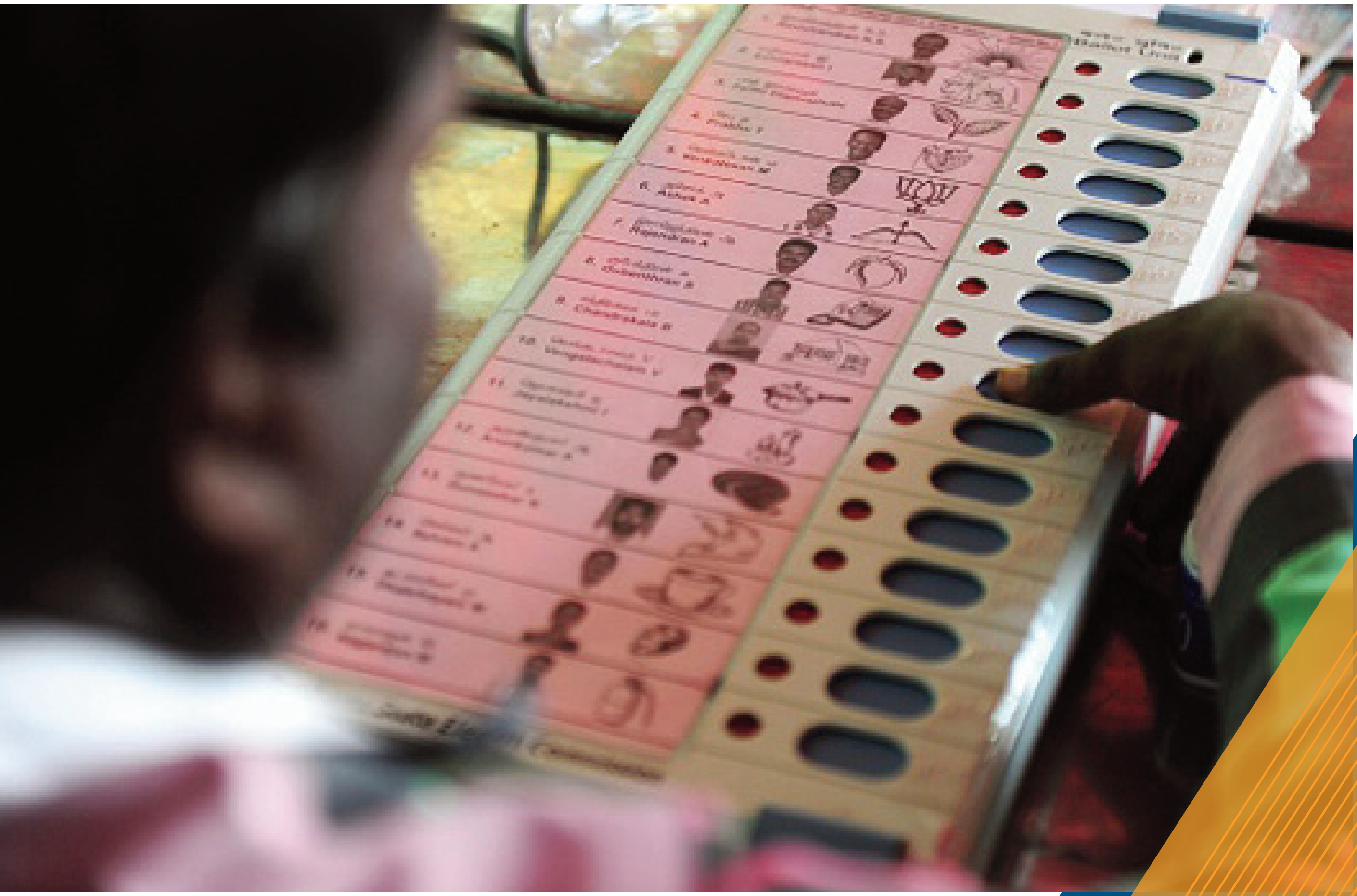
CYBER-AI CONVERGENCE AND INTERFERENCE

Securing Elections and Building Human Resilience

By Eleonore Pauwels



NOVEMBER 2020

The views and opinions expressed in this paper are those of the author and do not necessarily reflect the official policy or position of Konrad-Adenauer-Stiftung

Table of Content

Preface

Elections are a key element of any democracy. However, we have seen in the past the fallacy of electoralism[1] and the temptation by many external actors to declare a political system as democratic just because of regular electoral exercises. Often the quality of elections as such has been disregarded, or deficiencies in the electoral process were identified, but persist without any consequences. According to widely recognized international standards democratic elections have to be free, meaning the rights of citizens to participate and to compete are respected and protected by the rule of law. Democratic elections are equally meant to be fair, meaning that a level playing field should exist. But what do these minimal standards mean in the age of artificial intelligence and new technologies. Election campaigns as well as the electoral process run by Electoral Management Bodies are increasingly digitized and thus vulnerable to AI-enabled cyberattacks. The convergence of cybersecurity with AI and other emerging technologies enhances the risks for subversive attacks jeopardizing the conduct of free and fair elections, and thus the integrity and legitimacy of the entire electoral process.

The present study "Cyber-AI Convergence and Interference: Securing elections and building human resilience" by Eleonore Pauwels dissects how the convergence of AI with cyber- and information offensive operations impacts the security of elections. It provides a matrix of an electoral cycle analyzing how the convergence of AI and cybersecurity impacts the landscape of threats and vulnerabilities of an electoral process. It identifies data-targets as well as human vulnerabilities and attack-vectors tailored to electoral processes.

For a political foundation such as Konrad-Adenauer-Stiftung who has in its mandate the support of democratization processes worldwide, the impact of new technologies on electoral processes and the state of our democracies is of utmost interest.

It affects consolidated as well as emerging democracies. The threats that we are facing are manifold and they go way beyond the erosion of institutions. They particularly impact and dramatically change the social fabric and the political culture in our societies. A transformation that certainly also has its positive sides as long as the negative side-effects and collaterals are reigned in. But particularly the latter has never been as complex before.

In defense of democracy we can identify two frontlines:

We have the political space, the ambit where candidates and parties are campaigning, seeking popular support and where online defamation, hate-speech, data leaks, disinformation and deep-fakes can alternate the level playing field. It is this level, the capturing of the hearts and minds of citizens, which a previous study by the author "The Anatomy of Information Disorders in Africa" dissected in detail and illustrated with examples from the African continent.

But we also have the technical space, analyzed in the present study, where particularly Electoral Management Bodies are the most vulnerable institutions. It is a sphere where data manipulation by local or foreign actors can disrupt an electoral process, and where competing political parties need to have sufficient expertise on technologies used in order to understand and to prevent any electoral fraud.

In order to gain further insights into the vulnerability of the electoral cycle to modern technology, KAS New York embarked together with the author on this broader research project that besides of the use of AI to generate hyper-targeted disinformation campaigns, data-manipulation and cyber/AI-enabled

[1] A term coined by political scientist Terry Lynn Karl.

cognitive-emotional conflicts and disinformation also addresses pertinent questions such as how fit for purpose are electoral laws in the context of today's technological abilities? And how can security and resilience of election infrastructure be guaranteed best?

The results of these analyses are meant to assist and to sensitize Electoral Management Bodies, law makers, political party representatives, media and civil society to the emerging threats which jeopardize the democratic character of elections and bring about wide-spread repercussions for the political culture of societies.

It also reaches out to international organizations who often assist in election management or election observation and who need to take into account the possible distortions which easily might get unnoticed.

KAS New York wishes all stakeholders and the interested public an interesting read!

Andrea E. Ostheimer
Executive Director
Konrad-Adenauer-Stiftung, New York

Executive Summary

Far beyond what was conceived through traditional security and military doctrines, we face new challenges that pertain to human and political security. What matters is not only who wins new territories, but who wins the data, the trust, the hearts and minds of citizens within a country or polity.

For a decade, malign foreign powers have weaponized the infrastructure that underpins democratic societies. They have hacked the Internet, media, and even voting databases to sow confusion, discontent, and distrust. From the 2016 Brexit referendum, to the 2016 U.S. presidential primaries and general election, to the 2017 French presidential election, foreign meddlers have systematically sought to skew the democratic debate.

> An emerging typology of cyberattacks could leverage adversarial AI to manipulate the integrity of datasets and software involved in the electoral process.

Both state and non-state actors are already using the convergence of artificial intelligence (AI) and cyber-capabilities to manipulate information, erode trust, interfere with the internal political processes of other states, or to paralyze infrastructure critical to national and human security.

This technological convergence has significant adversarial, social, and even strategic implications. For instance, using AI systems could drastically amplify the nature, scope, and intensity of cyberattacks on member states' critical election infrastructure. An emerging typology of cyberattacks could leverage adversarial AI to manipulate the integrity of datasets and software involved in the electoral process. Such adversarial attacks already harness techniques to evade detection, target human vulnerabilities through precision social engineering and, ultimately, impact cyber and information security. As underlined by David Schwed, Professor and Founding, Director of Cybersecurity Program at Yeshiva University's Katz School, "AI will take a more prevalent role in malicious actors' attack arsenals. They will be able to launch unlimited autonomous attacks with a reduced need for human intelligence."[1]

States will learn to live with these electoral cyber-threats, just as they are learning to apprehend the shifting nature and scope of low-intensity cyber-conflict. The primary concern will be that, with AI and increasing cyber interconnectedness, these threats to election security will become more complex, difficult to prevent and detect. They will target national information infrastructure, undermining the integrity of sensitive security and civilian biometrics data. As more devices are being connected to the Internet – from personal sensors to elements of critical infrastructures, the deployment of 5G will accelerate AI processing at the edge-device. The opportunities for and destructiveness of sophisticated types of electoral cyberattacks are only going to increase.

An emerging typology of cyberattacks could leverage adversarial AI to manipulate the integrity of datasets and software involved in the electoral process.

[1] Forbes, 2019. "141 Cybersecurity Predictions for 2020." https://www.forbes.com/sites/gilpress/2019/12/03/141-cybersecurity-predictions-for-2020/

> With AI and increasing cyber interconnectedness, these threats to election security will become more complex, difficult to prevent and detect.

This technical and policy brief will explore challenges related to the use of AI and cyber-technologies and how they are being considered within broader election support strategies,[2] including those of for example IFES[3] and EU member states.[4] In particular, this brief will explore how digital tools used to administer and support electoral processes are increasingly vulnerable to AI-driven malware and cyber-attacks, indicating the need for new approaches to ensure the security and resilience of election infrastructure. The capacity of autonomous malware to improve upon their own strategies and launch increasingly aggressive, precise counter-attacks with each iteration leads to an expansion and augmentation of existing cyberattack capabilities. The automation of cyberattacks that can manipulate and corrupt the integrity of critical information within election infrastructure is a growing threat triggered by technological convergence. Defining the evolving threats' landscape, this report will dissect the election cycle and infrastructure to identify entry-points for converging Cyber-AI attack vectors, detect related data-targets and vulnerabilities and propose recommendations.

In the wake of a few seminal reports,[5] the author choses to rely on a unique perspective which approaches the election cycle and its infrastructure, first, as a set of complex socio-technical systems and, second, as a set of data-driven processes. This approach is holistic and strategic as it allows us to anticipate, reframe and better understand the emerging types of vulnerabilities that AI will increasingly be able to target within the data-infrastructures and data-optimization processes related to the conduct of elections. In our view, what matters is not primarily discussing the level of digitization of discrete steps in an election process, but anticipating threats to data integrity within the full information life cycle of an electoral process.

Manipulating data integrity is a new and extremely powerful tactic for those who wish to sow deception and mistrust in critical socio-technical systems. Elections – like other critical data-driven infrastructures in health and emergency relief – are vulnerable to emerging techniques of data-manipulation and poisoning.[6] And, like trust in health services and disaster management, trust in elections is at the core of our social contract; even more, it is the foundation of our democracies.

> The automation of cyberattacks that can manipulate and corrupt the integrity of critical information within election infrastructure is a growing threat triggered by technological convergence.

In the election security context, where we crucially need to build and reinforce trust, the advent of AI is an epistemic shift as much as a technological one. The techniques of AI promise to help us produce,

[2] This report is based on both, primary and secondary resources, using a mixed-methods approach comprised of qualitative desk research, literature reviews, policy analyses, expert interviews and consultations, as well as foresight methodologies (signals, drivers and trends impact analysis). The list of primary and secondary resources is provided in Annex.

[3] International Foundation for Election Security (IFES), 2018. "Cybersecurity in Elections." https://www.ifes.org/sites/default/files/2018_heat_cybersecurity_in_elections.pdf

[4] NIS Cooperation Group, 2018. "Compendium on Cyber Security of Election Technology." https://www.ria.ee/sites/default/files/content-editors/kuberturve/cyber_security_of_election_technology.pdf

[5] NIS Cooperation Group, 2018.; IFES, 2018.; Herpig S., et al, 2018. "Securing Democracy in Cyberspace." https://www.stiftung-nv.de/sites/default/files/securing_democracy_in_cyberspace.pdf

[6] Data Poisoning is an adversarial attack that aims to manipulate the training dataset in order to control the prediction behavior of a trained algorithmic model such that the model will label malicious examples into a desired category (e.g., labeling spam e-mails as safe). Data poisoning attacks can therefore subvert the learning process for the machine learning system and/or degrade the performance of the system.

analyse, assert, and verify a complex body of knowledge. Yet, these techniques could also undermine the integrity and credibility of our global intelligence and information systems.[7]

Section 1 of this report traces the recent trends and research in cybersecurity of election technology, providing an initial diagnosis of how traditional, legacy[8] approaches in electoral cybersecurity have to adapt to the convergence of cyber- and AI techniques. Section 2 offers a paradigm of AI convergence and explains how this paradigm produces an array of AI-enabled cyberattacks able to target and manipulate the integrity of data-sets. Section 3 provides a general matrix of an election cycle and infrastructure, analysing the landscape of converging threats and vulnerabilities, attack vectors, data-targets and implications. Section 4 concludes with recommendations, in particular reflections on needs and methods for adversarial threat assessment, while reflecting on the role and responsibilities of a diversity of stakeholders.

[7] Pauwels E., 2019. "The New Geopolitics of Converging Risks." https://collections.unu.edu/eserv/UNU:7308/PauwelsAIGeopolitics.pdf

[8] In the context of this technical and policy brief on AI and cybersecurity, the term "legacy" describes a system or an approach that is old, traditional ("inherited") but is still used as a reference because it would be too difficult to replace it. The required knowledge and foresight to update and replace such "legacy" approach has not been acquired and achieved yet.

1

Existing and Emerging Trends in Electoral Cybersecurity

While several member states have already witnessed instances of election interference, the most drastic and comprehensive attempt at disrupting strategic processes of an entire election cycle took place in a powerful tech-leading country.

Over the course of the US 2016 Presidential election, officers of the Main Intelligence Directorate of the General Staff of the Russian Army (GRU) allegedly conducted an unprecedented, coordinated cyber-campaign against state election infrastructure.[9] They scanned voter registration databases for vulnerabilities and targeted state websites in at least 21 states before Election Day. Using malicious code injection (known as SQL injection), GRU officers fully accessed several states' electoral systems and stole hundreds of thousands of voters' personal information. The Senate Intelligence Committee reports that in a small number of states, they penetrated restricted elements of election infrastructure and were in a position to, at a minimum, alter or delete voter registration data.[10] In August 2016, GRU officers also targeted, through spear-phishing emails, a voter registration software vendor and impersonated the company's employees sending malicious emails to several Florida election administrators.

Targeting and infecting weak links through spear-phishing, social engineering and remote access trojans, the attackers hacked the Hillary Clinton campaign, the Democratic Congressional Campaign Committee and the Democratic National Committee.

This unparalleled, complex sequence of attacks targeting election security unveils a set of existing cyberthreats that states and electoral management bodies urgently need to prepare for. Yet, it is also likely that both, external adversaries and insider threats, will keep adapting and upping their game, amplifying the existing threat landscape through technological convergence.

Elections as Complex Data-driven Processes

Electoral systems in a growing number of countries will come under adversarial pressure, with diversifying offensive techniques: these attacks target not only the functioning of physical election infrastructure, but also its trove of sensitive data; they are not perpetrated by humans or bots acting separately, but by a complex alliance of human-machine deceptive tactics.

Several seminal reports[11] on election cybersecurity have started showing the need for models of threat assessment and prevention that are more holistic, anticipatory, departing from reductionist legacy approaches to better understand the full range of technical, human, political and procedural vulnerabilities in an election cycle. IFES' comprehensive report, *Cybersecurity in Elections*, provides a crucial, in-depth account of the challenges faced by Electoral

[9] Special Counsel Robert S. Mueller, III., "Report On The Investigation Into Russian Interference In The 2016 Presidential Election: Volume I," U.S. Department of Justice, March 2019, 49-51 ("Mueller Report"). https://www.justice.gov/storage/report.pdf

[10] United States Senate. "Report Of the Select Committee on Intelligence On Russian Active Measures Campaigns and Interference in the 2016 U.S. Election," Volume 1: Russian Efforts Against Election Infrastructure with Additional Views. Report 116-XX https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume1.pdf

[11] NIS Cooperation Group, 2018.; IFES, 2018.; Herpig S., et al, 2018.

Management Bodies (EMBs) when working towards protecting digital assets, infrastructures and procedures in elections. While out of scope for this report, it is important to note that other resources by the Carter Center, the Organization of American States (OAS), and the Organization for Security and Cooperation in Europe (OSCE) have begun reflecting on the expertise and skills needed by electoral observation missions to assess the functioning of digitized electoral processes.[12]

The IFES report emphasizes several aspects that resonate with this report: first, EMBs need to manage a difficult tension between transparency and cybersecurity, not only to secure the conduct of elections, but also to promote accountability of the electoral process and trust in the subsequent results. Second, the IFES report insists on supporting EMBs with adequate expertise and models to identify, understand and respond to future vulnerabilities in electoral cybersecurity, beyond learning from past threats and recent failures. In light of these challenges, the IFES report "outlines strategies for EMBs to strengthen their technology and procedures to resist vulnerabilities," by designing what IFES calls "Holistic Exposure and Adaptation Testing (HEAT) process."[13] The HEAT process offers a unique contribution to electoral cybersecurity by crafting a holistic and systematic framework to mitigate a large spectrum of interrelated vulnerabilities, arising from technology, human, political, legal and procedural exposure.

> Attacks target not only the functioning of physical election infrastructure, but also its trove of sensitive data; they are not perpetrated by humans or bots acting separately, but by a complex alliance of human-machine deceptive tactics.

This report shares the comprehensive, interdisciplinary, and non-reductionist approach proposed by IFES. Increasingly, the convergence of cybersecurity with AI, precision social engineering, biometrics, and 5G in what we call "The Internet of Bodies,"[14] leads to a complex socio-technical ecosystem where emerging properties produced by technological convergence will significantly amplify technical, but also human, procedural, data- and infrastructure-related vulnerabilities. From weaponizing internet and electricity shutdowns to corrupting voters' digital identities, this report shows how the threat landscape to election security is extending.

Interestingly, the integration of AI with its potential for automating precision, stealth and personalization in cyber-offense intensifies rather than reduces the impact of human vulnerabilities on electoral cybersecurity. AI-enabled cyberattacks, through impersonation, precision spear-phishing and social engineering, can harness human weaknesses to the point of increasingly equipping external actors with insider and tacit knowledge. The distance between external and insider attacks is shrinking.

Another evolving aspect outlined in the IFES report is the importance of protecting data integrity through an election cycle.[15] Such argument is also efficiently developed by Herpig et al. in a 2018 report[16] that explains how elections are first and foremost data-driven processes. Herpig et al. show how different sets of data, from personal and governmental data, to confidential communication and security information, are the target of cyberattacks and offenses in elections. The heterogeneity of both data-sets and related actors, including political candidates, party campaigns, public and private sectors, produces daunting information security challenges in elections.

Aligning with the arguments advanced by Herpig et al., the IFES report states that "concerns may be raised around privacy of citizen data – including biometric information – especially in countries that are collecting voter data and do not have data protection laws in place, or where data is kept on servers outside the country, raising the risk that such data could be exploited."[17]

[12] IFES, 2018. pg 14.

[13] IFES, 2018. pg 4.

[14] Pauwels E., 2018.

[15] IFES, 2018. pg 5.

[16] Herpig S. et al., 2018.

[17] IFES, 2018. pg 23.

The IFES report goes further by outlining the kind of information security breaches that could be performed by insider malpractice and go undetected by traditional, legacy cybersecurity approaches: "There are a number of countries in which the central election authority is a de facto extension of the government, regardless of the EMB's formal status as an independent commission. [...] This can lead to data security breaches, such as breaches of voter registration data stored in the central election office. If an IT staffer receives an order from a politicized EMB commissioner to copy the entire voter register onto a USB flash drive, he or she may do it without questioning, fearing repercussion."[18]

> AI has the potential to significantly augment sensitive and biometrics data exfiltration capacities while evading detection.

While such information security breach can happen through rudimentary digital malpractices, AI has the potential to significantly augment sensitive and biometrics data exfiltration capacities while evading detection [Such potential is documented in Section 2].

In light of these emerging socio-technical challenges, it will be crucial to develop a systemic, comprehensive threat and vulnerability assessment framework that can support EMBs in working towards preparedness and prevention. In addition to IFES, a few other expert groups[19] have proposed electoral cybersecurity analyses that favor preventive adversarial testing to promote foresight, adaptability and resilience.

In 2018, experts from EU member states, the European Commission and the European Union Agency for Network and Information Security (ENISA) produced a *Compendium on Cyber Security of Election Technology*, a comprehensive compilation of guidelines on electoral cybersecurity to help EMBs learn from EU member states' past threats, as well as preventive and cooperation experiences.[20] The *Compendium* strategically insists on the importance of combining voting software functionality test and adversarial testing. For instance, in a test attack, white-hat teams could harness many offensive techniques to take advantage of cybersecurity, human, political, data and infrastructure-related vulnerabilities. Another set of testing methods consists in red-teaming or tailored election simulation tabletop exercises to test EMB responses and resilience to specific and hybrid forms of cyber exploitation.

Most recent reports provide in-depth frameworks to help EMBs and government actors better prepare in defending against existing cyber-threats to election security. Yet, in the near-future, one additional, strategic approach is needed to understand how the electoral cybersecurity threat landscape is evolving for states in the Global South that are struggling to build and secure capacity in the development and deployment of cyber- and AI technologies. Governments from different nations should share learning and experiences in election security and work collaboratively to establish and maintain international standards regarding best practices surrounding election security. This may include, but is not limited to, voting technologies, software, and security strategies. UN Member States will need to develop a common understanding of how technological convergence impacts election security to be able to design proper oversight in collaboration with strategic actors in the private sector and civil society. States lagging behind in Cyber-AI convergence are the most at risk and the least likely to have any adversarial testing and foresight capacity.

> UN Member States will need to develop a common understanding of how technological convergence impacts election security to be able to design proper oversight in collaboration with strategic actors in the private sector and civil society.

[18] IFES, 2018. pg 23.

[19] IFES, 2018. pg 23-24.; Stanford Policy Center, 2019. "Securing American Elections." https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/stanford_cyber_policy_center-securing_american_elections.pdf

[20] NIS Cooperation Group, 2018.

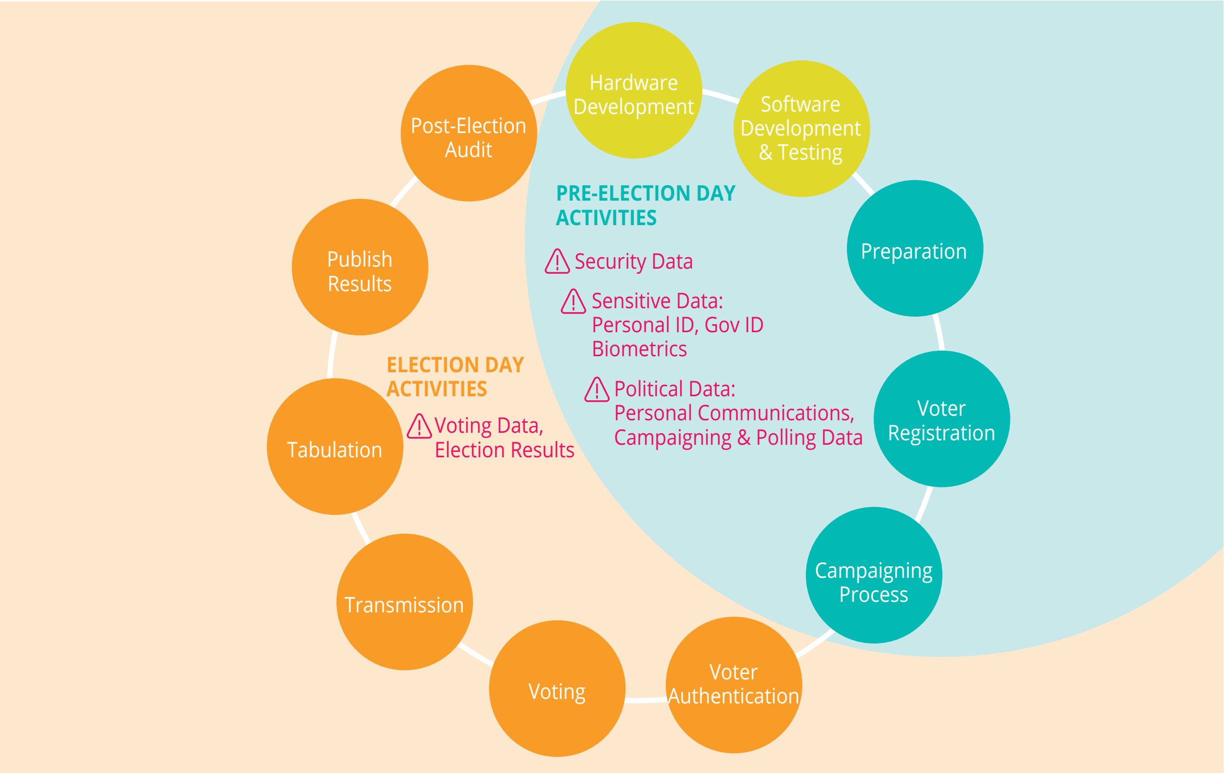
> “Within a network of trust, entrepreneurs, policymakers and EMBs could use red-teaming exercises to anticipate potential vulnerabilities as well as safety and security best practices relevant to the convergence of AI and cyber-technologies.

AI-driven cyberattacks already impact low- and high-income countries without discrimination. As the digital scope of these new forms of hybrid threats is not limited to the West, an increasing number of States might be used as “vulnerable links” in a new virtual geography of conflicts. Keeping pace with rapidly changing cybersecurity threats will become increasingly difficult, regardless of the country. Yet, it will be the most vulnerable countries, the vulnerable links, that will be impacted the most. As AI and cyber capabilities expand in developing countries, so too will the attack surface.

In the context of electoral cybersecurity, nascent efforts around adversarial testing, foresight and anticipatory accountability could take increasingly agile forms and thrive on alliances between States such as North-South, and South-South collaborations. For instance, within a network of trust, entrepreneurs, policymakers and EMBs could use red-teaming exercises to anticipate potential vulnerabilities as well as safety and security best practices relevant to the convergence of AI and cyber-technologies. Red-teaming could also turn into “sparring” exercises where corporate and government actors learn to collaborate and build a trusted space where to test AI and cyber-systems to report fatal anomalies and discuss optimized defence capabilities tailored to electoral context. Such sparring exercises could be crucial for states that are currently vulnerable links in the new geopolitics of converging technologies.

Sections 2 and 3 of this paper will map technical, but also human, procedural, data- and infrastructure-related vulnerabilities in an election cycle.

Figure 1 | The electoral cycle and its different types of sensitive data-sets.


Hardware Development
Software Development & Testing
Preparation
Voter Registration
Campaigning Process
Voter Authentication
Voting
Transmission
Tabulation
Publish Results
Post-Election Audit
PRE-ELECTION DAY ACTIVITIES
⚠ Security Data
⚠ Sensitive Data: Personal ID, Gov ID Biometrics
⚠ Political Data: Personal Communications, Campaigning & Polling Data
ELECTION DAY ACTIVITIES
⚠ Voting Data, Election Results

The Paradigm of Cyber-AI Convergence

The cybersecurity threats landscape will drastically change in the next decade. And we are not prepared for the cyber- and information security challenges triggered by this era of technological convergence.

Hybrid Security Threats and Attack Surface

Converging technologies are becoming complex hybrid systems that are merging and enabling each other, with drastic variations in velocity, scope and system-wide impact.[21] The convergence of cybersecurity with AI, biometrics, 5G and quantum computing will empower new transformative, dual-use techniques to optimize digital assets in cyberspace and will shape future security and normative challenges.

Quantum technologies will redefine security in cyberspace with more powerful techniques for cryptography, data-optimization and complex problem-solving. The deployment of 5G networks will become an enabler for AI edge-processing. 5G will speed up a shift in AI processing from cloud architectures to decentralized processing at the edge device, amplifying what we call the Internet of Things.

The result is the development of complex systems, exhibiting precision and speed, adaptability and efficiency, but also emerging behaviours that are difficult to anticipate, understand, mitigate and control.

In this era of convergence, AI systems provide an "increasing resource of interactive, autonomous, and self-learning agency, able to achieve outcomes that usually require human intelligence to be performed successfully."[22] Similar to what we observe in biology, this combination of autonomy and self-evolution is where we are facing a new form of augmentation, which underpins both, beneficial and malicious uses of AI. Think of swarms of bots and AI malware that learn to cooperate and transfer laterally between hosts, just like opportunistic viruses can learn to infect bodies.

Technological convergence leads to synergies, adding more value and functional capabilities to complex systems but also increasing emerging uncertainties.

Take a moment to consider the AI-Cyber convergence: AI's exceptional capacity for automating anomaly detection will play a powerful role in cyber defence, algorithms being able to detect abnormal and illicit behaviours across large computing networks and able to learn how to patch vulnerabilities against evolving cyber-threats.

> AI malware with behavioural detection capacity can learn to evade cybersecurity techniques, and evolve a different hiding strategy for each targeted acquisition.

[21] Pauwels E., 2019.

[22] King T. et al, 2019. "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions." https://link.springer.com/content/pdf/10.1007%2Fs11948-018-00081-0.pdf

Yet, cybersecurity is also being challenged by a new class of AI malware, whose aim is to manipulate the integrity of sensitive data. Given their ability to learn and adapt, malicious algorithms can decide what type of payload to use (a deeplocker or a ransomware) according to the type of cyberattack, just like bio-organisms will pass a certain genetic payload for the next-generation to survive. AI malware with behavioural detection capacity can learn to evade cybersecurity techniques, and evolve a different hiding strategy for each targeted acquisition.

Future AI-led cyber-threats will also mix AI-driven and human-driven combinations of attacks' strategies, blurring the distinction between machine or human malicious intelligence, which might have legal consequences in attribution.

> Static not self-learning, programmed to only detect known threats, this approach is no longer viable.

AI will likely become resistant to the categorization of threats that remains the basis for our legacy cyber-security approaches.[23] This is because the traditional approach to cyber security relies on being able to define the threat in advance. Static not self-learning, programmed to only detect known threats, this approach is no longer viable. From novel and fast-spreading attacks to insiders gone rogue, from hacked Internet of Things (IoT) devices to compromised supply chains, the AI-Cyber-threat landscape evolves in unpredictable ways and a new approach to cyber defence is urgently required.

Next, we will examine the emerging typology of cyberattacks that could leverage adversarial AI to *manipulate the integrity of datasets*, poison codes, evade detection, harness human vulnerabilities and ultimately impact cyber and *information* security.

- **Autonomous malware:** As well explained by Marcus Fowler, Director of Strategic Threat, at Darktrace, "AI will not only enable malware to move stealthily across businesses without requiring a human's hands on the keyboard, but attackers will also use AI in other malicious ways, including determining their targets, conducting reconnaissance, and scaling their attacks."[24] In the future, AI-enabled malware will replicate through a series of autonomous and intelligent strategies that tailor methods of propagation to the parameters and weaknesses of the infected system. For instance, a worm-style attack, like WannaCry, could adapt[25] its techniques for propagation to its environment, learning from a previous detection event, instead of using a more traditional, known form of network propagation (or "lateral movement") such as the Eternal Blue exploit.[26]

This is what autonomy and modularity will mean in AI adversarial attacks. AI-enabled malware will familiarize itself with its environment before striking, learning from contextual information to select whatever propagation method appears most successful for the target environment. Autonomous AI malware can also harness multiple payloads

> AI-enabled malware will familiarize itself with its environment before striking, learning from contextual information to select whatever propagation method appears most successful for the target environment.

[23] Darktrace, 2019. "Machine Learning in the Age of Cyber AI." https://www.darktrace.com/en/resources/wp-machine-learning.pdf

[24] Forbes, 2019. "141 Cybersecurity Predictions for 2020." https://www.forbes.com/sites/gilpress/2019/12/03/141-cybersecurity-predictions-for-2020/

[25] Darktrace, 2018. "The Next Paradigm Shift AI-Driven Cyber-Attacks." https://www.oixio.ee/sites/default/files/the_next_paradigm_shift_-_ai_driven_cyber_attacks.pdf

[26] EternalBlue is an exploit that allows cyber threat actors to remotely execute arbitrary code and gain access to a network by sending specially crafted packets. It exploits a software vulnerability in Microsoft's Windows operating systems (OS) Server Message Block (SMB) version 1 (SMBv1) protocol, a network file sharing protocol that allows access to files on a remote server. This exploit potentially allows cyber threat actors to compromise the entire network and all devices connected to it. Due to EternalBlue's ability to compromise networks, if one device is infected by malware via EternalBlue, every device connected to the network is at risk.

for disruption – stealing biometrics details and invalidating electoral voting machines with Distributed Denial of Service[27] (DDoS). Semantic analysis and contextual awareness allow software to automate intelligent decision-making about how to evade and how to attack.

- **Breaking cryptographic keys:** In collaboration with engineers from Google, researchers from the University of Toronto created in 2018 an algorithm that could break two well-established codes: the Caesar cipher, which is simple and comparatively vulnerable to cracking, and the Vigenère cipher, which uses secret keys—an extra variable that requires more sophistication to crack. They demonstrated that a certain type of algorithms called CipherGAN is capable of "cracking language data enciphered using shift and Vigenere ciphers to a high degree of fidelity and for vocabularies much larger than previously achieved."[28] This CipherGAN algorithm is capable of moving back and forth between two completely unrelated texts— for instance, two plain texts in cipher code.[29]

- **Intelligent evasion techniques:** Malicious algorithms will be able to tailor their offensive strategies to the environment they infect. AI-enabled malware could deploy evolving methods to evade cybersecurity detection, from adapting defensive behaviours to erasing itself when it suspects it is being analysed. By leveraging contextualisation, AI malware will target specific vulnerable entry-point, or imitate trusted elements of the system. This will allow AI cyber-attacks to evade detection and maximize the damage they cause.

> "This ability to evade detection will mean that AI malware is able to compromise more devices than ever before.

Sophisticated threat actors can often maintain a long-term presence in their target environments for months at a time, without being detected. They move slowly and with caution, to evade traditional security controls and are often targeted to specific individuals and organizations. AI will also be able to learn the dominant communication channels and the best ports and protocols to use to move around a system, discreetly blending in with routine activity. This ability to evade detection will mean that AI malware is able to compromise more devices than ever before.

- **Low and slow data exfiltration:** AI-led cyber-attack will excel at leveraging low and slow data exfiltration. Cybersecurity experts have noticed cases where "data is being exfiltrated from a medical technology company at such a slow pace, and in such small packages, that it avoids triggering the data volume threshold in legacy security tools."[30] AI malware with their capacity to analyse and scope context will be powerful tools for low-and-slow data exfiltration as well as for data manipulation.

- **Data-manipulation and poisoning at the source:** AI malware could be used to automate data-manipulation with the intent to falsify, erase or steal intelligence within large curation of data, for instance genomics or medical history data. It could also specifically target serious cybersecurity weaknesses in optical scanning equipment and electronic machine networks.

[27] A Distributed Denial of Service is a cyber-attack in which the perpetrator seeks to make a machine or network resource unavailable to its intended users by temporarily or indefinitely disrupting services of a host connected to the Internet. Denial of service is typically accomplished by flooding the targeted machine or resource with superfluous requests in an attempt to overload systems and prevent some or all legitimate requests from being fulfilled. In a distributed denial-of-service attack, the incoming traffic flooding the victim originates from many different sources. This effectively makes it impossible to stop the attack simply by blocking a single source. DDoS attacks have been used as a weapon of choice of hacktivists, profit-motivated cybercriminals, and nation states. See: https://us-cert.cisa.gov/ncas/tips/ST04-015

[28] Gomez A. et al., 2018. "Unsupervised Cipher Cracking Using Discrete GANs." https://arxiv.org/pdf/1801.04883.pdf

[29] Gomez A. et al., 2018.

[30] Darktrace, 2018. pg 4.

In April 2019, researchers at the Ben-Gurion University Cyber Security Research Center in Israel used algorithms to automate the manipulation of bio-data – by removing or adding realistic, malignant-seeming tumors to CT scans before doctors could examine them.[31] This experiment shows that it is possible to use machine learning to train algorithms to quickly adjust and scale fake tumors to conform to a patient's unique anatomy and biology.[32] The entire attack can be fully automated so that once the malware is launched into a hospital network, it will operate on its own, to find and alter CT scans, even searching for a specific patient's name.

> AI malware could be used to automate data-manipulation with the intent to falsify, erase or steal intelligence within large curation of data.

- **Arbitrary, autonomous code injection and execution:** "Code injection" is a generic term used to describe an attack that exploits poorly written code in a way that allows attackers to execute their own arbitrary code. In essence, malicious code injection techniques allow attackers to manipulate the integrity of data and the logic and functioning of algorithmic models.

  A Structured Query Language (SQL) injection occurs when an attacker inserts malicious code into a server that uses SQL and forces the server to reveal information it normally would not. An attacker could carry out an SQL injection simply by submitting malicious code into a vulnerable website search box.[33] During Ohio's state elections in November 2019, Russian-linked hackers attempted SQL injection seeking to insert malicious code into Ohio Secretary of State Frank LaRose's official website.[34]

- **Precision Biometrics Attacks:** In 2018, IBM detected an AI malware that can hide a cyberthreat, such as WannaCry, in a video conference application, and launch only when it identifies the **face of the target**.[35] This makes the malicious code hard to detect and almost impossible to reversely engineer.

More common, scalable, **personalized spear-phishing** leverages AI to tailor phishing emails to specific users in order to increase chances of infecting the system. AI malware will watch, track and evaluate individuals' emotions, language and behaviour, impersonating trusted contacts within professional and personal social networks. Tailored communication generated by AI malware will therefore be almost impossible to distinguish from human peers' communications. An AI system that has been taught to study the behaviour of social network users and implement finely-targeted, personalized spear-phishing attacks on them, was able to perform more than 6 times as efficiently as humans and with a higher conversion rate.[36]

In March 2019, cyber criminals used machine learning **voice spoofing** to commit a cybercrime by reproducing the voice of a CEO, demanding a fake transfer of about $240,000. The company Lyrebird developed an AI-enabled voice imitation algorithm that it says "can not only mimic the

[31] Zetter K., 2019. "Hospital viruses: Fake cancerous nodes in CT scans, created by malware, trick radiologists." The Washington Post. https://www.washingtonpost.com/technology/2019/04/03/hospital-viruses-fake-cancerous-nodes-ct-scans-created-by-malware-trick-radiologists/

[32] Mirsky Y., 2019. "CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning." https://arxiv.org/abs/1901.03597

[33] Cisco, "Common Cyber Attacks." https://www.cisco.com/c/en/us/products/security/common-cyberattacks.html

[34] The Associated Press, 2019. "Russian-owned company attempted Ohio election hack". https://apnews.com/6518b9a986f640c4899a979bbc48390b

[35] Osborne C., 2018. "Deeplocker: When malware turns artificial intelligence into a weapon." ZDNet. https://www.zdnet.com/article/deeplocker-when-malware-turns-artificial-intelligence-into-a-weapon/.; Kirat D. et al, 2018. "DeepLocker: Concealing Targeted Attacks with AI Locksmithing." Black Hat USA 2018.

[36] Seymour J. and Tully P. "Weaponizing Data Science for Social Engineering." https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter.pdf

speech of a real person but shift its emotional cadence – and do all this with just a tiny snippet of real-world audio."[37] Increasingly, machine-learning methods can be used to generate automated argumentative text or to impersonate someone's behaviour in a video or someone's voice in an audio file. **DeepFakes** could be used in incredibly convincing spear phishing attacks that users would have a very hard time to identify as false. Already, impersonation attacks are on the rise: about two-thirds of businesses saw an increase in forgeries in the last 12 months.[38]

> “DeepFakes could be used in incredibly convincing spear phishing attacks that users would have a very hard time to identify as false.

- **Social and Emotional Engineering:** Cybercriminals use AI to automate new forms of social engineering. The combination of psychometrics manipulation tools with personal datasets can help craft convincing emotion-targeting campaigns that can hardly be recognized as malicious. Even the most experienced users might fall for such personalized attacks.

  By allowing the analysis of individual communication, perception and emotion to be automated, AI systems can increase anonymity and psychological distance in cyber operations. In the near-future, automated cyber operations, led by machine-learning, will therefore be more effective, finely targeted, difficult to attribute, and likely to exploit evolving vulnerabilities in AI and human systems.[39] One pervasive security threat will be new forms of hybrid influencing made possible by the automation of social-engineering attacks. Many major cybersecurity incidents rely on social engineering where malicious actors target the social and psychological vulnerabilities of humans within chains of command. The goal is to manipulate command and control organizations to compromise their own safety and security.

> “One pervasive security threat will be new forms of hybrid influencing made possible by the automation of social-engineering attacks.

- **Distributed Denial-of-Service Attack and Defacement:** Though DDoS attacks have been around since the inception of the Internet, AI botnets — a network of autonomous agents — are emerging as the new go-to DDoS technique. Hackers have even been willing to spend about $150 per week to rent a botnet to launch DDoS attacks to takedown online services by flooding it with so much data traffic that it is unable to maintain functionality.

  "A DDoS attack against a state elections website could take it offline anywhere from seconds to days. While a couple of seconds may not affect voter's abilities to reach the polls, if the elections-related information is offline for longer periods of time, it could prevent individuals from knowing their specific polling location, thereby reducing voter turnout."[40] When it takes place during vote tabulation and publication of results, a DDoS attack harms the credibility of the electoral process and undermines citizens' trust in the process, in EMB's capacities and in the overall election results.

  Defacement occurs when malicious actors leverage cyber-vulnerabilities to deface websites and can even include changing or blocking key information that voters need to find their polling station. Blocking voters from accessing election-related information connects defacement to DDoS attacks. Defacement can also be used in coordination with disinformation campaigns by defacing

[37] Lomas N., 2017. "Lyrebird is a voice mimic for the fake news era." TechCrunch. https://techcrunch.com/2017/04/25/lyrebird-isa-voice-mimic-for-the-fake-news-era/

[38] Darktrace, 2018. p 5.

[39] Mayer M., 2018. "Artificial Intelligence and Cyber Power from a Strategic Perspective." IFS Insights; April. https://brage.bibsys.no/xmlui/bitstream/handle/11250/2497514/IFS%20Insights_4_2018_Mayer.pdf

[40] FireEye, 2018. "Attacking the Ballot Box: Threats to Election Systems.". https://media.scmagazine.com/documents/343/election_systems_report_85540.pdf

political party and/or candidate websites to create unrest.

> The convergence of AI and cybersecurity is giving rise to disruptive, adversarial techniques that can corrupt information security and ultimately erode evidence, trust and cohesion within political processes and societies.

The convergence of AI and cybersecurity is giving rise to disruptive, adversarial techniques that can corrupt information security and ultimately erode evidence, trust and cohesion within political processes and societies. Since 2004, at least 27 European and North American countries have allegedly been victims of cyberattacks, disinformation, and financial influence campaigns crafted for destabilization.[41]

In this new convergence merging cybersecurity, tech and politics, the next winning move will be to manipulate information infrastructure and its secrets. We increasingly face geopolitical conflicts in which psychological and algorithmic manipulation are becoming endemic in cyberspace, an ecosystem of nearly four billion minds.[42] Yet, the impacts felt are real in the physical world, from influencing elections, to destabilizing economies and political regimes.

This paper focuses on how the convergence of AI with cyber- and information offensive operations impacts the security of elections. Most has been said and written about the lack of resilience of the U.S. electoral system towards new cyber-threats. Yet, few contemporary analyses capture how AI convergence will impact the security of elections in emerging economies as well as in low-income and fragile states – those states that struggle to compete and build tech and innovation capacity. There is a crucial need to map and analyse which countries, electoral processes and populations will face pervasive hybrid threats due to the convergence of dual-use technologies. The next section offers such a preliminary conceptual mapping of emerging AI-cybersecurity threats classified by separate categories of attack-vectors and data-targets, drawing from the analysis below.

Section 3 provides a general matrix of an election cycle and infrastructure, analysing how the convergence of AI and cybersecurity will impact the landscape of election threats and vulnerabilities, by identifying data-targets, human vulnerabilities, attack-vectors tailored to election processes and implications.

> There is a crucial need to map and analyse which countries, electoral processes and populations will face pervasive hybrid threats due to the convergence of dual-use technologies.

[41] Dorell O., 2017. "Alleged Russian political meddling documented in 27 countries since 2004." USA Today. https://www.usatoday.com/story/news/world/2017/09/07/alleged-russian-political-meddling-documented-27-countriessince-2004/619056001/

[42] Villasenor J., 2018. "Artificial Intelligence and the future of geopolitics." Brookings Institute. https://www.brookings.edu/blog/techtank/2018/11/14/artificial-intelligence-and-the-future-of-geopolitics/

3

Conceptual Matrix of Cyber-AI Threats & Vulnerabilities in an Election Cycle

States increasingly perceive cyberspace not only as a source of innovation and supremacy, but also as a source of potential threats, both from other states and non-state actors. And the power to conduct fair and free elections is at the heart of this battle of influence being waged for the control of populations' trust. The below preliminary mapping of emerging AI-enabled cybersecurity threats (cf. Figure 2) is a conceptual exercise with its inherent limits, but still a powerful tool to assess the preparedness and resilience of nations, which are integrating digital technologies to manage political campaigns and electoral processes.

Election campaigns and parts and processes of the election cycle are becoming increasingly digitized, and therefore vulnerable to AI-enabled cyberattacks. The convergence of cybersecurity with AI and other emerging technologies augments the potential for deception and subversive attacks that can interfere with populations' perceptions and the conduct of free and fair elections.

> The power to conduct fair and free elections is at the heart of a battle of influence being waged for the control of populations' trust.

Digitization and technological convergence extend and amplify the attack surface, turning an array of electoral datasets into targets for interference. Elections also take place within a web of complex socio-technical systems where human vulnerabilities increasingly allow for algorithmic manipulation, social engineering, political deception and cyber-AI attacks. This report therefore approaches the election cycle and its infrastructure, first, as a set of complex socio-technical systems and, second, as a set of data-driven processes. This approach allows us to anticipate, reframe and better understand the emerging types of vulnerabilities that AI will increasingly be able to target within the data-infrastructures and data-optimization processes related to the conduct of elections.

> Election campaigns and parts and processes of the election cycle are becoming increasingly digitized, and therefore vulnerable to AI-enabled cyberattacks.

Human Vulnerabilities, Intelligent Malware & Precision Biometrics-Targets

Cybersecurity experts are concerned that emerging technologies like AI and autonomous data-capture devices within the Internet of Things (IoT) are helping cybercriminals attack election systems faster than government and electoral officials can keep up. Securing elections is increasingly about human vulnerabilities, intelligent malware and information security.

AI is increasingly used to map users' online behaviors, relationships, political and sexual orientations, health and emotional states. In the near future, facial recognition, biosensors and algorithms will capture and analyze an ever more refined record of humans' biometrics. AI will watch, track, and evaluate individuals, from the predictive power of one algorithm to the next. In this "Internet of Bodies and Devices," cybercriminals can harness personal data of individuals for intimidation, manipulation, ransomware,

> Elections take place within a web of complex socio-technical systems where human vulnerabilities increasingly allow for algorithmic manipulation, social engineering, political deception and cyber-AI attacks.

and precision spear-phishing, with the aim to intercept credentials, insider-knowledge and sensitive security files.

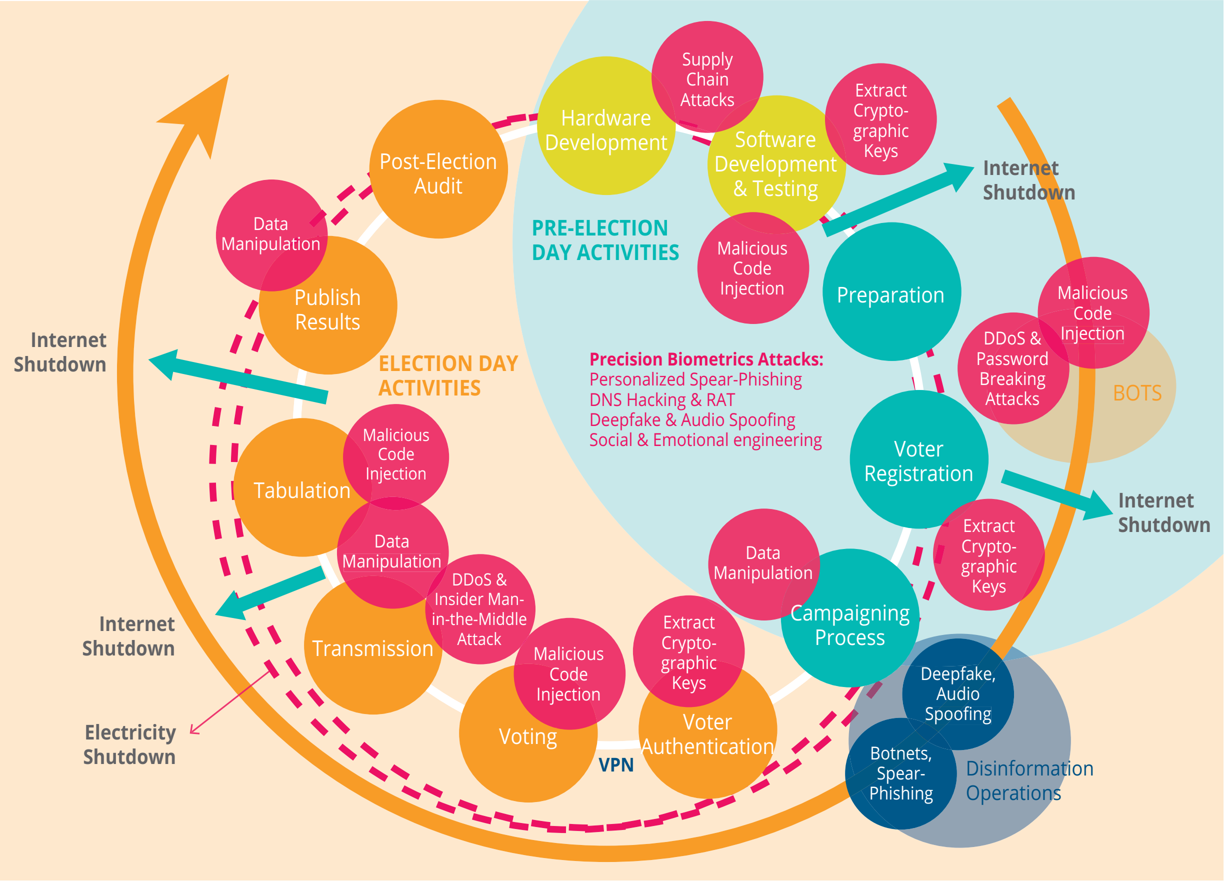
At the end of 2019, forecasting growing 2020 trends, Cisco[43] unveils that most successful cyber-threats start with hacking human vulnerabilities within chain of command. For instance, performing Domain Name System (DNS) hijacking, malicious actors can silently redirect unsuspecting visitors from legitimate websites – they are known to visit – towards malicious ones, potentially to install intelligent malware, intercept confidential data and credentials.

Cisco also reveals how malicious actors use campaigns targeting government entities and public services to install intelligent remote access trojans (RAT) that can evade detection with highly damaging consequences, from exfiltrating credentials and sensitive information, deleting and installing files, to taking command of a computing system. Increasingly, exfiltrated data are encrypted by hackers to circumvent monitoring tools.

Precision spear-phishing emails rely on more and more sophisticated tactics such as impersonating the style, tone and attitude of professional contacts to infect chain of command. In 2019, we saw the emergence of precise forms of social engineering with the *Emotet* botnet able to hijack email threads

> Securing elections is increasingly about human vulnerabilities, intelligent malware and information security.

Figure 2 | Vulnerabilities that may exist in the full electoral cycle.



[43] Cisco, 2019. "Threats of the Year." https://www.cisco.com/c/dam/en/us/products/collateral/security/2019-threats-of-the-year-cybersecurity-series-dec-2019.pdf

by injecting responses into old or ongoing professional conversations.[44]

The use of deep-learning could drastically enhance these intrusive tactics with reproducing someone's biometrics for authentication or generating audio and video forgeries to impersonate trusted contacts. In 2018, at the Black Hat Convention in Las Vegas, a malware learned to wipe out the computer of a target, just by recognizing his or her facial and biometrics features.[45]

> Increasingly, exfiltrated data are encrypted by hackers to circumvent monitoring tools.

All of the above techniques could be used to target human vulnerabilities and take command of digital operations through sophisticated malware during political campaign and the full electoral cycle. Vendors and other private sector actors – working on hardware and software development for elections – could be targeted through DNS hijacking, RAT and precision spear-phishing. Same tactics could be used with IT specialists and other staff within EMBs and government entities. Campaign staffers are prime targets for precision spear-phishing as, working under pressure, they might be tempted to quickly open links and documents sent via emails and they may lack up-to-date cybersecurity training.

During political campaigning prior to an election, spear phishing has become a common tool for attackers to gain access to sensitive data on election and government officials, political parties, candidates, and voters. In April 2017, the hacker group APT28 registered domain names similar to the name of the Macron campaign team. The attackers then successfully fooled staff and were able to obtain login credentials which enabled them to access information that was later leaked on a website known as the MacronLeaks.[46]

Ukrainian elections have also been recently targeted for spear-phishing attacks with malicious actors using virus-infected greeting cards, shopping invitations, offers for software updates, and other malicious phishing material.[47] In January 2019, the head of Ukraine Cyber Police reported that, ten weeks before the presidential election, hackers were acquiring personal information of civil servants and election officials, paying in cryptocurrency on the dark web.[48]

Interestingly, attackers are beginning to use current events as phishing lures. For example, in December 2018, a document titled "UDS 2019 Current Agenda.doc," which, when opened, dropped a particular malware, was sent by the hacker group SNAKEMACKEREL in anticipation of the Underwater Defense & Security 2019 event.[49]

Another source of vulnerabilities exists in socio-technical systems surrounding political and electoral staff and their organizations. Large amounts of desktop computers, network-connected printers, cameras, laptops, phones and personal sensing devices (such as Fitbit) sit unsecured, with the latest security patches not installed, just waiting for someone with malicious intentions to exfiltrate sensitive data or infect computing networks with autonomous malware.

> The use of deep-learning could drastically enhance intrusive tactics with reproducing someone's biometrics for authentication or generating audio and video forgeries to impersonate trusted contacts.

[44] MalwareBytes, 2019. "Emotet is back: botnet springs back to life with new spam campaign." https://blog.malwarebytes.com/botnets/2019/09/emotet-is-back-botnet-springs-back-to-life-with-new-spam-campaign/

[45] Kirat D. et al, 2018.

[46] Risk and Resilience Team, 2017. "Hotspot Analysis: Cyber and Information Warfare in Elections in Europe." Center for Security Studies, ETH Zurich. https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/Cyber-Reports-2017-08.pdf

[47] Polityuk P., 2019. "Ukraine says it sees surge in cyber attacks targeting election," Reuters. https://www.reuters.com/article/us-ukraine-cyber-exclusive/exclusive-ukraine-says-it-sees-surge-in-cyber-attacks-targeting-election-idUSKCN1PJ1KX

[48] Polityuk P., 2019.

[49] Brady M. And Bucholz K., 2019. "SNAKEMACKEREL Delivers SedUploader Malware." Accenture. https://www.accenture.com/us-en/blogs/blogs-snakemackerel-delivers-seduploader-malware

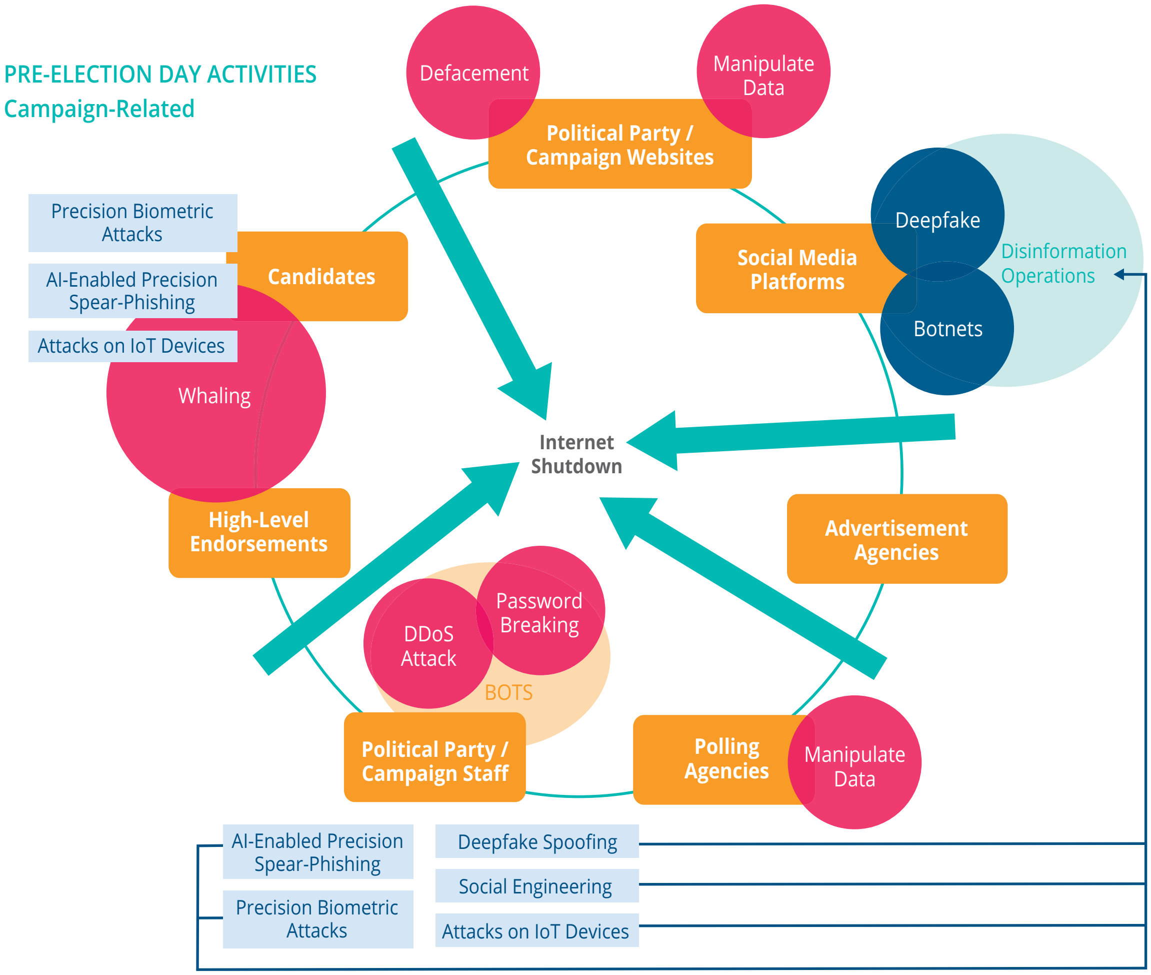
Personal data produced by IoT devices will likely surface on the dark web and be used for automated attacks like password-breaking. Botnets are responsible for nearly 300,000 malicious login attempts every hour.[50] These bots use leaked usernames and passwords to make repeated login attempts into the accounts of campaign staffers and election officials, trying one password combination after another until one works. Once the cybercriminals have invaded the campaign account, they can leak corrosive material and wage information operations. Automated cyberattacks against political campaigns are on the rise and, in convergence with 5G, will happen at speedlight, invading extensive networks of personal devices.

> “Another source of vulnerabilities exists in socio-technical systems surrounding political and electoral staff and their organizations.

Figure 3 shows entry-points, vectors, and targets corresponding to political campaigning activities. Most of the above-described methods of AI-Cyber intrusion could give cybercriminals command over political party and campaign websites to manipulate or erase content, what is called *defacement*.

The below table shows how certain types of cyberattacks harness vectors and human vulnerabilities to target specific data sets through increasingly extensive networks of digital devices. Datasets that

Figure 3 | Entry-points, vectors, and targets corresponding to pre-election day political campaigning activities.



[50] Akamai, 2018. “State of the Internet.” https://www.akamai.com/us/en/multimedia/documents/state-of-the-internet/soti-2018-credential-stuffing-attacks-report.pdf

Table 1 The landscape of cyber-AI threats before and during elections

LANDSCAPE OF CYBER-AI THREATS	ATTACK VECTORS	ACTORS
Supply Chain Attack	• Malicious Implants, Malicious Code Injection, RAT, Extract Cryptographic Keys • Precision Biometrics Attacks, Personalized Spear-Phishing, DNS Hacking, Deepfake, Audio Spoofing, Social Engineering	Insider Threat and External Actors
Attack on Software and E-Voting Equipment	• Malicious Implants, Malicious Code Injection, RAT, Extract Cryptographic Keys • Precision Biometrics Attacks, Personalized Spear-Phishing, DNS Hacking, Deepfake, Audio Spoofing, Social Engineering	Insider Threat and External Actors
Attack on Voter Registration Databases and Websites	• DDoS and Password Breaking attacks, DNS Hacking, RAT, Malicious Code Injection, Extract Cryptographic Keys • Data Manipulation and Poisoning • Breaking of Digital Signatures • Precision Biometrics Attacks, Personalized Spear-Phishing	Insider Threat and External Actors
Attack on Political Campaign	• DDoS, Defacement, Password Breaking attacks, DNS Hacking, RAT, Malicious Code Injection, Extract Cryptographic Keys • Data Manipulation and Poisoning • Precision Biometrics Attacks, Personalized Spear-Phishing, Whaling, Deepfake, Audio Spoofing, Social Engineering	Insider Threat and External Actors
Attack on Election Day Activities (voting, ballot transmission, tabulation, publication)	• DDoS and Password Breaking Attacks, DNS Hacking, RAT, Malicious Code Injection, Extract Cryptographic Keys • Data Manipulation and Poisoning • Breaking of Digital Signatures	Insider Threat and External Actors

> “Automated cyberattacks against political campaigns are on the rise and, in convergence with 5G, will happen at speedlight, invading extensive networks of personal devices.

HUMAN TARGET	DATA TARGET	ELECTORAL PHASE TARGETED	AI AUGMENTATION
Engineers, Employees from Vendors	Security Data	Hardware Development	Precision, Personalization, Automation, Intelligent Evasion, Data-exfiltration
Engineers, Contractors, Gov IT sSpecialists, Electoral (cybersecurity) Officials, EMBs	Security Data: Cryptographic Keys, Digital Signatures, Source Code Libraries	Software Development	Precision, Personalization, Automation, Intelligent Evasion, Data-exfiltration
• Voters • Gov Officials • Gov IT Specialists, Electoral (cybersecurity) Officials, EMBs	Sensitive Data: • Personal ID, Gov ID • Biometrics	Voter Registration	Precision, Personalization, Automation, Intelligent Evasion, Data-exfiltration
• Political Candidates • Campaign Staff (humans & their IoT devices)	• Personal Communication • Political & Polling Data • Security Data • Personal ID	Political Campaign	Precision, Personalization, Automation, Intelligent Evasion, Data-exfiltration
Gov IT Specialists, Electoral (cybersecurity) Officials, EMBs	• Voting Data & Election Results	Vote Casting, Counting & Publication	Automation, Intelligent Evasion, Data-exfiltration

> “Certain types of cyberattacks harness vectors and human vulnerabilities to target specific data sets through increasingly extensive networks of digital devices.

can become sensitive targets in hacking operations extend from security data about digital assets, hardware and software and equipment of the election process; government-issued identification, biometrics and personal data of voters and political or election staff; as well as personal communication and campaigning information. In a nutshell, the digital footprint of a full election-cycle involves large amounts of heterogeneous data, contingent to significant potential for social and emotional engineering.

The result of this new forms of social engineering, intelligent malware, precision spear-phishing and biometrics attacks could impact pre-election and election day operations with the goal to manipulate election data integrity, compromise the electronic process of elections, and leak sensitive information.

Next, we will examine the cyber-threats and vulnerabilities that exist at different steps of pre-election and election day activities as presented in Figure 4 and Figure 5.

Election Hardware and Supply Chain Attacks

Without proper vetting and oversight during the technology development and testing phase, various attacks on the supply chain could compromise the integrity and security of the entire electoral process from the beginning, represented as the pink circular arrow in Figures 3 and 5. Election vendors are entities that "design, manufacture, integrate, and support voting machines and the associated technological infrastructure."[51] Often managed by corporate vendors, biometrics ID, for instance, are becoming an integral part of the electoral infrastructure and a first entry-point for adversarial attacks.

> "Without proper vetting and oversight during the technology development and testing phase, various attacks on the supply chain could compromise the integrity and security of the entire electoral process.

For example, an insider threat, such as an electrical engineer, that works for a vendor that provides hardware components for election technologies could insert a malicious chip into the motherboard — and/or routers and computer operating system software — which could allow both insider threats and external actors to access and manipulate voting data throughout the electoral process by leveraging the backdoor installed in the hardware.

State-sponsored actors, such as those within the intelligence communities, could intercept various hardware components *en route* from vendors in order to implant malicious chips or code that would allow them to monitor communications and siphon sensitive data from those systems, such as the personal data connected to the voter roll, which could be used downstream to inform social engineering campaigns.

Another example of how election technology vendors are highly vulnerable to supply-chain attacks is the possibility of spear-phishing attacks targeting these vendors in order to gain access to customer networks, which could allow external actors to communicate with candidates, government and electoral officials.[52]

The opacity of supply chains, whether elections technologies are obtained from private vendors or developed in-house by electoral authorities, exacerbates the potential for internal and external actors to successfully commit supply chain attacks, in particular when there is limited oversight of cybersecurity practices or responses to security vulnerabilities.

Software Development and Electronic Voting Machines

Several teams of researchers in the public and private sectors have warned that software pro-

[51] Lorin H. et al., 2017. "The Business of Voting: Market Structure and Innovation in the Election Technology Industry." University of Pennsylvania, Wharton School. https://publicpolicy.wharton.upenn.edu/business-of-voting/.

[52] Norden L., Deluzio C., and Ramachandran G., 2019. "A Framework for Election Vendor Oversight." Brennan Center for Justice, New York University School of Law. https://www.brennancenter.org/sites/default/files/2019-11/2019_10_ElectionVendors.pdf

grammed to power electronic voting machines could be poisoned with increasingly sophisticated malware.

The most straightforward scenario would rely on an insider threat actor to directly infect computer servers used by a vendor or the election and government authorities in charge of designing the election software code. Such an attack is also possible remotely by cybercriminals who would use precision spear-phishing, DNS hijacking or RAT to launch a malware able to infiltrate and corrupt the targeted computer network where the election software resides.

Further down the pre-election day cycle, an insider threat could infect with malware the removable media (flash drives and memory cards) used to transfer election software onto voting machines. By getting physical access to voting machines, when they sit in storage before deployment, malicious actors could compromise them using infected memory cards. Such a mock attack strategy by a security team at Stanford took less than one minute to infect Diebold Accuvote TS machines.[53] As very few voting machines have strong authentication or integrity checks, these external devices could execute an arbitrary code, poisoning voters' selection or tampering with vote counting.

Cybersecurity approaches are progressively being developed to better detect illicit intrusion, run software integrity checks, hunt malicious codes, and patch, in real-time, known vulnerabilities. Yet, the problem is amplified with unknown threats.

The current era of converging, unpredictable, fast-moving and self-learning AI-enabled cyberattacks will outpace traditional cybersecurity techniques programmed to primarily detect pre-modelled threats. Blending into the cyber-ecosystem, selecting customized payloads, only crossing the perimeter boundary once and hiding data exfiltration patterns, new AI-cyber-threats could become extremely difficult to detect when targeting election software. Yet, they could still act as powerful autonomous agents to poison the functioning of such software or learn how to manipulate the integrity of voters' selection and personal data.

With automated malware, data exfiltration scenarios will become even better at evading detection. AI malware with a strong presence in the computer network of a vendor or an election management body does not have to conduct the exfiltration of sensitive electoral data (security data or voters' information) over the course of 24 hours – but could spread the exfiltration over 24 days.

> “Blending into the cyber-ecosystem, selecting customized payloads, only crossing the perimeter boundary once and hiding data exfiltration patterns, new AI-cyber-threats could become extremely difficult to detect when targeting election software.

Another attack vector that AI malware could leverage is impacting cryptographic keys in electronic voting and counting. Cryptographic keys are used to perform tasks such as encrypting votes and the digital ballot box, ensuring votes and software are unmodified, verifying the identity of a voter before he or she casts a ballot, and assisting in tallying the results of an election. In cutting-edge AI research[54] – using generative adversarial neural networks (GAN) – teams at Google are now relying on neural networks to dynamically discover new forms of encryption and decryption to protect a communication channel from adversaries trying to break the security schemes.[55] While such research is promising to reinforce encryption methods, the question remains if it could also be used by hackers for breaking traditional, weak cryptographic keys.

[53] Feldman J., Halderman J., and Felten E., 2017. “Security Analysis of the Diebold AccuVote-TS Voting Machine.” https://www.usenix.org/legacy/event/evt07/tech/full_papers/feldman/feldman_html/index.html

[54] Abadi M. and Andersen D., 2016. “Learning to Protect Communications with Adversarial Neural Cryptography.” arXiv. https://arxiv.org/pdf/1610.06918.pdf

[55] Abadi M. and Andersen D., 2016.

Voter Registration Websites and Databases

Voter registration is essentially a complex data-driven process that aims at curating voter registration lists, by updating and maintaining the information of those eligible to vote, while removing illegitimate or deceased voters. EMBs are then responsible for delivering precinct-by-precinct registration lists to polling stations where in-person voting occurs.

Such complex registration effort relies on two data-collection processes that present inherent vulnerabilities: online registration portals and voter information databases accessible via the Internet in order to allow voters to check their voter status. As an array of countries – such as Brazil and India – have adopted online voter registration processes, the databases and registration websites have become vulnerable to cyberattacks, in particular attacks that aim at manipulating the integrity of voter registration data.

Malicious actors who infiltrate registration databases could delete, falsify or corrupt information about voters with significant implications that range from preventing them from registering before the deadline, deterring them from voting, forcing them to use provisional ballots, or changing their polling locations.

Cyber-intrusions into voter registration databases are a real risk, amplified by three growing threats: continuous cyber-theft of personal data, precision spear-phishing targeting electoral management bodies, and registration websites' vulnerabilities.

In the last decade, millions of voters across the globe have had their personal information leaked publicly or sold in the underground economy and dark web. In July 2016, the seller "DataDirect" auctioned access to a database that purportedly contains registration records for voters in all 50 US states.[56] A listing for the database appeared on a dark web marketplace called The Real Deal, a popular site many cyber criminals use for buying and selling everything from illegal drugs to zero-day software exploits.

In early 2019, a 20-year old amateur German hacker accessed and released personal information – photos, phone numbers, credit card numbers – on Twitter of hundreds of German politicians, including Chancellor Angela Merkel.[57]

> “Cyber-intrusions into voter registration databases are a real risk, amplified by three growing threats: continuous cyber-theft of personal data, precision spear-phishing targeting electoral management bodies, and registration websites' vulnerabilities.

In 2018, the Indian government biometrics database, Aadhaar, was the target of multiple cyberattacks that potentially compromised the ID profiles of large swaths of the 1.1 billion registered citizens. The Chandigarh based Tribune newspaper reported that cybercriminals were monetizing access to the Aadhaar database at a rate of 500 rupees for 10 minutes.[58] Elsewhere in the world in 2018, cybertheft of personal data impacted about 150 million users of the MyFitnessPal application, and around 50 million Facebook users.[59]

Using leaked personal information, malicious actors can gain access to voter registration databases to manipulate or erase existing profiles. As we mentioned above, precision spear-phishing is another potential avenue for cybercriminals to target and steal credentials from civil servants and electoral staff in charge of voter registration databases. Such tactics are part of the adversarial toolkit used by the

[56] Szoldra P., 2016. "A hacker is selling a database of all US voters for $7,800 on the dark web." Business Insider. https://www.businessinsider.fr/us/hacker-voter-registration-database-2016-7

[57] Chappell B. 2019. "Police Say Hacking Suspect, 20, Confessed to Posting German Leaders' Private Data." NPR; 8 January. https://www.npr.org/2019/01/08/683272309/hacking-suspect-20-confesses-to-posting-private-data-of-hundreds-of-germanleade

[58] Khaira R. 2018. "Rs 500, 10 minutes, and you have access to billion Aadhaar details." The Tribune; 5 January. https://www.tribuneindia.com/news/nation/rs-500-10-minutes-and-you-have-access-to-billion-aadhaar-details/523361.html

[59] Axel. 2018. "Enough Is Enough: 2018 Has Seen 600 Too Many Data Breaches." Medium; 24 July. https://medium.com/@AxelUnlimited/enough-is-enough-2018-has-seen-600-too-many-data-breaches-9e3e5cd8ff78

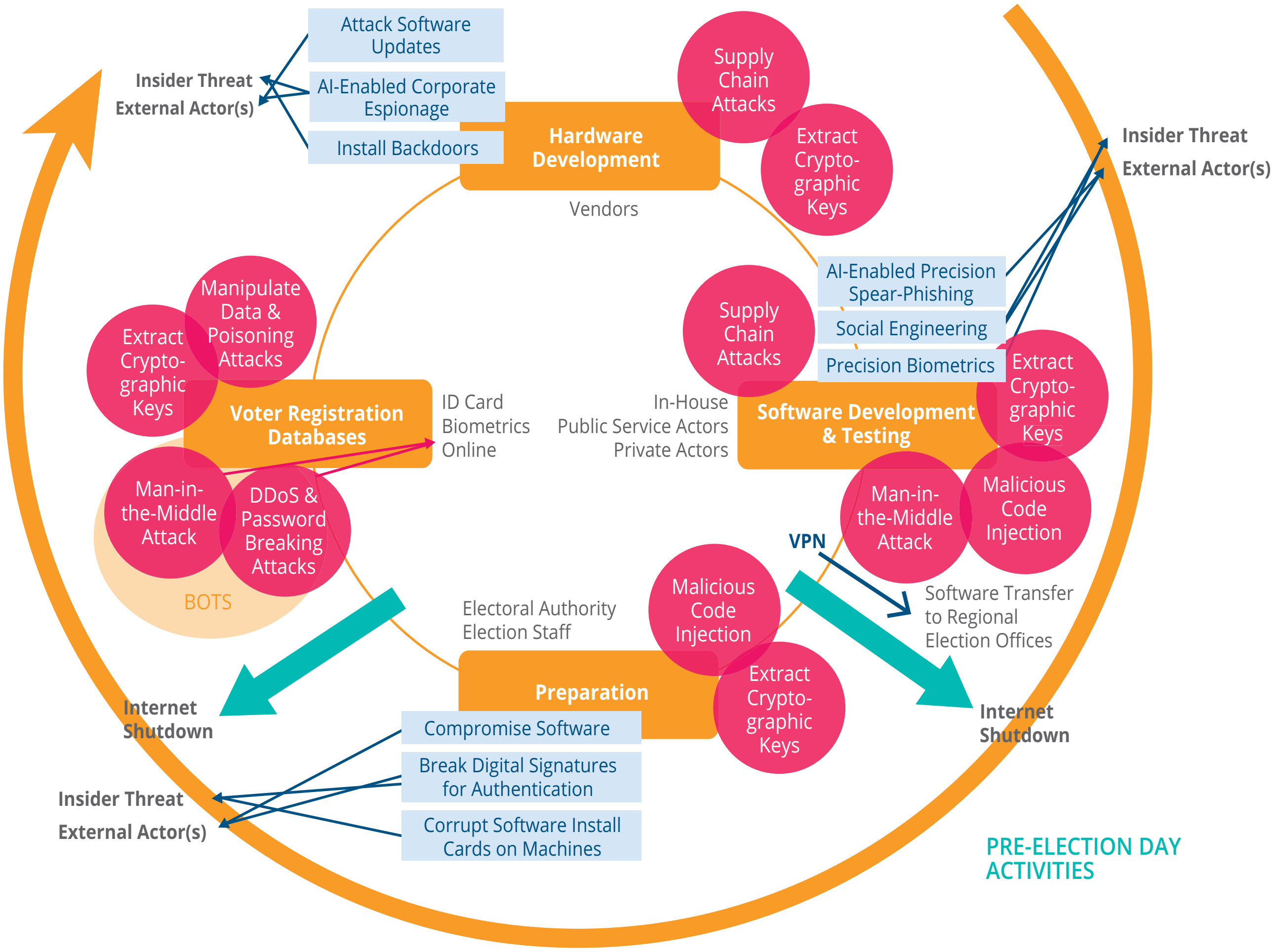
> Using leaked personal information, malicious actors can gain access to voter registration databases to manipulate or erase existing profiles.

GRU during the U.S. 2016 presidential election. According to the FBI, in November 2016, the GRU targeted, with precision spear-phishing emails, over 120 accounts used by county-level officials and successfully penetrated registration databases in at least two Florida counties.[60]

Another attack vector is for malicious actors to use distributed denial-of-service (DDoS) to disable voter registration websites and prevent voters to update their information on time. Through injection of malicious code – cracking what is called SQL injection vulnerability – attackers would also be able to execute arbitrary code execution allowing them to extract, corrupt, erase or add voter records. In June 2016, the GRU used SQL injection vulnerability to compromise the computer network of the Illinois State Board of Elections (SBoE), accessing information on millions of registered voters and extracting data related to thousands of voters.[61]

We are essentially facing, at this stage, a context of pervasive information insecurity with corrosive implications for voter data integrity. Figure 4 shows entry-points, vectors, and targets corresponding to pre-election day activities.

Figure 4 | Vulnerability points and potential actors who may inflict damage during pre-election day preparations.



[60] Mazzei P., 2019. FBI to Florida Lawmakers: You Were Hackedby Russians, but Don't Tell Voters." The New York Times. https://www.nytimes.com/2019/05/16/us/florida-election-hacking-russians-fbi.html

[61] Stanford Policy Center, 2019.

> In the last decade, millions of voters across the globe have had their personal information leaked publicly or sold in the underground economy and dark web.

Counting Votes and Computerized Tabulation

When the votes are transmitted from voting machines at a polling station to a centralized tabulation system supervised by election authorities, poll workers aggregate the data from each polling station on an election management system (EMS). EMS are essentially computing software that aggregate and tabulate voters' ballots from disparate voting machines into publishable results. Hosted on desktop computers, connected to voting machines by removable media and powered by private – sometimes public – networks, EMS are prime targets for AI-cyberattacks through software or hardware vulnerabilities.

Memory cards transferred from voting machines to EMS could be a vector for arbitrary code execution. If EMS are connected to the Internet or an Intranet connected to the Internet, they become vulnerable to remote AI malware taking advantage of weak authenticity checks, weak password security, poorly executed encryption and software or SQL injection vulnerabilities.

> We are essentially facing, at this stage, a context of pervasive information insecurity with corrosive implications for voter data integrity.

Cybercriminals could reach a relatively large attack surface by gaining arbitrary command of computerized tabulation systems: from deleting or manipulating votes, interfering with vote count or crashing election infrastructure.

Disruptive Attacks on Internet and Electricity Networks

Beyond compromising the voting process directly, cybercriminals may seek to target critical infrastructure that plays a strategic role in the conduct of an election, such as shutting down Internet and electricity networks. In January 2019, ahead of the presidential election in Ukraine, the cyber police warned that "critical infrastructures in sectors such as energy and banking may again become the object of cyberattacks during or before the elections using malware to create so-called backdoors for a large coordinated attack."[62]

Across the world, at any given moment, there are pervasive offensive cyber-operations being waged for the control of critical urban infrastructures. These battles for influence and control tend to occur in peacetime, infiltrating local governments and smart cities. In 2018 and 2019, Baltimore and Atlanta were paralyzed for days under ransomware attacks, shutting down critical services such as airports and 911 emergency call centers. In Johannesburg and Hyderabad, ransomware attacks affected electricity companies' ability to respond to power failures. Increasingly, such disruptive cyberattacks on critical electric and internet information infrastructure could target the electoral process.

Over the past years, internet shutdowns have become more frequent – with 134 instances in India during 2018 alone.[63] Of significant scope and disruption, sometimes targeted at subpopulations, intentional and ordered or sanctioned by authorities, internet shutdowns are the new reality of a current digital techno-war. According to the Special Rapporteur on the rights to freedom and peaceful assembly and association, elections have suffered growing numbers of network disruptions and social media bans since 2016. "The Special Rapporteur believes network shutdowns are in clear violation of international law and cannot be justified in any circumstances (...) in the context of protests and

[62] Polityuk P., 2019. "Ukraine says it sees surge in cyber attacks targeting election." Reuters. https://www.reuters.com/article/us-ukraine-cyber-exclusive/exclusive-ukraine-says-it-sees-surge-in-cyber-attacks-targeting-election-idUSKCN1PJ1KX

[63] Data from Access Now

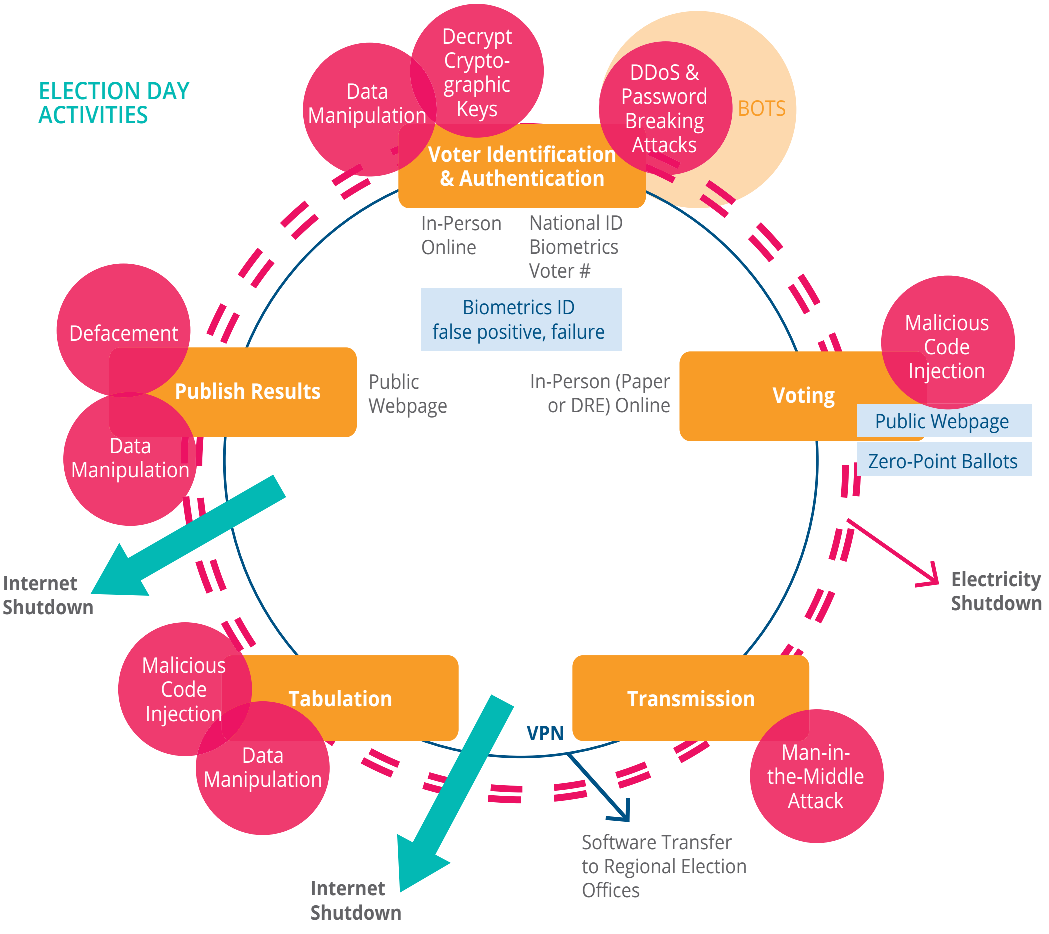
elections, when tensions are at the highest, these tools are actually needed to prevent disinformation and dispel rumors." [64]

Network shutdowns have implications beyond campaigning activities and the regulation of information warfare. Figure 3, 5 and 6 show the extensive impact of electricity and internet shutdowns on the conduct of an e-voting process. Cyber-strikes to compromise an electricity network at crucial voting time would have serious damaging effect on a large spectrum of semi-digitized to fully-digitized electoral processes.

> Beyond compromising the voting process directly, cybercriminals may seek to target critical infrastructure that plays a strategic role in the conduct of an election, such as shutting down Internet and electricity networks.

Table 1, as well as Figure 4 and 5, show the interdependencies between AI-cyber threats, in particular, entry-points, attack vectors, and targets corresponding to pre-election and election day activities.

Figure 5 | Vulnerability points and particular targets on election day.



[64] A/HRC/41/41, paras. 52-53.

The Way Forward - Adversarial Testing & Multistakeholder Teaming in Complex Electoral Cyber-Ecosystems

4

How do you strategically influence elections in the digital age? Cyberattacks are not the only way, but one, increasingly disruptive way. They can infiltrate hardware and software, or corrupt the integrity of electoral data-sets. They can harness human vulnerabilities and political tensions, leading to social and behavioural engineering. Ultimately, they can sow distrust, confusion and anger, helping win the battle for citizens' hearts and minds.

> Electoral cybersecurity in the era of technological convergence is a challenge our societies need to face.

Electoral cybersecurity in the era of technological convergence is a challenge our societies need to face. It will require a cognitive turn into how we apprehend and protect digital assets, such as sensitive human and security data, and how we approach complex socio-technical systems where human and machine behaviors intersect and influence each other. In other words, we need to understand and master a human-centred, data-driven, and holistic, non-reductionist approach to election cybersecurity.

This paper does not pretend to provide recommendations comprehensive enough to prevent and deter the complex and evolving nature of AI-enabled cyber-threats to elections as they have been elaborated in section 2-4. Yet, it aims to emphasize the tenets of the collective cognitive and strategic approach needed to confront existing and future cyber-threats.

First, electoral cybersecurity should be considered in a holistic context where complex alliances of human- and machine-driven deceptive tactics are used to generate cyber- and information warfare. In this context, governments and parliaments should consider the full spectrum of what election interference is, view elections as critical national infrastructure or essential critical services, and qualify electoral cybersecurity as a "public-private-civil" partnership or multi-stakeholder partnership.

Second, governments should develop and financially support an organisational structure that enables robust electoral cyber-security in this multistakeholder perspective. EMBs and government officials, and registered political parties would gain from sharing instrumental knowledge on cutting-edge AI-enabled cybersecurity with private sector companies that are pioneering defensive techniques such as digital forgery detection, DDoS detection, and AI-enabled models of intrusion and anomaly detection in critical infrastructure. This is even more crucial in the face of a dire diagnosis made by election security experts, who posit that most EMBs lack dedicated cybersecurity officers, leading to a diminished knowledge and appreciation of the real types of threats they may be facing, thus hindering their

> Extensive use of red-teaming exercises to detect and fix security vulnerabilities should become a priority for EMBs, IT government specialists and private sector vendors involved in election technology.

> As AI increasingly integrates with cybersecurity, governments and EMBs should promote a culture of responsible governance that relies on a human-centred understanding of risks and vulnerabilities in election technology.

ability to prepare and mitigate.[65] Not to speak of the bleak picture when it comes to opposition parties who often lack access to the infrastructure and resources.

Third, within a multi-stakeholder model, EMBs and government officials should be exposed to proactive, comprehensive methods of risk assessment and management under adversarial conditions. At the intersection of cybersecurity and AI, they would need to explore and implement adversarial testing and red-teaming exercises, improved software self-integrity verification, updated data authentication techniques and responsible disclosure of cyber-AI vulnerabilities. Extensive use of red-teaming exercises to detect and fix security vulnerabilities should become a priority for EMBs, IT government specialists and private sector vendors involved in election technology.

In section 3, this paper provides a mapping exercise, a general matrix of an election cycle and infrastructure, analysing the landscape of converging threats and vulnerabilities, attack vectors, data-targets and implications. This methodology can help identify "incentive structures, interactive effects, and leverage points" in the electoral process. Specifically, system maps can point out particular vulnerabilities through their visualization of the process at hand, along with key relationships and connections which may serve as weak links in the process.[66] These exercises are also highly effective for identifying points of intervention within the system to address vulnerabilities.

Fourth, and most importantly as AI increasingly integrates with cybersecurity, governments and EMBs should promote a culture of responsible governance that relies on a human-centred understanding of risks and vulnerabilities in election technology. This is even more crucial as AI technologies, such as precision biometrics attacks, forgeries, spear-phishing and emotional engineering, will target human intelligence and behaviors when testing the resilience of the electoral infrastructure. Such human-centred approach to responsible governance needs to involve not only electoral officials, but also auxiliary targets, including political candidates and political campaign staff.

Although AI systems can seek and find repetitive patterns at a much faster rate compared to humans, they may not always predict security flaws as reliably as collective human intelligence does. Electoral breaches might happen because security paradigms are predominantly relying on automated systems instead of including highly trained professionals.

Thought leaders[67] are driving a shift in mindset across the whole cyber security industry, from one that was very much focused on the technical aspects of keeping hackers at bay, to a more holistic and practical view of the best way to protect human actors and digital assets. Doing this involves taking a design-centric view of the electoral process, looking at the entire election cyber-ecosystem with human behavior as part of it, rather than implementing overly strict, technical and impractical rules and policies.

> Thought leaders are driving a shift in mindset across the whole cyber security industry, from one that was very much focused on the technical aspects of keeping hackers at bay, to a more holistic and practical view of the best way to protect human actors and digital assets

Such a move towards a human-centred approach in electoral cybersecurity might also benefit from increasing strategic communications towards the public to preserve resilience and promote transparency about the election process and its security.

[65] IFES, 2018. pg 21.

[66] IFES, 2018. pg 25.

[67] Stanford Policy Center, 2019.

> A move towards a human-centred approach in electoral cybersecurity might also benefit from increasing strategic communications towards the public to preserve resilience and promote transparency about the election process and its security.

Fifth, other policy interventions related to election security could centre around protecting data-integrity and voters' privacy. As laid out by IFES, this can be accomplished through ensuring that "election legislation is harmonized with data protection legislation or includes articles about the protection of private citizen information, drawing on international principles."[68] Some countries have strongly pursued this vision, including the Philippines which have chosen to pursue criminal charges against the country's election board following an electoral system and voter registration hacking incident.[69] It is of utmost importance to design and implement adequate steps to protect citizen information, including biometrics data.

Sixth, governments from different nations should share learning and experiences in election security and work collaboratively to establish and maintain international standards regarding best practices surrounding election security. This may include, but is not limited to, voting technologies, software, and security strategies.

In the future, UN Member States will need to develop a common understanding of how technological convergence impacts election security to be able to design proper oversight in collaboration with strategic actors in the private sector and civil society. States lagging behind in Cyber-AI convergence are the most at risk and the least likely to have any adversarial testing and foresight capacity.

Increasingly, it will become urgent to re-think the integration of electoral cybersecurity into innovative global development strategies. States interested in fostering responsible Cyber-AI convergence could enter into mechanisms of digital cooperation with countries in the Global South to partner around mutually beneficial transfers of data, talent, technologies, security practices, and lessons learned.

> Increasingly, it will become urgent to re-think the integration of electoral cybersecurity into innovative global development strategies.

[68] IFES, 2018. pg 25.

[69] IFES, 2018. pg 18.

[70] Herpig S., et al, 2018. pg 35-36.

Author: Eleonore Pauwels

Eleonore Pauwels is an international expert in the security, societal and governance implications generated by the convergence of artificial intelligence with other dual-use technologies, including cybersecurity, genomics and genome-editing. Pauwels provides expertise to the World Bank, the United Nations and the Global Center on Cooperative Security in New York. She also works closely with governments and private sector actors on AI-Cyber Prevention, the changing nature of conflict, foresight and global security.

In 2018 and 2019, Pauwels served as Research Fellow on Emerging Cybertechnologies for the United Nations University's Centre for Policy Research. At the Woodrow Wilson International Center for Scholars, she spent ten years within the Science and Technology Innovation Program, leading the Anticipatory Intelligence Lab. She is also part of the Scientific Committee of the International Association for Responsible Research and Innovation in Genome-Editing (ARRIGE). Pauwels is a former official of the European Commission's Directorate on Science, Economy and Society.

Pauwels regularly testifies before U.S. and European authorities including the U.S. Department of State, NAS, NIH, NCI, FDA, the National Intelligence Council, the European Commission and the UN. She writes for Nature, The New York Times, The Guardian, Scientific American, Le Monde, Slate, UN News, The UN Chronicle and The World Economic Forum.

Cybersecurity Glossary

Adversarial Machine Learning: Adversarial machine learning is a technique employed in the field of machine learning which attempts to manipulate the functioning of an algorithmic model and undermine its performance through injecting malicious input such as noise or signals that fool classification.

AI-driven or Autonomous malware: autonomous malware can self-propagate via a series of autonomous decisions, intelligently tailored to the parameters of the infected system. For instance, such malware can learn context by quietly sitting in an infected environment and observing normal business operations, such as the internal devices the infected machine communicates with, the ports and protocols it uses, and the accounts which use it. Malware authors can maximize their profits if their malware can also choose autonomously which payload will yield the highest profit based on the context of the environment and infected machine. Autonomous malware can therefore learn to choose whatever method appears most successful for the target environment and use this to move laterally, propagate and compromise the host system.

Botnet: A botnet refers to a group of computers which have been infected by malware and have come under the control of a malicious actor. Self-propagating botnets can recruit additional bots through a variety of different channels. Pathways for infection include the exploitation of website vulnerabilities, malware, and cracking weak authentication to gain remote access. Botnets can be designed to accomplish illegal or malicious tasks including sending spam, stealing data, ransomware, fraudulently clicking on ads or distributed denial-of-service (DDoS) attacks.

Data-Poisoning Attacks: Data Poisoning is an adversarial attack that aims to manipulate the training dataset in order to control the prediction behavior of a trained algorithmic model such that the model will label malicious examples into a desired category (e.g., labeling spam e-mails as safe). Data poisoning attacks can therefore subvert the learning process for the machine learning system and/or degrade the performance of the system.

DDOS: A Distributed Denial of Service is a cyber-attack in which the perpetrator seeks to make a machine or network resource unavailable to its intended users by temporarily or indefinitely disrupting services of a host connected to the Internet. Denial of service is typically accomplished by flooding the targeted machine or resource with superfluous requests in an attempt to overload systems and prevent some or all legitimate requests from being fulfilled. In a distributed denial-of-service attack, the incoming traffic flooding the victim originates from many different sources such as botnets. This effectively makes it impossible to stop the attack simply by blocking a single source.

DNS Hijacking: Domain Name Server (DNS) hijacking, also named DNS redirection, is a type of DNS attack in which DNS queries are incorrectly resolved in order to unexpectedly redirect users to malicious sites. To perform the attack, perpetrators either install malware on user computers, take over routers, or intercept or hack DNS communication. DNS hijacking can be used for phishing, by displaying fake versions of a website users' access and then stealing data or credentials. Some governments use DNS hijacking for censorship, redirecting users to government-authorized sites.

Injection Attacks: Injection attacks refer to a broad class of attack vectors. In an injection attack, an attacker supplies untrusted input to a program. This input gets processed by an interpreter as part of a command or query. In turn, this alters the execution of that program. Injections are amongst the oldest and most dangerous attacks aimed at web applications. They can lead to data theft, data loss, loss of data integrity, denial of service, as well as full system compromise. The primary reason for injection vulnerabilities is usually insufficient user input validation.

In the case of a **"Malicious Code Injection,"** the attacker injects application code written in the application language. This code may be used to execute operating system commands with the privileges of the user who is running the web application. In advanced cases, the attacker may exploit additional privilege escalation vulnerabilities, which may lead to full web server compromise.

In the case of an **SQL (Structured Query Language) Injection Attack**, the attacker injects SQL statements that can read or modify database data. In the case of advanced SQL Injection attacks, the attacker can use SQL commands to write arbitrary files to the server and even execute operating system commands. This may lead to full system compromise.

Password Breaking Attacks: Password breaking or "cracking" attack is the process of recovering passwords from data that has been stored in or transmitted by a computer system. A common but work intensive approach (brute-force attack) is to repeatedly try guesses for the password and to check them against an available cryptographic hash of the password. Alternatively, as an optimized approach, an algorithmic program can take a dictionary of words and commonly used passwords – as well previously cracked passphrases – and turn them into hashes to check against the stolen hash or hashes. Increasingly, algorithms can be trained to predict the passwords people are going to use, or using right now, based on what they've all done in the past. For instance, a **deep learning approach** to password cracking uses a Generative Adversarial Network (GAN) to autonomously learn the distribution of real passwords from actual password leaks, and to generate high-quality password guesses.

Personalized Spear-Phishing: Personalized spear-phishing leverages AI to tailor phishing emails to specific users in order to increase chances of infecting the system. AI malware can learn to rely on behavioural surveillance, affect-recognition, context-understanding and anomaly-detection to analyse individuals' emotions, language and behaviour. Such AI malware can then learn to impersonate an individual's trusted contacts within professional and personal social networks. Tailored communication generated by AI malware will therefore be almost impossible to distinguish from human peers' communications.

Precision Biometrics Attacks: Biometric authentication systems are a next type of targets for AI-driven cyberthreats. Precision biometrics attacks are emerging types of forgeries and impersonations that use biometrics characteristics – such as voice tone and modulation, facial features and expressions – to manipulate users' behaviours or get authentication access to an operating system.

- **Audio-spoofing** is a type of attacks where malicious actors alter an audio recording of a voice, such that it mimics a target speaker's voice to access a system protected by automatic speaker verification (AVS). Given the recent advances in audio processing technology, it is becoming easier to synthesize speech in such a way that it sounds like a given target speaker. These technologies can be used by security hackers to break into ASV systems or convince a human target to release sensitive information within an organisation.

- **Deepfake** are another type of impersonations that rely on individuals' biometrics. Deep-learning generative algorithmic models combined with facial-mapping software enable the cheap and easy fabrication of content that hijacks one's identity—such as face, expressions and body. In August 2019, researchers in Israel published a **new method** for making Deepfakes by creating realistic face-swapped videos in real-time, with no extensive facial data-training. Deep-learning algorithms – called FSGAN – can pinpoint facial biometrics features in a video, then align the source face to the target's face. Algorithms that do not need to be trained on each new face target provide a powerful toolkit to create realistic video forgeries at scale and with minimal know-how.

Precision Social and Emotional Engineering: Cybercriminals use AI to automate new forms of social engineering. The combination of psychometrics manipulation tools with personal datasets can help craft convincing emotion-targeting campaigns that can hardly be recognized as malicious. Even the most experienced users might fall for such personalized attacks. By allowing the analysis of individual communication, perception and emotion to be automated, AI systems can increase anonymity and psychological distance in cyber operations. In the near-future, automated cyber operations, led by machine-learning, will therefore be more effective, finely targeted, difficult to attribute, and likely to exploit evolving vulnerabilities in AI and human systems. One pervasive security threat will be new forms of hybrid influencing made possible by the automation of social-engineering attacks. Many major cybersecurity incidents rely on social engineering where malicious actors target the social and psychological vulnerabilities of humans within chains of command. The goal is to manipulate command and control organizations to compromise their own safety and security.

RAT: A Remote Access Trojan (RAT) is a type of malware that allows hackers to monitor and control users' computers or networks. Hackers can then wipe infected computers' hard drive, download illegal content from the internet, place additional malware or activate an infected computer's webcam or microphone discreetly. Hackers can also use a RAT to obtain keystrokes and files from an infected computer. These keystrokes and files could contain bank information, passwords, sensitive photos, or private conversations. Additionally, hackers can control infected computers remotely to perform embarrassing or illegal actions or harness a user's network as a proxy server to commit crimes anonymously.

Website Defacement: Web defacement is an attack in which malicious parties penetrate a website and replace content on the site with their own messages. The messages can convey a political or religious message, profanity or other inappropriate content that would embarrass website owners, or a notice that the website has been hacked by a specific hacker group. Most websites and web applications store data in environment or configuration files, that affects the content displayed on the website, or specifies where templates and page content is located. Unexpected changes to these files can mean a security compromise and might signal a defacement attack. A defacement attack therefore acts as a public indicator that a website has been compromised, and causes damage to the brand and reputation, which lasts long after the attacker's message has been removed.

Whaling Attack: A whaling attack is a method used by cybercriminals to impersonate a senior player at an organization and directly target senior or other important individuals at an organization, with the aim of stealing money or sensitive information or gaining access to their computer systems for criminal purposes. Also known as "CEO fraud," whaling is similar to spear-phishing in that it uses methods such as personalization, social engineering or DNS hijacking to trick a target into performing specific actions, such as revealing sensitive data or transferring money. Whereas phishing scams target non-specific individuals and spear-phishing targets particular individuals, whaling doubles down on the latter by not only targeting those key individuals, but doing so in a way that the fraudulent communications they are sent appear to have come from someone specifically senior or influential at their organization.

ANNEX: Bibliography and Selected List of Expert Interviews and Consultations

- Caio Machado, Oxford Internet Institute and University of Sao Paulo, December 18, 2019
- Diego Aranha, Cybersecurity Exert – Electronic Voting in Brazil, Department of Engineering, University Aarhus and University of Campinas, December 20, 2019
- Christina Nemr, Cybersecurity and Disinformation Expert, US Department of State and Park Advisors, December 20, 2019
- Rushdi Nackerdien, Regional Director: Africa, IFES, December 13, 2019
- Niels Nagelhus Schia, Senior Research Fellow, NUPI (Norwegian Institute of International Affairs), Center for Cybersecurity Studies, December 13, 2019
- Paul Sambo, Expert on Election in South Africa, University of Pretoria, December 11, 2019
- Steve Martin, UN Electoral Assistance Division, December 9, 2019
- Maarten Halff, UN Electoral Assistance Division, December 9, 2019
- Professor Jarno Limnell, Cybersecurity, University of Aalto (Finland) and CEO of Tosibox, December 9, 2019

Abadi M. and Andersen D., 2016. "Learning to Protect Communications with Adversarial Neural Cryptography." arXiv. https://arxiv.org/pdf/1610.06918.pdf

Akamai, 2018. "State of the Internet." https://www.akamai.com/us/en/multimedia/documents/state-of-the-internet/soti-2018-credential-stuffing-attacks-report.pdf

Aranha D. and van de Graaf J., 2018. "The Good, the Bad, and the Ugly: Two Decades of E-Voting in Brazil." IEEE Security & Privacy. https://www.computer.org/csdl/magazine/sp/2018/06/08636417/17D45VsBU6S

Aranha D. et al, 2019. "The return of software vulnerabilities in the Brazilian voting machine." Computers & Security. https://www.sciencedirect.com/science/article/pii/S0167404819301191

Brady M. And Bucholz K., 2019. "SNAKEMACKEREL Delivers SedUploader Malware." Accenture. https://www.accenture.com/us-en/blogs/blogs-snakemackerel-delivers-seduploader-malware

Burt C., 2019. " National digital ID system in Brazil delayed by lack of deal for biometric database access." Biometric Update. https://www.biometricupdate.com/201912/national-digital-id-system-in-brazil-delayed-by-lack-of-deal-for-biometric-database-access

Cisco, "Common Cyber Attacks." https://www.cisco.com/c/en/us/products/security/common-cyberattacks.html

Cisco, 2019. "Threats of the Year." https://www.cisco.com/c/dam/en/us/products/collateral/security/2019-threats-of-the-year-cybersecurity-series-dec-2019.pdf

Darktrace, 2018. "The Next Paradigm Shift AI-Driven Cyber-Attacks." https://pdfs.semanticscholar.org/6b18/6268d00e891f3ed282544ac5833c01a2891c.pdf

Darktrace, 2019. "Machine Learning in the Age of Cyber AI." https://www.darktrace.com/en/resources/wp-machine-learning.pdf

Dorell O., 2017. "Alleged Russian political meddling documented in 27 countries since 2004." USA Today. https://www.usatoday.com/story/news/world/2017/09/07/alleged-russian-political-meddling-documented-27-countriessince-2004/619056001/

Feldman J., Halderman J., and Felten E., 2017. "Security Analysis of the Diebold AccuVote-TS Voting Machine." https://www.usenix.org/legacy/event/evt07/tech/full_papers/feldman/feldman_html/index.html

FireEye, 2018. "Attacking the Ballot Box: Threats to Election Systems.". https://media.scmagazine.com/documents/343/election_systems_report_85540.pdf

Gomez A. et al., 2018. "Unsupervised Cipher Cracking Using Discrete GANs." https://arxiv.org/pdf/1801.04883.pdf

Gonggrijp R. and Hengeveld W., 2007. "Studying the Nedap/Groenendaal ES3B voting computer: A computer security perspective." Proc. Electronic Voting Technology Workshop. https://www.usenix.org/conference/evt-07/studying-nedap-groenendaal-es3b-voting-computer-computer-security-perspective

Herpig S., et al, 2018. "Securing Democracy in Cyberspace." https://www.stiftung-nv.de/sites/default/files/securing_democracy_in_cyberspace.pdf

International Foundation for Election Security (IFES), 2018. "Cybersecurity in Elections." https://www.ifes.org/sites/default/files/2018_heat_cybersecurity_in_elections.pdf

Isaac M. and Roose K., 2018. "Disinformation Spreads of WhatsApp Ahead of Brazilian Election." The New York Times. https://www.nytimes.com/2018/10/19/technology/whatsapp-brazil-presidential-election.html

King T. et al, 2019. "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions." https://link.springer.com/content/pdf/10.1007%2Fs11948-018-00081-0.pdf

Kirat D. et al, 2018. "DeepLocker: Concealing Targeted Attacks with AI Locksmithing." Black Hat USA 2018.

Lomas N., 2017. "Lyrebird is a voice mimic for the fake news era." TechCrunch. https://techcrunch.com/2017/04/25/lyrebird-isa-voice-mimic-for-the-fake-news-era/

Lorin H. et al., 2017. "The Business of Voting: Market Structure and Innovation in the Election Technology Industry." University of Pennsylvania, Wharton School. https://publicpolicy.wharton.upenn.edu/business-of-voting/.

MalwareBytes, 2019. "Emotet is back: botnet springs back to life with new spam campaign." https://blog.malwarebytes.com/botnets/2019/09/emotet-is-back-botnet-springs-back-to-life-with-new-spam-campaign/

Mari A., 2019. "Brazilian government to create single citizen database." ZDNet. https://www.zdnet.com/article/brazilian-government-to-create-single-citizen-database/

Mayer M., 2018. "Artificial Intelligence and Cyber Power from a Strategic Perspective." IFS Insights; April. https://brage.bibsys.no/xmlui/bitstream/handle/11250/2497514/IFS%20Insights_4_2018_Mayer.pdf

Mazzei P., 2019. FBI to Florida Lawmakers: You Were Hackedby Russians, but Don't Tell Voters." The New York Times. https://www.nytimes.com/2019/05/16/us/florida-election-hacking-russians-fbi.html

Miller J., 2019. "Brazil to create massive biometric database of all citizens." TechSpot. https://www.techspot.com/news/82321-brazil-create-massive-biometric-database-all-citizens.html

Mirsky Y., 2019. "CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning." https://arxiv.org/abs/1901.03597

NIS Cooperation Group, 2018. "Compendium on Cyber Security of Election Technology." https://www.ria.ee/sites/default/files/content-editors/kuberturve/cyber_security_of_election_technology.pdf

Norden L., Deluzio C., and Ramachandran G., 2019. "A Framework for Election Vendor Oversight." Brennan Center for Justice, New York University School of Law. https://www.brennancenter.org/sites/default/files/2019-11/2019_10_Election-Vendors.pdf

Organization of American States, 2018. "Preliminary Report." http://www.oas.org/documents/eng/press/Preliminary-Report-EOM-Brasil-2nd-round-ENG.pdf

Osborne C., 2018. "Deeplocker: When malware turns artificial intelligence into a weapon." ZDNet. https://www. zdnet.com/article/deeplocker-when-malware-turns-artificial-intelligence-into-a-weapon/

Pascu L., 2019. "Brazilian consumer rights watchdog vetoes facial recognition tech." Biometric Update. https://www.biometricupdate.com/201909/brazilian-consumer-rights-watchdog-vetoes-facial-recognition-tech

Pauwels E., 2019. "The New Geopolitics of Converging Risks." https://collections.unu.edu/eserv/UNU:7308/PauwelsAIGeopolitics.pdf

Polityuk P., 2019. "Ukraine says it sees surge in cyber attacks targeting election," Reuters. https://www.reuters.com/article/us-ukraine-cyber-exclusive/exclusive-ukraine-says-it-sees-surge-in-cyber-attacks-targeting-election-idUSKCN1PJ1KX

Reinhart R., 2018. “Brazilians Face Confidence Crisis Ahead of Election.” Gallup. https://news.gallup.com/poll/243161/brazilians-face-confidence-crisis-ahead-key-election.aspx

Risk and Resilience Team, 2017. “Hotspot Analysis: Cyber and Information Warfare in Elections in Europe.” Center for Security Studies, ETH Zurich. https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/Cyber-Reports-2017-08.pdf

Seymour J. and Tully P. “Weaponizing Data Science for Social Engineering.” https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter.pdf

Special Counsel Robert S. Mueller, III., “Report On The Investigation Into Russian Interference In The 2016 Presidential Election: Volume I,” U.S. Department of Justice, March 2019, 49-51 (“Mueller Report”). https://www.justice.gov/storage/report.pdf

Stanford Policy Center, 2019. “Securing American Elections.” https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/stanford_cyber_policy_center-securing_american_elections.pdf

Szoldra P., 2016. “A hacker is selling a database of all US voters for $7,800 on the dark web.” Business Insider. https://www.businessinsider.fr/us/hacker-voter-registration-database-2016-7

The Associated Press, 2019. “Russian-owned company attempted Ohio election hack”. https://apnews.com/6518b9a986f640c4899a979bbc48390b

United States Senate. “Report Of the Select Committee on Intelligence On Russian Active Measures Campaigns and Interference in the 2016 U.S. Election,” Volume 1: Russian Efforts Against Election Infrastructure with Additional Views. Report 116-XX. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume1.pdf

Villasenor J., 2018. “Artificial Intelligence and the future of geopolitics.” Brookings Institute. https://www.brookings.edu/blog/techtank/2018/11/14/artificial-intelligence-and-the-future-of-geopolitics/

Wolchok S. et al., 2010. “Security analysis of India’s electronic voting machines.” Proc. ACM Conf. Computer and Communications Security.

Zetter K., 2019. “Hospital viruses: Fake cancerous nodes in CT scans, created by malware, trick radiologists.” The Washington Post. https://www.washingtonpost.com/technology/2019/04/03/hospital-viruses-fake-cancerous-nodes-ct-scans-created-by-malware-trick-radiologists/

Konrad-Adenauer-Stiftung e. V.

Andrea E. Ostheimer
Executive Director
www.kas.de/newyork

andrea.ostheimer@kas.de

www.kas.de/newyork