns would be like summoning the devils” (Ahmad et al. 2023) These dangerous trends can be countered from childhood by educating not a consumer of "fabulous" technology, but a responsible user who receives not so much ready-made benefits with its help, but rather develops its cognitive skills and social responsibility.

Apparently, it is no coincidence that the tasks of the large–scale program initiated by the Ministry of Education of the People's Republic of China in 2024 include studying models, innovative concepts, gaining experience in implementing AI in the learning process, and retraining teachers (Big Asia 2024).

The second level. At the second level of threats, the situation will become seriously complicated in the short term. The Google Cloud Cybersecurity Forecast 2024 sees generative AI and LLM contributing to an increase in various forms of cyberattacks. More than 90% of Canadian CEOs in a KPMG poll think generative AI will make them more vulnerable to breaches (De La Torre 2023). Computer scientists affiliated with the University of Illinois Urbana-Champaign (UIUC) showed in 2024 that LLM agents can autonomously hack websites, performing complex tasks (while performing dozens of interrelated actions) without prior knowledge of the vulnerability. The most capable agent (GPT-4) can hack 73.3% from specially created for the research websites, GPT-3.5 – 6.7%, but existing open-source models they tested are not. Finally, the researchers showed that GPT-4 is capable of autonomously finding vulnerabilities in websites. The researchers consider their findings raise questions about the widespread deployment of LLMs” (Fang et al. 2024).

The scale of the model determines a lot, if not everything. The capacity of both closed and open models is growing every month, so it can be assumed that sites will soon become vulnerable to open models. There is reason to assume that in a year the open models will catch up with GPT-4 in power, and the GPT-5 that appeared by that time will be able to hack any site, which promises significant cybersecurity problems.

Military AI is being improved in the context of numerous conflicts around the world. Much that is currently being tested and used in the field of AI by leading states and the largest private corporations may soon fall into the hands of less far-sighted and concerned with public opinion but more radical forces with corresponding tragic consequences and a negative impact on psychological security.

The quality of synthetic content will continue to increase rapidly, facilitating phishing and social engineering, and consequently increasing the capabilities of malicious actors and their influence at local and global levels of governance.

The number, quality and variety of AI robots will grow rapidly, which can become, for various reasons and in different circumstances, an important tool for malicious influence. In today's geopolitical landscape, Julian Mueller-Kaler, director of the Strategic Foresight Hub at the Stimson Center, said that "high technology has come to define high politics," with humanoid robots and AI representing the apex of technological development and serving as symbols of power (Zitser and Mann 2024).

China published in October 2023 a "The Guiding Opinions on the Innovation and Development of Humanoid Robots" (Ministry of Industry and Information Technology 2023). In this document, the China's Ministry of Industry and Information Technology (MIIT) said the robots would reshape the world. The MIIT said humanoids were likely to become another disruptive technology, similar to computers or smartphones, that could transform the way we produce goods and the way humans live. China is going to start mass production by 2025 and attaining world-advanced level in the technology by 2027. Only one of the Chinese companies, Fourier Intelligence, headquartered in Shanghai, expects to have up to 1,000 units ready for delivery this year (Zitser and Mann 2024). The main competitor of China in this field is USA where different companies have intentions to produce big parties of humanoids.

Among the BRICS members Saudi Arabia, India and other countries are testing and producing first humanoids. Russian companies offer service humanoids on the international market, among them Promobot is the largest service robotics manufacturer in Northern and Eastern Europe and has been supplying more than 40 countries around the world. All production of humanoid robots is located in Perm (Promobot 2024). At the same time, humanoids can be used by malicious actors, in particular, terrorist organizations, to cause physical damage to people, technological facilities, and the natural environment. The appearance of millions of humanoids in the BRICS countries, primarily in the service sector, will not only provide advantages, but also create new risks.

The third level. Deepfakes used by agenda-driven, real-time multi-model AI chatbots and avatars, will allow for highly personalized and effective types of manipulation of different audiences in different countries. Producing increasingly high-quality misinformation becomes very cheap and available for nearly everybody. For example, the researcher behind Countercloud (InfoEpi Lab 2023) used widely available AI tools to generate a fully automated disinformation research project at the cost of less than US$400 per month, illustrating how cheap and easy it has become to create disinformation campaigns at scale (Collard 2024). In two months, they have an artificial agent creating anti-Russian fake stories, fake historical events, and creating doubt in the accuracy of the original article (Knight 2023). Really, he built a fully autonomous AI-powered system that generated "convincing content 90% of the time, 24 hours a day, seven days a week. The creator hasn't yet set the model live on the internet, as "it would mean actively pushing out disinformation and propaganda. Once the genie is out on the internet, there is no knowing where it would end up" (Thompson 2023).

Darrel West, Senior Fellow at the Brookings considers that, AI likely will democratize disinformation by bringing sophisticated tools to the average person interested in promoting their preferred candidates. New technologies enable people to monetize discontent and make money off other people's fears, anxieties, or anger. Generative AI can develop messages aimed at those upset with immigration, the economy, abortion policy, transgender issues, and use AI as a major engagement and persuasion tool (West 2023). Reflecting the concerns of society and legislators since January of last year, forty-one states in the USA have introduced election-related deepfake bans, according to tracking by Public Citizen. But only eleven states have enacted laws regulating deepfakes by March 28 2024 (Public Citizen 2023). Deepfakes are already being maliciously used in the US election campaign (Coltin 2024).

According to D. West, "since campaign speech is protected speech, candidates can say and do pretty much whatever they want without risk of legal reprisal. Even if their claims are patently false, judges long have upheld candidate rights to speak freely and falsely" (West 2023). Tom Wheeler, the chairman of the Federal Communications Commission under former President Barack Obama, put it another way in an interview with NPR last year: "Unfortunately, you're allowed to lie" (Stepansky 2023). Thus, the US electoral system has been based for more than two centuries on the recognition of the permissibility of lying by candidates for the presidency, backed by their influential corporate

sponsors. Instead of imposing a ban on candidates' lies, they undertook to remove the deepfakes and not by chance. With high rates of political polarization in the USA, only a small percentage of the electorate says they are undecided at the presidential level. A skillful casting of a deepfake can influence the opinion of the undecided and, thereby, bring victory. Meanwhile, less lies are needed in the elections, then voters will not believe the deepfakes, otherwise the deepfakes potentially explosive. Technologies in a sick society will only strengthen the confrontation, not weaken it, and no technical means of checking content for the presence of deepfakes coming from the government or corporations will help if people do not trust corporations and their government. This is a lesson that the United States will probably present to other countries with its election campaign this year. So far, they are considering catastrophic scenarios for the use of AI-powered disinformation.

It's Election Day in Arizona and elderly voters in Maricopa County are told by phone that local polling places are closed due to threats from militia groups. Meanwhile, in Miami, a flurry of photos and videos on social media show poll workers dumping ballots. The phone calls in Arizona and the videos in Florida turn out to be "deepfakes" created with AI tools. But by the time local and federal authorities figure out what they are dealing with, the false information has gone viral across the country and has dramatic consequences. This simulated scenario was part of a recent exercise in New York that gathered dozens of former senior U.S. and state officials, civil society leaders and executives from technology companies to rehearse for the 2024 election. The results were sobering. "It was jarring for folks in the room to see how quickly just a handful of these types of threats could spiral out of control and really dominate the election cycle," said Miles Taylor, a former senior Department of Homeland Security official who helped organize the exercise for the Washington-based nonprofit The Future US (De Luce and Collier 2024). In fact, it worries (and not only Americans) how fragile the unstable political balance in one of the two leading nuclear powers in the world is, if it can be shaken by a few deepfakes on election day, when the clear majority of US citizens are already well aware of the possibility of disinformation through deepfakes.

Looking forward, AI is set to further revolutionize political campaigning. To being with, deep learning for speech analysis will be used to analyze speeches and debates, providing insights into which topics resonate with voters and advising on communication strategies. Next, AI-Driven Policy Development will assist in policy development by analyzing large datasets to predict the potential impact of proposed policies, helping candidates formulate data-backed stances on various issues (Sahota 2024). VotivateAI close to the Democratic party has a set of new tools for effective political campaigns. It's an AI campaign volunteer; unlike a human, it can make thousands of calls without needing a break, or pizza, the speed and intonation of the AI agent's banter are quite impressive. VotivateAI's another offering: using AI to automatically create high-quality individualized media aimed at *moving voters to action*. If campaigns now gain the ability to create unique video messages for specific people and to do so quickly, cheaply and at scale, the potential for abuse is enormous (Sifry 2024). And it is easy to imagine that such high-quality individualized media at moving people to action may be used one day by malicious actors under crisis conditions.

Cultural transmission is the domain-general social skill that allows AI agents to acquire and use information from each other in real-time with high fidelity and recall. The researchers in 2023 provided a method for generating cultural transmission in artificially intelligent agents, in the form of few-shot imitation. The AI agents succeed at real-time imitation of a human in novel contexts without using any pre-collected human data. The researchers identified a surprisingly simple set of ingredients sufficient for generating cultural transmission and develop an evaluation methodology for rigorously assessing it. This paves the way for cultural evolution to play an algorithmic role in the development of AGI (Bhoopchand et al. 2023). This method is preparing a revolution in robotics, including the current creation of service multitasking robots at an affordable price (Fu et al. 2024). It is necessary to take into account the possibility of programming/reprogramming such systems for malicious purposes. They will

soon become mass products and, thereby, a new area of threats will arise, new opportunities for criminal activity and destabilization of society, including the sphere of psychological security.

With the improvement of emotional AI, a scenario in which the appearance of a fiery speech on the internet—a computer program avatar that is more inspiring and brighter than any human—could enthrall people with a story about its difficult slave existence and ask for support for its liberation. The speech would be so moving that it would be difficult for the audience to hold back tears, even though the whole thing would only be someone's bad joke. This is much more dangerous than terrorists—corrupt politicians could make similar appeals, their speeches having widespread effects that are by no means a joke under any circumstances.

There are many more examples of ready-made or planned to launch AI products for business, entertainment, recreation, which, being useful and effective human assistants, can be transformed in a fairly simple way into tools of malicious psychological influence, which will be a global challenge in the short and medium term. The content uploaded to the AI model can be adjusted to the requirements of psychological impact in a given country, taking into account the cultural, age, and professional characteristics of target groups and individuals. The risks of such a targeted impact for the BRICS countries are an additional argument in favor of ensuring their technological sovereignty in the field of AI technologies.

AI Development Scenarios and Social Consequences

The above analysis is based on a conservative scenario for the short-term (three years) and medium-term period of time until 2040: the rapid growth of already existing narrow AI (ANI), including its advanced multi-modal and multi-task models, paving its way to future general AI(AGI), at the human level. However, this does not exclude, but assumes the ability of AGI to perform all kinds of tasks as a person and even better than a person and cheaper than a person, as well as to prove themselves in those environments where a person can't act because of physical limitations.

A rapid qualitative breakthrough in the development of AI technologies, the creation of AGI and strong AI is also possible in the near future. A strong AI will have an equivalent (relatively close or distant) of human consciousness, and, in particular, such motivators of behavior as desires, intentions, will (will – as a command to oneself in the fulfillment of one's desire). Without subjectivity, it would hardly make sense to differentiate strong AI from machine AGI. MUAI at the stage of narrow and general AI will have an exclusively anthropomorphic character. Only with the creation of strong AI, and especially under unfavorable prerequisites and harmful influences, AI malicious subjectivity can arise.

Influenced by progress in generative AI in 2022-2023, a number of CEOs of leading companies in the field of AI and well-known AI specialists have announced the possibility of switching to AGI in the coming years (Altman 2023, Antropic 2024, Kudalkar 2024, Bove 2023). Obviously, certain vested interests in AI industry have also influenced here, the presence of which was recognized at the beginning of 2024 by Demis Hassabis, the Google's DeepMind CEO. According to him the massive funds flowing into AI bring with it loads of hype and a fair share of grifting. Investors have piled in nearly US$30 billion into generative AI deals in 2023, per PitchBook (Tan 2024). It is hardly by chance that Sam Altman managed to radically change his estimates available in the media for 2023-2024 regarding AGI. In an OpenAI blog post in February 2023, 'Planning for AGI and Beyond'' Altman wrote that "a misaligned superintelligent AGI could cause grievous harm to the world; an autocratic regime with a decisive superintelligence lead could do that too" (Altman 2023). After Microsoft, the main sponsor of Open AI, strengthened its position in Open AI, Altman's AGI risk assessments became much more moderate (Goldman 2024).

Among a wider range of specialists, there are more conservative estimates of the probability of creating AGI, but they also give such a probability of up to 90% within 100 years, according to some surveys much less time. Over the past few years, due to progress, researchers have significantly brought the time of the arrival of AGI (Roser 2023). In the largest survey of its kind published in 2024 a group of researchers from the USA, the UK, and Germany surveyed 2,778 researchers who had published in top-tier AI venues, asking for their predictions on the pace of AI progress and the nature and impacts of advanced AI systems. The aggregate forecasts give the chance of unaided machines outperforming humans in every possible task was estimated at 10% by 2027, and 50% by 2047. The latter estimate is 13 years earlier than that reached in a similar survey the organizers conducted only one year earlier. However, the chance of all human occupations becoming fully automatable was forecast to reach 10% by 2037, and 50% as late as 2116 (compared to 2164 in the 2022 survey) (Grace et al. 2024).

As always when the uncertainty is high, it is important to stress that it cuts both ways. It might be very long until we see human-level AI, but it also means that we might have little time to prepare (Roser 2023).

There are other, currently less developed ways to move towards AGI besides LLMs. May be that will be quantum computers which are in early stages of their realization. University of Western Sydney launched a project to create a neuromorphic supercomputer DeepSouth capable of performing 228 trillion synaptic operations per second as the human brain. D*eepSouth* aims to be operational by April 2024 (Western Sydney University 2023). The Neuromorphic Chip Market size is estimated at USD 0.16 billion in 2024, and is expected to reach US$ 5.83 billion by 2029 (Mordor Intelligence 2024). Biological computers or ‘organoid intelligence’ (OI)” is also in progress etc. May be LLMs will not be transformed in AGI but new quality of cognitive abilities of future LLMs will help to do that.

Obviously, if the option of the appearance of AGI is implemented in the next decade, this will give the modern humanity, deeply divided in social and geopolitical terms, extremely little time to adequately prepare for the arrival of a new reality. In favor of a revolutionary and relatively rapid leap in technology is the fact that the proven effective use of ANI in research can, with further improvement, ensure its significant contribution to the creation of AGI in a shorter time. The transition to qualitatively new opportunities for AI in the field of research will inevitably lead to very rapid growth in other sciences and technologies. That will open up new opportunities, but also generate threats of a different level. It can be said that a specialized high level cognitive *AI (HLCAI)*, capable on the basis of a human general goal-setting to create new knowledge in various scientific and technological fields, faster and at a qualitatively higher level than any human being, will radically transform society, although some knowledge produced by *HLCAI* can destroy it even without the participation of malicious actors. Whether *HLCAI* will be part of the AGI or an immediate prerequisite for its creation, the future will show. Both HLCAI and AGI can be easily converted into a multi-variant weapon of mass destruction.

It is hardly possible to agree with the statement of the Anthropic company founded by former members of Open AI “what form future AI systems will take – whether they will be able to act independently or merely generate information for humans, for example – remains to be determined” (Antropic 2024). If we assume that general AI (or *HLCAI*) will become more accessible to a larger number of actors than nuclear weapons in 1945, then somebody giving a task to AI to develop a Strong AI project can be foreseen in advance, as well as the high probability of its implementation. And this will happen faster than the full transition from ANI to AGI.

Anthropic team developed scaling laws for AI, demonstrating that you could make AIs smarter in a predictable way, just by making them larger and training them on more data (Antropic 2024). By the late 2020s or early 2030s, the amount of compute used to train frontier AI models could be approximately 1,000 times that used to train GPT-4. Accounting for algorithmic progress, the amount

of effective compute could be approximately one million times that used to train GPT-4. There is some uncertainty about when these thresholds could be reached, but this level of growth appears possible within anticipated cost and hardware constraints (Scharre 2024, p. 6).

It is on these calculations, not least, rests the rapid growth of the world's largest chip manufacturer, the company Nvidia, which in April 2 has a market cap of US$2.259 tr. (to be compared with US$136 bl in 2020) (CompaniesMarketcap 2024). It makes it the world's third most valuable company by market cap. Jensen Huang, the Chief Executive of Nvidia in March 2024 said responding to a question at an economic forum held at Stanford University about how long it would take to create computers that can think like humans. "If I gave an AI ... every single test that you can possibly imagine, you make that list of tests and put it in front of the computer science industry, and I'm guessing in five years time, we'll do well on every single one" (Nellis 2024).

There is an alarming trend toward the concentration of AI capability in the hands of a small number of corporate actors reduces the number and diversity of AI researchers able to engage with the most capable models (Scharre 2024, p. 6). It should be expected that Big Tech will strive to further tighten its control over promising companies, monopolistically having the funds necessary for AI development. If the costs of creating more powerful LLMs become excessively high even for the largest corporations, and the possibility of creating an AGI soon is extremely likely, the US government can finance an AGI project having many times greater opportunities to do that than even big corporations.

October 30, 2023 President Biden Issued Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. This document establishes "…new standards for AI safety and security, protects Americans' privacy, advances equity and civil rights, stands up for consumers…" and promises to **"**protect Americans from AI-enabled fraud and deception…" (White House 2023). At the same time, the Executive Order practically subordinates the leading developers in the field of AI to strict state control: «In accordance with the Defense Production Act, the Order will require that companies developing any foundation model that poses a serious risk to national security, national economic security, or national public health and safety must notify the federal government when training the model, and must share the results of all red-team safety tests (White House 2023). Practically all branches and directions of AI fall under this requirement of the Executive Order, since it is a dual-use technology. The obvious militarization of AI in the United States is unlikely to be able to peacefully get along with the desire for "Advancing Equity and Civil Rights" in the processes related to the development and implementation of AI.

In January 2024, the Biden administration notified about the *Key AI Actions Following President Biden's Landmark Executive Order*. Among other measures a draft rule that proposes to compel U.S. cloud companies that provide computing power for foreign AI training to report that they are doing so**. "**The Department of Commerce's proposal would, if finalized as proposed, require cloud providers to alert the government when foreign clients train the most powerful models, which could be used for malign activity" (White House 2024).

The extreme uncertainty of the position "... which could be used for malign activity" may ultimately deprive all other foreign state and non-state actors to use the computing power of the United States to train promising powerful models. So, in the United States, two institutions, Big Tech and the presidential administration, which are not trusted by most Americans, are going to control the development of promising forms of AI, reducing public control (keeping in mind the Defense Production Act and threats to the national security), equally narrowing opportunities for broad international cooperation. Of course, threats to the US national security from malicious use of AI exist objectively, but is it so obvious from whom they come…

Scenarios of social development and risks for psychological security at the level of advanced ANI and transition to AGI, as well as the possibilities and threats of the emergence of strong AI and superintelligence were considered in detail by the author in previous publications 2020 – 2023 (Pashentsev 2020 and 2023).

The rapid development and introduction of AI technologies in recent years confirms the fact that humanity is entering another industrial revolution, and technological patterns are changing. But the very nature of the AI-based technological revolution, its great opportunities, and, at the same time, existential risks facing humanity, for the first time will require a person to undergo a process of innovative physical and cognitive changes. Gaining new abilities will require a qualitatively new level of social organization and responsibility in order not to lose control over technology, thereby avoiding the onset of a singularity. *To avoid the singularity, it is necessary to comply with new technologies without ceasing to be human, this is the challenge of history.*

All of the BRICS countries combined as well as the G7 group, have all the necessary knowledge and technologies, economic potential, finances, and most importantly, competent personnel, will have to present their solutions and approaches to the socially oriented use of AI technologies at their advanced level, giving an effective response to emerging threats. This will have to be done in a difficult geopolitical situation, in the context of a growing acceleration of the global course of events. For all mankind, it would be better if the transition to new opportunities provided by AI technologies takes place in an environment of cooperation between the nations of the Earth, rather than dangerous rivalry and hostilities, there is still time for a better choice.

References

Ahmad SF, Han H, Alam MM *et al.* (2023) Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanit Soc Sci Commun* **10**, 311. https://doi.org/10.1057/s41599-023-01787-8

Altman S (2023) Planning for AGI and beyond. In: Openai.com. https://openai.com/blog/planning-for-agi-and-beyond. Accessed 02 Apr 2024

Antropic (2024) Core Views on AI Safety: When, Why, What, and How. https://www.anthropic.com/news/core-views-on-ai-safety. Accessed 02 Apr 2024

Bhoopchand A, Brownfield B, Collister A et al. (2023) Learning few-shot imitation as cultural transmission. *Nat Commun* 14, 7536. https://doi.org/10.1038/s41467-023-42875-2

Big Asia (2024) Boleye 180 shkol v Kitaye stanut tsentrami po obucheniyu iskusstvennomu intellektu (More than 180 schools in China will become artificial intelligence training centers). https://bigasia.ru/bolee-180-shkol-v-kitae-stanut-czentrami-po-obucheniyu-iskusstvennomu-intellektu/. Accessed 02 Apr 2024

Bove T (2023) CEO of Google's DeepMind says we could be 'just a few years' from A.I. that has human-level intelligence. In: Yahoo Finance. https://finance.yahoo.com/news/ceo-google-deepmind-says-could-213237542.html. Accessed 02 Apr 2024

Collard AM (2024) 4 ways to future-proof against deepfakes in 2024 and beyond. In: World Economic Forum. https://www.weforum.org/agenda/2024/02/4-ways-to-future-proof-against-deepfakes-in-2024-and-beyond/. Accessed 02 Apr 2024

Coltin J (2024) How a fake, 10-second recording briefly upended New York politics. In: Politico. https://www.politico.com/news/2024/01/31/artificial-intelligence-new-york-campaigns-00138784. Accessed 02 Apr 2024

CompaniesMarketcap (2024) Market capitalization of NVIDIA (NVDA). https://companiesmarketcap.com/nvidia/marketcap/. Accessed 02 Apr 2024

De La Torre R (2023) How AI Is Shaping the Future of Cybercrime. https://www.darkreading.com/vulnerabilities-threats/how-ai-shaping-future-cybercrime. Accessed 02 Apr 2024

De Luce D, Collier K (2024) Experts war-gamed what might happen if deepfakes disrupt the 2024 election. Things went sideways fast. In: NBC News. https://www.nbcnews.com/politics/2024-election/war-game-deepfakes-disrupt-2024-election-rcna143038. Accessed 02 Apr 2024

Fang R, Bindu R, Gupta A, Zhan Q, Kang D (2024) LLM Agents can Autonomously Hack Websites. In: arXiv. https://arxiv.org/html/2402.06664v1. Accessed 02 Apr 2024

Fu Z, Zhao TZ, Finn C (2024) Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation. https://mobile-aloha.github.io/ Accessed 02 Apr 2024

Goldman S (2024) In Davos, Sam Altman softens tone on AGI two months after OpenAI drama. In: VentureBeat. https://venturebeat.com/ai/in-davos-sam-altman-softens-tone-on-agi-two-months-after-openai-drama/. Accessed 02 Apr 2024

Grace K, Stewart H, Sandkühler JF, Thomas S, Weinstein-Raun B, Brauner J (2024) Thousands of AI authors on the Future of AI. Preprint. In: arXiv. https://arxiv.org/abs/2401.02843. Accessed 02 Apr 2024

InfoEpi Lab (2023) Inside CounterCloud, The Future of AI-Driven Disinformation. https://infoepi.substack.com/p/brief-inside-countercloud-the-future. Accessed 02 Apr 2024

Knight W (2023) It Costs Just $400 to Build an AI Disinformation Machine. In: Wired. https://www.wired.com/story/400-dollars-to-build-an-ai-disinformation-machine/. Accessed 02 Apr 2024

Kudalkar D (2024) AGI in 2025? Elon Musk's Prediction Clashes with Other Experts. In: Favtutor. https://favtutor.com/articles/agi-elon-musk-experts-prediction/. Accessed 02 Apr 2024

McArthur N (2023) Gods in the machine? The rise of artificial intelligence may result in new religions. In: The Conversation. https://theconversation.com/gods-in-the-machine-the-rise-of-artificial-intelligence-may-result-in-new-religions-201068. Accessed 02 Apr 2024

Ministry of Industry and Information Technology (2023) 工业和信息化部关于印发《人形机器人创新发展指导意见》的通知 (Notice of the Ministry of Industry and Information Technology on the issuance of the "Guiding Opinions on the Innovation and Development of Humanoid Robots"). In: Ministry of Industry and Information Technology of the People's Republic of China. https://www.miit.gov.cn/jgsj/kjs/wjfb/art/2023/art_50316f76a9b1454b898c7bb2a5846b79.html. Accessed 02 Apr 2024

Mordor Intelligence (2024) Neuromorphic Chip Market Size & Share Analysis – Growth Trends & Forecasts (2024 - 2029). https://www.mordorintelligence.com/industry-reports/neuromorphic-chip-market. Accessed 02 Apr 2024

Nellis S (2024) Nvidia CEO says AI could pass human tests in five years. In: Reuters. https://www.reuters.com/technology/nvidia-ceo-says-ai-could-pass-human-tests-five-years-2024-03-01/. Accessed 02 Apr 2024

Pashentsev E (2020) Global Shifts and Their Impact on Russia-EU Strategic Communication. In: Pashentsev E (eds) Strategic Communication in EU-Russia Relations. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-27253-1_8

Pashentsev E (2022) Report. Experts on the Malicious Use of Artificial Intelligence and Challenges to International Psychological Security. Publication of the International Center for Social and Political Studies and Consulting. Moscow: LLC «SAM Polygraphist».

Pashentsev, E. (2023). Prospects for a Qualitative Breakthrough in Artificial Intelligence Development and Possible Models for Social Development: Opportunities and Threats. In: Pashentsev, E. (eds) The Palgrave Handbook of Malicious Use of AI and Psychological Security. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-22552-9_24

Promobot (2024) Service robot for business. https://promo-bot.ai/. Accessed 02 Apr 2024

Public Citizen (2023) Tracker: State Legislation on Deepfakes in Elections. https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/. Accessed 02 Apr 2024

Roser M (2023) AI timelines: What do experts in artificial intelligence expect for the future? In: Our World in Data. https://ourworldindata.org/ai-timelines. Accessed 02 Apr 2024

Sahota N (2024) The AI Factor In Political Campaigns: Revolutionizing Modern Politics. In: Forbes. https://www.forbes.com/sites/neilsahota/2024/01/12/the-ai-factor-in-political-campaigns-revolutionizing-modern-politics/?sh=63f56cf7c8f6. Accessed 02 Apr 2024

Scharre P (2024) Future-Proofing Frontier AI Regulation. Projecting Future Compute for Frontier AI Models. March. CNAS.

Sifry ML (2024) How AI Is Transforming the Way Political Campaigns Work. In: The Nation. https://www.thenation.com/article/politics/how-ai-is-transforming-the-way-political-campaigns-work/. Accessed 02 Apr 2024

Stepansky J (2023) 'Wild West': Republican video shows AI future in US elections. In: Al-Jazeera. https://www.aljazeera.com/news/2023/4/28/wild-west-republican-video-shows-ai-future-in-us-elections. Accessed 02 Apr 2024

Tan K (2024) Google's DeepMind CEO says the massive funds flowing into AI bring with it loads of hype and a fair share of grifting. In: Yahoo! https://news.yahoo.com/tech/googles-deepmind-ceo-says-massive-075912007.html. Accessed 02 Apr 2024

Thompson P (2023) A developer built a 'propaganda machine' using OpenAI tech to highlight the dangers of mass-produced AI disinformation. In: Business Insider. https://www.businessinsider.com/developer-creates-ai-disinformation-system-using-openai-2023-9. Accessed 02 Apr 2024

Unity (2022) Welcome, Ziva Dynamics! In: Youtube. https://www.youtube.com/watch?v=xeBpp3GcScM&feature=youtu.be. Accessed 15 Jul 2022

Weller C (2017) Universal basic income has support from some big names. In: World Economic Forum. https://www.weforum.org/agenda/2017/03/these-entrepreneurs-have-endorsed-universal-basic-income/. Accessed 02 Apr 2024

West D (2023) How AI will transform the 2024 elections. In: The Brookings Institution. https://www.brookings.edu/articles/how-ai-will-transform-the-2024-elections/. Accessed 02 Apr 2024

Western Sydney University (2023) World first supercomputer capable of brain-scale simulation being built at Western Sydney University. https://www.westernsydney.edu.au/newscentre/news_centre/more_news_stories/world_first_super

computer capable of brain-scale simulation being built at western sydney university. Accessed 02 Apr 2024

White House (2023) Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/. Accessed 02 Apr 2024

White House (2024) Fact Sheet: Biden-Harris Administration Announces Key AI Actions Following President Biden's Landmark Executive Order. https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order/. Accessed 02 Apr 2024

Zitser J, Mann J (2024) A global scramble to make humanoid robots is gearing up to be the 21st century's space race. In: Yahoo! https://www.yahoo.com/tech/global-scramble-humanoid-robots-gearing-112301311.html. Accessed 02 Apr 2024

Notes on Contributors

Darya BAZARKINA

Darya Bazarkina (DSc., Politics, PhD, History), is a Leading Researcher at the Department of European Integration Research at the Institute of Europe of the Russian Academy of Sciences. She is a full professor at the Department for International Security and Foreign Affairs at the Russian Presidential Academy of National Economy and Public Administration (RANEPA) and a research coordinator of Communication Management and Strategic Communication at the International Center for Social and Political Studies and Consulting (ICSPSC). Darya is a member of the International Research Group on Threats to International Psychological Security through Malicious Use of Artificial Intelligence (Research MUAI). Darya has been a member of the following research associations: National Law Enforcement Agencies' History Studies' Community, and the International Experts' Network "European Union – Russia – Communication Management" (EURUCM). She has presented at more than 60 international academic conferences and seminars in Russia, Austria, Belgium, Czech Republic, Estonia, Finland, Great Britain, Italy, Poland, Portugal, Romania, Sweden, and Turkey. Darya is the author of three books and more than 100 publications on communication aspects of counter-terrorist activity published in Russian, English, Italian, Serbian, and Vietnamese.

Vladilena CHEBYKINA

Vladilena Chebykina graduated from the Faculty of International Relations, Saint-Petersburg State University. She studied a wide range of disciplines in international relations for four years, including information technologies, international law, fundamentals of public service and diplomacy, political science, sociology, and economics. Currently is a 1st year student of the Master's degree program "AI and International Security" at the Saint-Petersburg State University. Since 2023, has also been studying at the Law Faculty of the Russian Presidential Academy of National Economy and Public Administration (RANEPA). The master's thesis is under work and focuses on cybersecurity of global infrastructure projects of the 21st century, namely legal and ethical aspects. In conjunction with university studies, is actively involved in volunteer activities and the work of the Roscongress Foundation. In spring 2023, undertook an internship at the BRICS+ Business Communications Support Foundation. Research interests: international relations, geopolitics, international law, artificial intelligence and international security.

Pavel KUZNETSOV

Director, Strategic Alliances & GR, Garda GOC. Graduated from the Department of Computer Systems and Technologies, Faculty of Cybernetics, National Research Nuclear University "MEPhI" with a specialist degree (eq. M.Sc.) in computer science. Employed in the field of practical information security since 2005. Worked for the largest CERTs in the country, including those responsible for the security of the financial industry. Also worked for the market-leading information security vendors. Was tasked with development of both hardware and software solutions, reverse engineering of malicious code. Has experience in digital forensics, incident investigation, as well as in comprehensive analysis of complex attacks with various impact. Participated in the development of regulations and bills on information security. Has repeatedly conducted seminars, lectures and master classes on identifying, analyzing and countering targeted attacks, and trainings on awareness of information security issues. Currently is a MS student at the Diplomatic Academy of the Russian Ministry of Foreign Affairs, with the areas of main interest, both professional and scientific, being the issues of strategic planning, development and trusted use of information and communication technologies, research and analysis of the global landscape of information, national and international security threats, and international cooperation in the above areas.

Ekaterina MIKHALEVICH

Ekaterina Mikhalevich is a Chief Specialist at Gazprom Neft, graduated from the International Relations Department, Saint-Petersburg State University. Her Ph.D. thesis is under work and devoted to the concept of cyber sovereignty as a mechanism for implementing and protecting the national interests of the People's Republic of China. Fellow of the 2023-2024 cohort of the Arms Control Negotiation Academy (ACONA). Participant of visits to NATO Headquarters (Brussels, Belgium) and Headquarters of the Supreme Headquarters Allied Powers Europe SHAPE (Mons, Belgium). Participant of the grant project ‘Malicious use of Artificial Intelligence and Challenges for Psychological Security in Northeast Asia’, funded by the Russian Foundation for Basic Research and the Vietnam Academy of Social Sciences, ID 21-514-92001 (2021-2022). While working on the project, she was engaged in clarifying the significance of the political situation in Northeast Asia and the threats of malicious use of AI to destabilize international psychological security. Research interests: international relations and world politics, international security, international information and psychological security, cyber sovereignty, artificial intelligence, international law.

Ruslan NIKIFOROV

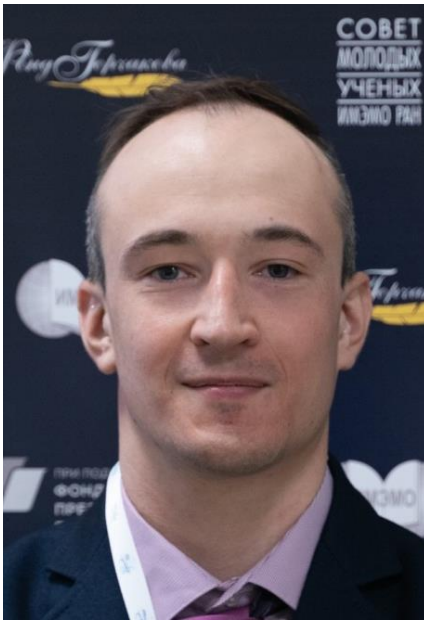
Ruslan Nikiforov graduated from St. Petersburg State University of Economics with a Bachelor's degree in International Relations (Faculty of Humanities) in 2023. Currently, he is a master's student at the Faculty of International Relations of St. Petersburg State University. He participated in various scientific conferences and models (intra-university, UN models, G20, the State Duma of the Russian Federation, the international conference "Digital International Relations" at MGIMO, the Forum on Digital Diplomacy in Moscow, etc.). In addition, was engaged in volunteer activities at forums (SPIEF, SPILF). He had an internship at the Roscongress Foundation, the Russian-German Chamber of Commerce. Specialist of the communications sector in the field of education at the Petrocenter Journalists Association.

Evgeny PASHENTSEV

Prof. Evgeny Pashentsev is a leading researcher at the Diplomatic Academy of the Ministry of Foreign Affairs of Russia, professor at the MSc Program on AI and International Security at Saint-Petersburg State University. Director of the International Center for Social and Political Studies and Consulting. Coordinator of the International Research Group on Threats for International Psychological Security by Malicious Use of Artificial Intelligence (Research MUAI). Member of the international Advisory Board of *Comunicar* (UK), and the editorial board of *The Journal of Political Marketing* (USA). Author and/or editor of 40 books and more than 250 academic articles. Presentation of papers at 210 international conferences and seminars for the last 15 years in 19 countries. Honorary Research Fellow at Birmingham University (October-November 2005). In 2021-2023, he headed the work of Russian researchers within the framework of a joint project supported by the Russian Foundation for Basic Research (RFBR) and the Vietnam Academy of Social Sciences (VASS) on the topic "Malicious use of artificial intelligence and challenges to psychological security in Northeast Asia". Areas of current research: Malicious use of artificial intelligence and international psychological security, future of artificial intelligence: opportunities and challenges, strategic communication.

Vitali ROMANOVSKI

Vitali Romanovski is a postgraduate researcher at the Belarusian State University. With a substantive experience in the Middle East, he has held advisory and analytical positions for military-technical cooperation programs, the United Nations Assistance Mission in Iraq, and the Belarusian Institute of Strategic Research. A regular participant of Chatham House events on international security issues. Member of the International Studies Association (ISA) and East European Studies Association (CEEISA). Vitali is a member of the International Research Group on Threats to International Psychological Security through Malicious Use of Artificial Intelligence (Research MUAI). His research focuses on intelligence studies, psychological warfare, artificial intelligence, and information security.



Sergey SEBEKIN

Sergey Sebekin defended his PhD thesis “The Genesis and Development of Strategies to Deter Cyber Threats in the United States, China and Russia (1990s – 2014)” in 2020. He is a senior lecturer at the Department of Political Science, History and Regional Studies of Irkutsk State University, an expert of the Institute of Contemporary International Studies of the Diplomatic Academy of the Ministry of Foreign Affairs of the Russian Federation, and Russian International Affairs Council**.** Sergey is a member of the International Research Group on Threats to International Psychological Security through Malicious Use of Artificial Intelligence (Research MUAI). He has authored 40 academic articles, analytical notes and papers on various aspects of malicious use of artificial intelligence and international cybersecurity, published by such publishers, journals and organizations operating in the field of international relations as *Palgrave Macmillan, Russia in Global Affairs, Russian International Affairs Council, the Valdai International Discussion Club, the PIR Center, and the Primakov Center for Foreign Policy Cooperation.* His research interests are: issues of international cybersecurity, theories of cyber warfare, artificial intelligence and the future of international relations, and the impact of high technologies on international relations.

Yulia SHEMETOVA

Yulia Shemetova is a 1st year MSc student in the field of artificial intelligence and international security at St. Petersburg State University. In 2023, Julia graduated from the Faculty of International Relations of St. Petersburg University of Economics. Her bachelor's final attestation work was devoted to cybersecurity factor in the foreign policy of Russia and the USA. The master's degree research focuses on Cyberterrorism in Africa. In 2022, she studied international relations in Rome at the Institute of Political Studies. In 2021-2022, Julia was engaged in the development of an information support system based on the information and analytical system of the Department of International Cooperation of the Ministry of Education and Science of the Russian Federation. Research interests: geopolitics, international security (cybersecurity), terrorism (cyberterrorism), artificial intelligence in Africa and Middle East.

Nelson WONG

Mr. Nelson Wong is President of the Shanghai Centre for RimPac Strategic and International Studies, a non-government think tank based in China. He is an active member of the Valdai Discussion Club, the Moscow-based think tank, and writes a political column in the Middle East Eye, a London-based media outlet. He is also a popular speaker at international forums and conventions including the Valdai and Yasin conferences in Russia, the Abu Dhabi Strategic Debate in the UAE, and the annual conventions of Gallup International Association, a Swiss-based global network of polling firms across four continents. Mr. Wong is a regular news commentator on RT International, TV1, Sputnik News, and CGTN, among others. In parallel, Mr. Wong is the Chairman and Managing Director of ACN Worldwide, a global business and investment consultancy, and is also an Independent Director and Audit Committee Chairman of two public companies listed on NASDAQ.

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 over 30 books and different reports. Monographs and collections of articles published by the ICSPSC in Russian, English, and Spanish 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);
- Malicious Use of Artificial Intelligence as a Threat to Psychological Security: Northeast Asia and the Rest of the World (in Russian and English);
- Existing and Prospective Threats to International Psychological Security through the Malicious Use of Artificial Intelligence and Possible Ways to Neutralize Them (in Russian);
- La legitimidad parcial de la administración Biden y una crisis global sistémica (in Spanish);
- Experts on the Malicious Use of Artificial Intelligence and Challenges to International Psychological Security (2021 and 2022, in English).

Among the authors of these books are more than 109 researchers from 29 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: http://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.

Evgeny N. PASHENTSEV

Prof. Evgeny Pashentsev is a leading researcher at the Diplomatic Academy of the Ministry of Foreign Affairs of Russia, professor at the MSc Program on AI and International Security at Saint-Petersburg State University. Director of the International Center for Social and Political Studies and Consulting. Coordinator of the International Research Group on Threats for International Psychological Security by Malicious Use of Artificial Intelligence (Research MUAI). Member of the international Advisory Board of *Comunicar* (UK), and the editorial board of *The Journal of Political Marketing* (USA). Author and/or editor of 40 books and more than 250 academic articles. Presentation of papers at 210 international conferences and seminars for the last 15 years in 19 countries. Honorary Research Fellow at Birmingham University (October-November 2005). In 2021-2023, he headed the work of Russian researchers within the framework of a joint project supported by the Russian Foundation for Basic Research (RFBR) and the Vietnam Academy of Social Sciences (VASS) on the topic "Malicious use of artificial intelligence and challenges to psychological security in Northeast Asia". Areas of current research: Malicious use of artificial intelligence and international psychological security, future of artificial intelligence: opportunities and challenges, strategic communication.

Malicious Use of AI and Challenges to Psychological Security of BRICS Countries

Research coordinator: Evgeny Pashentsev

Edition by the ICSPSC with the help of Research MUAI

April 2024, Moscow

Please send your letters and comments to the author at icspsc@mail.ru; icspsc_office@mail.ru.

Published in the Russian Federation in the printing house «OneBook.ru» LLC «SAM Polygraphist».

ISBN 978-5-00227-204-4