UNITED NATIONS
OFFICE OF COUNTER-TERRORISM
UN Counter-Terrorism Centre (UNCCT)

unicri
United Nations
Interregional Crime and Justice
Research Institute

ALGORITHMS AND TERRORISM:

THE MALICIOUS USE OF ARTIFICIAL INTELLIGENCE FOR TERRORIST PURPOSES

ALGORITHMS AND TERRORISM:

THE MALICIOUS USE OF ARTIFICIAL INTELLIGENCE FOR TERRORIST PURPOSES

A Joint Report by UNICRI and UNCCT

Disclaimer

The opinions, findings, conclusions and recommendations expressed herein do not necessarily reflect the views of the United Nations, the Government of the Kingdom of Saudi Arabia or any other national, regional or global entities involved.

The designation employed and material presented in this publication does not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area of its authorities, or concerning the delimitation of its frontiers or boundaries.

Acknowledgements

This report is the product of a joint research initiative on counter-terrorism in the age of artificial intelligence of the Cyber Security and New Technologies Unit of the United Nations Counter-Terrorism Centre (UNCCT) in the United Nations Office of Counter-Terrorism (UNOCT) and the United Nations Interregional Crime and Justice Research Institute (UNICRI) through its Centre for Artificial Intelligence and Robotics. The joint research initiative was funded with generous contributions from the Kingdom of Saudi Arabia.

Copyright

United Nations Office of Counter-Terrorism
S-2716
United Nations
405 East 42nd Street
New York, NY 10017
Website: www.un.org/counterterrorism/

Viale Maestri del Lavoro,10, 10127 Torino – Italy
Website: www.unicri.it
E-mail: unicri.publicinfo@un.org

FOREWORD

Over the past decade, we saw a speedy adoption of Artificial Intelligence (AI) solutions within various industries, by both public and the private sectors. It is predicted that the global AI market will exceed $100 billion by 2025 and AI enabled systems will continue to support many sectors – healthcare, education, commerce, banking and financial services, critical infrastructure, and security, among many others.

As stated by the United Nations Secretary-General António Guterres in his 2018 Strategy on New Technologies, "While these technologies hold great promise, they are not risk-free, and some inspire anxiety and even fear. They can be used to malicious ends or have unintended negative consequences". The potential benefits of AI to humanity are undeniable, and yet, research on the malicious use of AI is still in its infancy.

Terrorists have been observed to be early adopters of emerging technologies, which tend to be under-regulated and under-governed, and AI is no exception. Given the international linkages and cross-border implications of many technological systems, a regional and international approach becomes vital to ensure terrorists do not have the opportunity to exploit regulatory gaps that can expose vulnerabilities in AI systems. We need to build resilient governing structures that can quickly and effectively respond to and mitigate the impact of the malicious use of AI by terrorists.

The United Nations is responding to this need with a wide array of initiatives. The Secretary-General's Roadmap for Digital Cooperation sets "Supporting Global Cooperation on Artificial Intelligence" as one of the eight key areas for action. In line with this Roadmap, the United Nations Counter-Terrorism Centre in the United Nations Office of Counter-Terrorism is also responding to this challenge through its Global Counter-Terrorism Programme on Cybersecurity and New Technologies.

This report developed jointly with the United Nations Interregional Crime and Justice Research Institute should serve as an early warning for potential malicious uses and abuses of AI by terrorists and help the global community, industry and governments to proactively think about what we can do collectively to ensure new technologies are used to bring good and not harm.

I would like to take this opportunity to thank the international experts who were involved in shaping the recommendations of this report. My Office stands ready to support Member States and other counter-terrorism partners in countering the threat of AI by terrorists.

Vladimir Voronkov
Under-Secretary-General
United Nations Office of Counter-Terrorism
Executive Director
United Nations Counter-Terrorism Centre

FOREWORD

Artificial intelligence (AI) is arguably the quintessential emerging technology of our time. For several years now, the United Nations Interregional Crime and Justice Research Institute, through its Centre for AI and Robotics, has been exploring AI and what we have seen is immensely promising. In the context of today, AI has, for instance, played a role in helping to significantly speed up the development of messenger ribonucleic acid (mRNA) based vaccines, such as those now being used to rein in the COVID-19 pandemic. In our areas of work in the fields of justice, crime prevention, security and the rule of law, we have seen promising uses of AI, including its ability to help locate long-missing children, scan illicit sex ads to identify and disrupt human trafficking rings, and flag financial transactions that may indicate money laundering.

But we have also seen the dark side of AI – a side that has not received as much attention and remains underexplored. The reality is that AI can be extremely dangerous if used with malicious intent. With a proven track record in the world of cybercrime, it is a powerful tool that could conceivably be employed to further or facilitate terrorism and violent extremism conducive to terrorism, by, for instance, providing new modalities for physical attacks with drones or self-driving cars, augmenting cyberattacks on critical infrastructure, or enabling the spread of hate speech and incitement to violence in a faster and more efficient way.

Is AI the future of terrorism? As this report indicates, this remains to be seen. Even still, we must never forget that terrorism is an evolving threat that should not be underestimated. More than two decades into the 21st century, we have seen many examples of terrorists turning to new and emerging technologies such as drones, virtual currencies and social media. With AI increasingly becoming more accessible, it is imperative to stay ahead of the curve and be prepared for any eventuality involving its misuse.

We are therefore proud to present this report together with the United Nations Counter-Terrorism Centre at the United Nations Office on Counter-Terrorism, which has been made possible with the generous support of the Kingdom of Saudi Arabia. It is our aspiration that it is the beginning of the conversation on the malicious use of AI for terrorist purposes.

Antonia Marie De Meo
Director
United Nations Interregional Crime and Justice Research Institute

EXECUTIVE SUMMARY

New technologies and artificial intelligence (AI) in particular, can be extremely powerful tools, enabling big advances in medicine, information and communication technologies, marketing, transportation among many other research fields. However, they can also be used for malicious purposes when falling into the wrong hands. The scope of this report – *Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes* – is to contribute to understanding the potential risk of AI falling into the hands of terrorists.

Although terrorist organizations have, to a certain degree, traditionally tended to employ various forms of "low-tech terrorism" such as firearms, blades and vehicles, terrorism itself is not a stagnant threat. As soon as AI becomes more widespread, the barriers to entry will be lowered by reducing the skills and technical expertise needed to employ it. Therefore, the questions this report strives to answer are whether – or perhaps better "when" – AI will become an instrument in the toolbox of terrorism and, if that occurs, what the international community might reasonably expect.

This report is organized into nine chapters:

Chapter one provides a general overview, providing statistics that demonstrate growing concerns amongst experts regarding the malicious use of this technology, including by terrorists.

Chapter two describes the general landscape of AI. It begins by defining AI and related terms, including machine learning and deep learning, and concepts such as narrow and general intelligence. It then provides an overview of the different areas that currently benefit from AI algorithms and applications, such as natural language processing and image recognition, as well as possible future trends in the use of this technology.

Chapter three demonstrates the potential threat of terrorist groups and individuals using new technologies by presenting several examples of terrorist attacks where technologies such as the Internet and social media have been valuable and powerful tools.

Chapter four seeks to further contextualize the malicious use of AI by examining three categories of threats – cyber, physical and political – that have been identified in existing literature in order to demonstrate how AI might be maliciously used.

Chapter five proceeds to address the question of whether AI-enabled terrorism could be a conceivable reality, or if it is little more than mere science fiction. For that purpose, it presents examples of terrorist groups that have demonstrated interest in AI or related technologies, including in videos using facial recognition or unmanned aerial systems, also known as "drones".

Following this, chapter six provides an in-depth overview of the present and possible future malicious uses of AI by terrorist groups and individuals. This overview includes both malicious uses that are documented and have been identified through research, and those that, despite the lack of evidence or literature, could become a future reality.

Chapter seven presents three fictional scenarios to support visualizations of how AI could be maliciously employed for terrorist purposes. These scenarios focus on the use of AI-powered password guessing, ransomware, drones with facial recognition, deepfakes and morphed passports made available through an underground forum in the "crime-as-a-service" business model.

Building upon the information presented in previous chapters, chapter eight assesses whether there is cause for concern about terrorist groups and individuals directly employing AI, for instance, to improve or amplify an attack. In this regard, the concepts of *intent* and *capability* are analyzed to reach objective conclusions.

Chapter nine brings the report to a conclusion by offering a set of recommendations for counter-terrorism bodies and law enforcement agencies, as well as policymakers, industry and academia to consider for the future, and suggesting several follow-up actions for capacity-building to prepare for the possible future of AI-enabled terrorism.

In the preparation of this report, UNOCT and UNICRI have relied predominantly on desk-based research and open-source information, such as articles, official reports and media reports. An Expert Group Meeting was organized virtually on 9 February 2021 to complement the conclusions reached on the basis of open-source information and collect insights for the strategic recommendations and follow-up actions presented.

CONTENTS

I. INTRODUCTION

Artificial intelligence (AI) is a powerful tool. It is being used throughout the public and private sector to make people, and society at large, happier, healthier, wealthier and safer. The United Nations Secretary-General, António Guterres, has indicated that, if harnessed appropriately and anchored in the values and obligations defined by the Charter of the United Nations and the Universal Declaration of Human Rights, AI can play a role in the fulfilment of the 2030 Agenda for Sustainable Development, by contributing to end poverty, protect the planet and ensure peace and prosperity for all.[1] Yet, AI can also have a dark side: as a general-purpose technology, AI can, just as equally, be used or misused by malicious actors. A recent report by Europol, Trend Micro and UNICRI highlighted some of the many ways in which cybercriminals are already using AI as both an attack vector and an attack surface.[2, 3, 4] Just as AI can be used for criminal purposes, it can also be maliciously used by groups and individuals to enhance the intensity of terrorist attacks or to amplify the potential of these groups or individuals to disseminate extremist propaganda and incite violence.[5]

In August 2020, MIT Technology Review Insights surveyed 301 senior business leaders and academics on a wide range of AI-related matters, including their concerns about AI. The survey indicated that, while issues such as lack of transparency, bias, lack of governance in the development of AI, and the potential for automation to cause significant unemployment were a source of concern, the participants were most worried about AI falling into the wrong hands.[6]

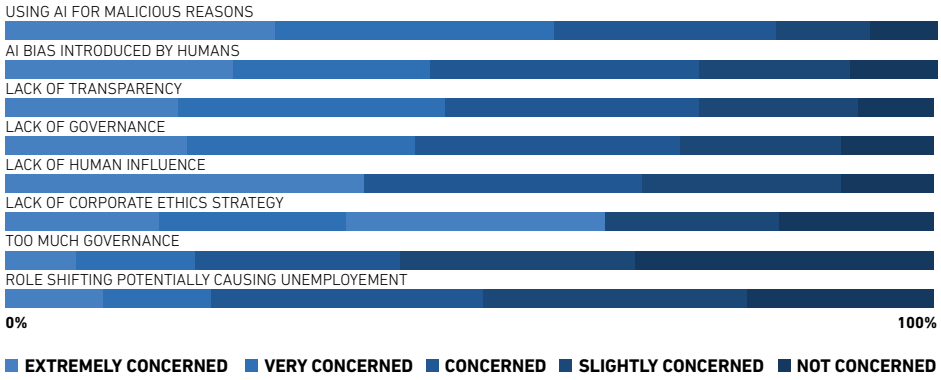


Figure 1: MIT Technology Review Insights demonstrates a prevalence of concerns about the malicious use of AI.

1 António Guterres. (Sept. 2018). Secretary-General's Strategy on New Technologies. United Nations. Accessible at https://www.un.org/en/newtechnologies/images/pdf/SGs-Strategy-on-New-Technologies.pdf

2 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philipp Amann, Aglika Klayn, Odhran McCarthy and Maria Eira. (Nov. 19, 2020). Malicious Uses and Abuses of Artificial Intelligence. Trend Micro Research. Accessible at http://unicri.it/sites/default/files/2020-11/AI%20MLC.pdf
The findings of this report serve as a point of departure for the present report and are based on the contributions from Trend Micro, Europol and UNICRI, which have been combined with input collected during a focused workshop in March 2020 that was organized by Trend Micro, Europol and UNICRI. Workshop participants included members from the Joint Cybercrime Action Taskforce (J-CAT), the International Criminal Court, and several members of Europol's European Cybercrime Centre (EC3) Advisory Groups.

3 Further noteworthy studies on the criminal use of AI have also be prepared by the Dawes Centre for Future Crime at University College London: Matthew Caldwell, Jerone T. A. Andrews, Thomas Tanay and Lewis Griffin. (Jul. 2020). Policy brief: AI-enabled future crime. Dawes Centre for Future Crime at University College London. Accessible at https://www.ucl.ac.uk/jill-dando-institute/sites/jill-dando-institute/files/ai_crime_policy_0.pdf

4 Link11. AI and Cyber Resilience: A Race between Attackers and Defenders. (Nov. 5, 2020). Link11. Accessible at https://www.link11.com/en/downloads/ai-and-cyber-resilience-a-race-between-attackers-and-defenders/

5 For the purposes of this report: Terrorist use of technology is considered to entail the use of a given technology for its intended purpose by terrorists, i.e. a terrorist using a messaging application to communicate with his or her associates; Terrorist misuse of technology is considered to entail the use of technology in conflict with its set terms and conditions, i.e. terrorists using social media platforms to incite violence; Malicious use of technology is considered to be a broader term that can encompass both use and misuse, and mainly refers to the intent with which a given technology is used.

6 MIT Technology Review Insights. (Nov. 30, 2020). A New Horizon: Expanding the AI landscape. MIT. Accessible at https://www.technologyreview.com/2020/11/30/1012528/a-new-horizon-expanding-the-ai-landscape/

A survey carried out by United Nations Counter-Terrorism Centre (UNCCT) of the United Nations Office of Counter-Terrorism (UNOCT) and UNICRI, through its Centre for Artificial Intelligence and Robotics, at an Expert Group Meeting designed to review and validate the present report, served to demonstrate a similar concern. From the 27 representatives from government, industry, academia and international and regional organizations, 44% felt that the malicious use of AI for terrorist purposes was "very likely" and 56% felt it was "somewhat likely". Tellingly, no surveyed participants felt that the malicious use of AI in this manner was "unlikely".[7]

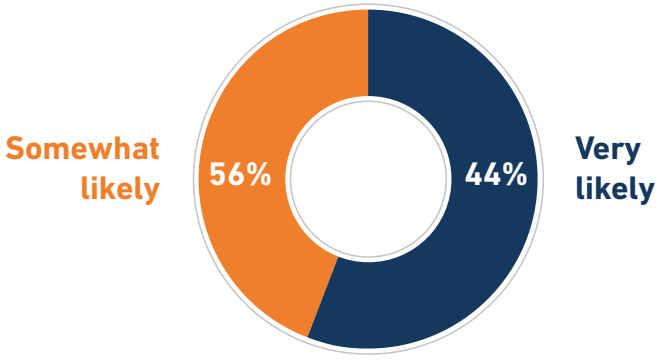


Figure 2: UNOCT-UNICRI Survey on perceived likelihood of malicious use of AI for terrorist purposes

In the discussion that followed the survey, four factors were identified by participants as contributing significantly to their concerns surrounding the potential malicious use of AI for terrorist purposes:

First, the "democratization" of new technologies such as AI. The notion of "democratization" refers to the fact that what was once an exceptionally advanced technology understood and used only by a very limited community with substantial resources and expertise, is becoming increasingly accessible to all and can be used without large investment or even with limited technical knowledge. In fact, many of the more popular algorithms are already open source and do not require an exceptionally high degree of expertise to be used. Even though the democratization of technology can in general be a driver of development and prosperity, the risk of the possible malicious use is equally heightened as a result. Moreover, considering also the possibility of such groups outsourcing in the form of the increasingly relevant "crime-as-a-service" business model used by criminal groups which "drives the digital underground economy by providing a wide range of commercial services that facilitate almost any type of cybercrime", the barriers for entry for the use of AI have been significantly lowered for actors with malicious intent.[8]

Second, the scalability of AI. Scalability can be understood as the ability of a technology to "grow" in size or volume and manage increased demand. Typically, scalability refers to making a piece of technology bigger and broader in terms of its use. It was observed by participants that in light of the particularly scalable nature of AI, those responsible for defending against potential malicious uses of AI would have to prepare for and defend against not only the threat of individual attacks, but an increased volume of attacks at any one point in time. A prime example of this is the threat of drone swarms autonomously flying in unison.

Third, the inherent asymmetry in terrorism/counter-terrorism. It was suggested that even if the malicious use of AI for terrorist purposes fails due, for example, to the lack of technical experience of the individual attackers, there can still be a considerable psychological impact in terms of instilling fear. For instance, a failed bomb attack nevertheless conveys a powerful message. Accordingly, as is often observed in the context of counter-terrorism, "we have to be lucky every time, they only have to be lucky once". At the same time, this asymmetry can be seen in the challenges

7 Insights from the UNOCT-UNICRI Expert Group Meeting. (Feb. 9, 2021). Participants in the Expert Group Meeting included representatives from the Austrian Institute of Technology; AWO; Chemonics; the Council of Europe; the European Commission - Directorate-General for Migration and Home Affairs (DG HOME); the European Union Agency for Law Enforcement Cooperation (Europol); the Foreign Ministry of the Russian Federation; Chatham House; the Geneva Centre for Security Policy; the Organization for Security and Cooperation in Europe (OSCE); Link11 Cyber Resilience; MalwareBytes; the North Atlantic Treaty Organization (NATO); Trend Micro; the United Kingdom Research and Innovation (UKRI) Trustworthy Autonomous Systems (TAS) Hub; the United Nations Counter-Terrorism Committee Executive Directorate (CTED); the United Nations University; and the Universities of College London, Bristol, Cambridge and Southampton

8 Internet Organised Crime Threat Assessment (IOCTA). (Sept. 29, 2014). Europol. Accessible at https://www.europol.europa.eu/sites/default/files/documents/europol_iocta_web.pdf

counter-terrorism entities and terrorists face with respect to the use of AI. For many entities seeking to leverage AI, careful consideration on its use is required to navigate civil liberties and fundamental human rights and freedoms. However, malicious actors such as terrorist groups and individuals are likely not to dwell on such concerns, thereby simplifying in many respects their possible use of AI.

Photo by Thomas Jensen on Unsplash

Fourth, the growing societal dependency on data and technology. It was noted that society as a whole is increasingly dependent on the integrity and availability of the Internet and the reliability of data for its functioning. With advancements in AI in recent years, there has been a rapid integration of AI into daily life, through smart devices and smart cities, including critical infrastructures like healthcare providers, energy providers, and biological and nuclear facilities. Although this presents many advantages, it equally presents a heightened vulnerability to AI-enabled cyber-attacks or more traditional attacks on AI systems within such infrastructures or the data upon which these systems operate.

While, as this report will describe, there is no clear evidence or indication of the actual direct use of AI by terrorist organizations, it is possible to lean on and learn from trends and developments in the malicious use of AI in other areas – in particular cybercriminals, who have long been early adopters of technology. In this regard, and considering the exponential growth of the AI industry in recent years and the abovedescribed factors, the potential for the malicious use of this technology for terrorist purposes merits the close attention of the international community going forward.

II. WHAT IS AI?

To appreciate how AI can be used, or, as the case may be, how it can be misused for terrorist purposes, it is essential to start by establishing a foundational understanding of the technology itself.

i. Key Terminology and Basic Concepts

AI is a field of computer science dedicated to the theory and development of computer systems that are able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, translation between languages, decision-making, and problem-solving.[9] These intelligent systems can be, for instance, software applications, robots and autonomous cars. AI is an umbrella term comprising many different sub-fields, the most prominent of which are described below.

Photo by Joshua Hoehne on Unsplash

Machine learning is a sub-field of AI that comprises the algorithms that can "learn" from data, i.e., progressively improve performance on a specific task. In contrast with other computer software, machine learning algorithms do not require explicit instructions from humans. Instead, they extract patterns and learn implicit rules from a considerable number of examples included in a database.[10] An AI system may therefore include a machine learning algorithm to perform a certain task, as well as sensors and external devices necessary to execute that task. For example, a computer vision AI system is composed of an image recognition software and one or more cameras to capture the image that the algorithm will process.

Deep learning is, in turn, a sub-field of machine learning that deals with a smaller family of algorithms, known as neural networks. These are algorithms inspired by the human brain that seek to learn from large amounts of data by performing a task repeatedly, each time making minor modifications to its internal features to improve the outcome. The term "deep learning" comes from the several (or "deep") layers of the neural network.[11]

9 Stuart J. Russell and Peter Norvig. (2009). Artificial Intelligence: A Modern Approach (3rd ed.). Prentice Hall.

10 Tom Mitchell. (1997). Machine Learning. McGraw Hill.

11 Ian Goodfellow, Yoshua Bengio and Aaron Courville. (2016). Deep Learning. MIT Press. Accessible at www.deeplearningbook.org

The image below captures the relationship between AI, machine learning, and deep learning.

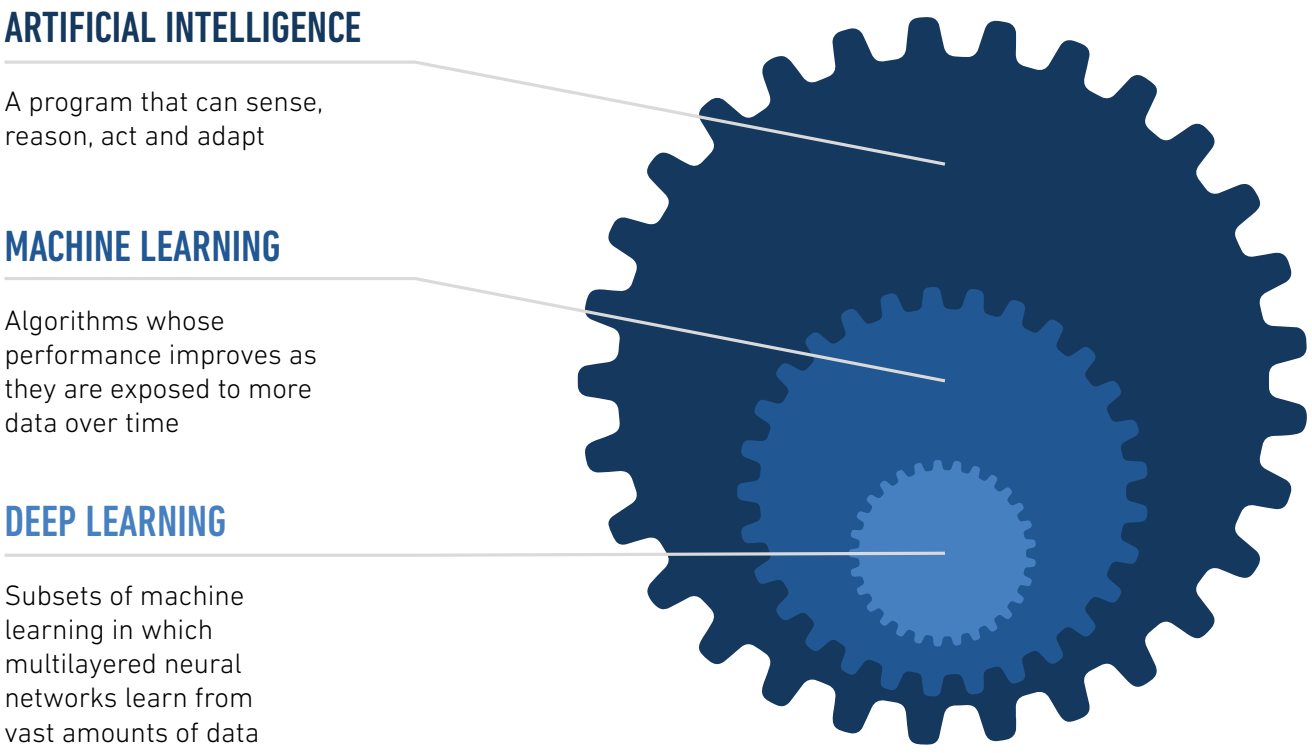


Figure 3. Relationship between AI and notable subfields

The AI systems that exist today consist of what are known as "narrow" AI applications. These are AI systems that are programmed to perform a single task, such as predicting the weather, playing chess, or analyzing medical images. As a result of their "narrow" programming, these systems do not perform well outside of the single task that they are designed to perform. However, by excelling in specific tasks, these systems can act as the building blocks of more intelligent AI systems that could be developed in the near future.

In connection with this, another concept commonly arising in the literature is artificial general intelligence (AGI), which refers to systems that can successfully perform any intellectual task that a human being can. Unlike narrow AI, which is purpose-built to perform a limited task, AGI would be able to learn, plan, reason, communicate in natural language, and integrate all these skills and apply them to any task. AGI has been the holy grail of AI for a long time, with experts extensively discussing *whether* and, if so, *when* AGI will arrive.[12]

Moving beyond AGI is the concept of artificial super intelligence (ASI). This is a concept attributed to machines that will be able to surpass human intelligence in every aspect.[13] From creativity to problem-solving, super-intelligent machines would overcome human intelligence as both individuals and as a society. This type of AI has generated a great amount of philosophical debate, with some experts arguing that it may even present an existential threat to humanity.[14]

12 Hal Hodson. (Mar. 1, 2019). DeepMind and Google: the battle to control artificial intelligence. 1843 Magazine. Accessible at https://www.economist.com/1843/2019/03/01/deepmind-and-google-the-battle-to-control-artificial-intelligence

13 Nick Bostrom. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

14 Rory Cellan-Jones. (Dec. 2, 2014). Stephen Hawking warns artificial intelligence could end mankind. BBC. Accessible at https://www.bbc.com/news/technology-30290540

ii. Algorithms and Applications

There are multiple architectures of neural networks within the deep learning family which enable different applications.

Convolutional Neural Networks (CNNs or ConvNets) are a class of neural networks most often applied to analyze images.[15] Inspired by the animal visual cortex,[16] these algorithms use multiple layers of single units or nodes to progressively extract higher-level features from the raw input. For instance, if the input is an image, the first layers of the neural network may identify lines and curves, while the last layers may identify letters or faces. This characteristic allows CNNs to identify objects,[17] which in turn enables object recognition and subsequently facial recognition.[18]
Another area that has received a lot of attention is natural language processing (NLP). The type of architecture most frequently used in NLP is known as Recurrent Neural Networks (RNNs)[19]. In an RNN, the network nodes are connected along a temporal sequence. Relying on an internal memory that processes sequences of inputs, a machine using an RNN can perform speech recognition by understanding sequences of words and their morphosyntax and semantic function.[20]

Besides speech recognition, NLP can also be used for the generation of text, which forms the basis of "chatbots"[21] – software programmes that operate online and can be programmed to simulate basic conversation. The generation of content such as text and images is possible due to another type of neural network known as Generative Adversarial Networks (GANs). This innovative architecture, invented in 2014, has since revolutionized the deep learning field.[22] GANs models consist of two artificial neural networks: a generative network and a discriminative network. While the generative network creates new data with the same characteristics as the training set, the discriminative network separates generated data from training data. A GAN trained on photographs can, for instance, generate new images that look similar to the dataset of stored pictures. The discriminator randomly receives photographs produced by the generator based on the training dataset and tries to identify them from the original images. The objective of the generator is to "fool" the discriminator network by producing increasingly better novel candidates.[23]

GANs have many applications, such as generating text, images, songs and various forms of art.[24] GANs are also behind the widely publicized and hotly debated phenomenon known as "deepfakes". A portmanteau of "deep learning" and "fake media", deepfakes are a type of synthetic media invented in 2017. They involve the use of AI techniques to manipulate or generate fake visual and audio content that humans or even technological solutions cannot immediately distinguish from authentic content.[25]

15 Ian Goodfellow, Yoshua Bengio and Aaron Courville. (2016). Deep Learning. MIT Press. Accessible at www.deeplearningbook.org

16 David Hubel and Torsten Wiesel. (Oct. 1959). Receptive fields of single neurones in the cat's striate cortex. J. Physiol. 148 (3): 574–91.

17 Dan Ciresan, Ueli Meier, Jonathan Masci, Luca M. Gambardella, Jurgen Schmidhuber. (2011). Flexible, High Performance Convolutional Neural Networks for Image Classification. Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence-Volume Volume Two. 2: 1237–1242.

18 Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, Andrew D. Back. (1997). Face Recognition: A Convolutional Neural Network Approach. IEEE Transactions on Neural Networks. 8 (1): 98–113. CiteSeerX 10.1.1.92.5813

19 Richard Socher, Cliff Lin, Andrew Y Ng, Christopher D Manning. Parsing Natural Scenes and Natural Language with Recursive Neural Networks. 28th International Conference on Machine Learning (ICML 2011).

20 Samuel Dupond. (2019). A thorough review on the current advance of neural network structures. Annual Reviews in Control. 14: 200–230.

21 Barbora Jassova. (Jan. 2, 2020). Natural Language Processing Chatbots: The Layman's Guide. Landbot.

22 Ian J. Goodfellow, et al. (Jun. 10, 2014). arXiv. "Generative Adversarial Networks." Accessed on Jul. 1, 2020, at https://arxiv. org/abs/1406.2661.

23 ibid.

24 Jason Brownlee. (Jul. 12, 2019). Machine Learning Mastery. Impressive Applications of Generative Adversarial Networks (GANs). Accessible at https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/

25 Oscar Schwartz. (Nov. 12, 2018). The Guardian. You thought fake news was bad? Deep fakes are where truth goes to die. Accessible at https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth

Photo by Ilya Pavlov on Unsplash

Coupled with the reach and speed of the Internet, social media and messaging applications, deepfakes can quickly reach millions of people in an extremely short period of time. Because of this, deepfakes have been identified as a powerful weapon in today's disinformation wars, whereby people can no longer rely on what they see or hear.[26]

iii. An Evolving Technology

The estimated market of AI-related products is growing every day as the integration of AI expands into fields such as medicine, economics, communications, insurance, financing and many other fields. AI is used, for instance, in autonomous vehicles, such as drones and self-driving cars, search engines, online smart assistants, spam filtering and targeted online advertising. It is also used to create poetry and other forms of art, prove mathematical theorems, play games like chess or Go, and, somewhat controversially, to inform and predict the outcome of judicial decisions.[27]

With the increasing investment by private industry, academia and governmental entities, machine learning algorithms will certainly continue to become even more sophisticated. Allied with the continuous collection of data by sensors and mobile apps, the data available to train machine algorithms will grow exponentially, leading to considerable progress in the field of AI, that will in turn further push the frontiers of what machines can do.[28]

The development in AI systems will likely allow for huge advances in the intelligence of robotics. Increasingly capable robots or robotic tools with enhanced "senses" and dexterity will be able to take on tasks once thought too delicate or uneconomical to automate. Moreover, with the expansion of capabilities of advanced AI, machines will likely be able to take over not only manual tasks but also cognitive tasks, for instance, in accounting, marketing and human resources.[29]

The whirlwind development and integration of AI in daily activities has already transformed many industries and sectors and this trend is projected to continue. AI is obliging actors in all sectors to rethink how decisions are taken and how everyday tasks are performed. At least for now, it appears that AI is here to stay and the potential of this game-changing technology is available for anyone willing to seize it.

---

26 Than Thi Nguyen, et al. (Jul. 28, 2020). arXiv. Deep Learning for Deepfakes Creation and Detection: A Survey. Accessible at https://arxiv.org/abs/1909.11573

27 Michael Haenlein, Andreas Kaplan. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. California Management Review. 61 (4): 5–14.

28 Xiaomin Mou. (Sept. 2019). Artificial Intelligence: Investment Trends and Selected Industry Uses. International Finance Corporation.

29 James Manyika, Susan Lund, Michael Chui, Jacques Bughin, Jonathan Woetzel, Parul Batra, Ryan Ko, and Saurabh Sanghvi. (2017, Nov 28). Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages. McKinsey & Company. Retrieved from https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages

III. THE THREAT OF ALGORITHMS AND TERRORISM

Although the tactics of terrorism vary from group to group and individual to individual, it can be said that to a certain degree, terrorist organizations tend to be risk averse with their repertoire, favouring the tried and tested effectiveness of weapons such as guns and bombs.[30, 31] Nevertheless, terrorism is undoubtedly not a stagnant threat. Terrorist groups and individuals have shown that they can adapt very well and have evolved considerably over the decades. They have demonstrated the potential to innovate, for instance, in their organizational structure, becoming decentralized, franchised and global. They have also evolved significantly in terms of tactics, moving from irregular guerrilla warfare to indiscriminate attacks.[32, 33]

Technology-wise, this trend to innovate is especially pronounced with respect to the Internet and social media, which have proven to be extremely valuable for terrorists. The Internet and social media, as well as by extension other ecosystems such as the online gaming platform, have become powerful tools for terrorist groups to radicalize, inspire, and incite violence; claim responsibility for attacks; recruit; raise and move funds; buy and transfer weapons; and make tutorials or instruments available to their members.[34, 35] For example, in 2019, the Christchurch shooting in New Zealand was live-streamed by the attacker on Facebook. Although the video was taken down several minutes later, the attack was broadcasted across the globe, amplifying its impact and effects on the victims.[36]

The scope of this growing phenomenon can be seen from efforts such as Europol's Referral Action Day. As part of 2020 Referral Action Day, Europol and 17 countries identified and assessed for removal as many as 1,906 URLs linking to terrorist content on 180 platforms and websites in only one day.[37] Over the course of two years, Facebook itself has removed more than 26 million pieces of content from groups such as the Islamic State of Iraq and the Levant (ISIL) and Al-Qaida and, in the first three months of 2020, it removed approximately 4.7 million pieces of content connected to "organized hate", including an increase of over 3 million pieces of content from the last quarter of 2019.[38, 39]

30 United Nation Office on Drugs & Crime. Terrorism and Conventional Weapons. Accessible at https://www.unodc.org/images/odccp/terrorism_weapons_conventional.html

31 Bruce Hoffman. (1994). Responding to Terrorism Across the Technological Spectrum. RAND Corporation.

32 Bruce Hoffman. (2017). Inside Terrorism, Third Edition. Columbia University Press.

33 Karlheinz Steinmüller. (2017). The World in 2040. Framework Conditions for New Kinds of Terrorism. In T.J. Gordon et al. (Eds.). Identification of Potential Terrorists and Adversary Planning: Emerging Technologies and New Counter-Terror Strategies.

34 United Nations Security Council Counter-Terrorism Committee and ICT4Peace. (Dec. 2016). Private Sector Engagement in Responding to the Use of the Internet and ICT for Terrorist Purposes: Strengthening Dialogue and Building Trust. United Nations. Accessible at https://ict4peace.org/wp-content/uploads/2016/12/Private-Sector-Engagement-in-Responding-to-the-Use-of-the-Internet-and-ICT-for-Terrorist-Purposes-1.pdf

35 United Nation Office on Drugs & Crime and United Nations Counter-Terrorism Implementation Task Force. (Sep. 2012). The use of the Internet for terrorist purposes. United Nations. Accessible at https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_Internet_for_terrorist_purposes.pdf

36 Tech Against Terrorism. (Mar. 26, 2019). Analysis: New Zealand attack and the terrorist use of the Internet. Tech Against Terrorism. Accessible at https://www.techagainstterrorism.org/2019/03/26/analysis-new-zealand-attack-and-the-terrorist-use-of-the-Internet/

37 Europol. (Jul. 3, 2020). Terrorist "How-to" Guides - Focus of Latest Europol Referral Action Day [Press release]. Europol. Accessible at https://www.europol.europa.eu/newsroom/news/terrorist-%E2%80%98how-to%E2%80%99-guides-focus-of-latest-europol-referral-action-day

38 Facebook. (Sept. 17, 2019). Combating Hate and Extremism. Facebook. Accessible at https://about.fb.com/news/2019/09/combating-hate-and-extremism/

39 Facebook. (May 12, 2020). An Update on Combating Hate and Dangerous Organizations. Accessible at https://about.fb.com/news/2020/05/combating-hate-and-dangerous-organizations/

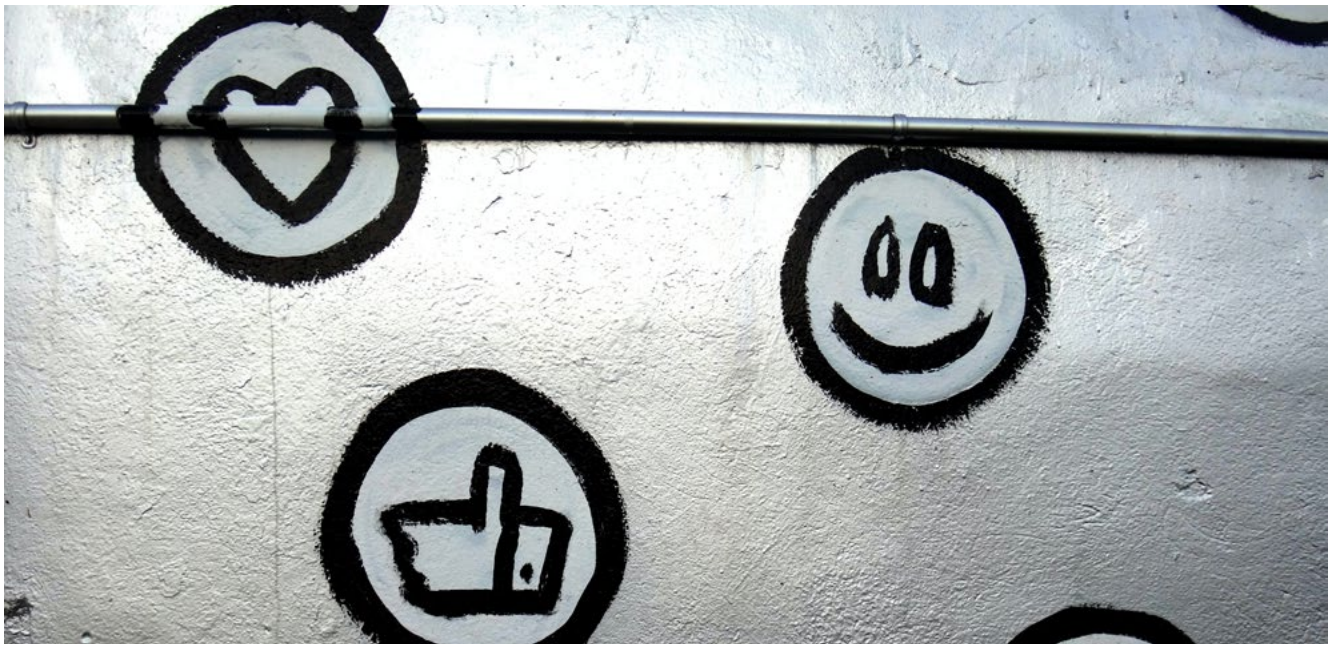


Photo by George Pagan III on Unsplash

The misuse of the Internet and social media is also a domain where the ability of terrorist groups to adapt, in particular to challenges, is quite evident. In response to the efforts of social media platforms and law enforcement agencies to take down terrorist content online, there have been several developments in how terrorists use the Internet and social media, from encrypted communication to other quite innovative methods. For instance, in an attempt to avoid detection, a recent video containing terrorist content uploaded to Facebook included a 30-seconds introduction of the France 24 news channel before the true 49-minute long propaganda video began.[40] Also notably, in May 2020, Hay'at Tahrir al-Sham – formerly Al-Nusrah Front for the People of the Levant – encouraged its members and militant groups in Syria to stop using platforms such as Telegram, Facebook Messenger and Viber and instead use other encrypted applications such as Conversations, Riot, Signal and Wire.[41] Indeed, the rolling out of end-to-end encryption (E2EE) by such platforms have prompted considerable concern amongst policy makers and counter-terrorism practitioners with respect to it enabling the potential for terrorists to "go dark" and communicate securely, thereby evading detection. Numerous investigations and counter-terrorism operations have indicated the use of encryption by both ISIL and Al-Qaida affiliated individuals.[42] While technology has certainly not been the exclusive trigger for the considerable evolution of terrorism, it has played an important role in it. Recent history is replete with examples that demonstrate the increasingly sophisticated use of diverse technologies by terrorist and violent extremist groups.[43] In many ways, this should not come as a surprise. Terrorists, just as every individual, are after all "children of their time".[44] As a collective, they rely on and make use of the tools and instruments that are available to them, just as any ordinary individual would. And indeed, all technology, analogue and digital, progresses to a point where malicious individuals can exploit it to perpetrate crimes.[45]

40 Gordon Corera. (Jul. 13, 2020). ISIS "still evading detection on Facebook", report says. BBC. Accessible at https://www.bbc.com/news/technology-53389657

41 United Nations Security Council. (Jul. 23, 2020). Twenty-sixth report of the Analytical Support and Sanctions Monitoring Team submitted pursuant to resolution 2368 (2017) concerning ISIL (Da'esh), Al-Qaida and associated individuals and entities. United Nations. Accessible at https://www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/s_2020_717.pdf, para. 99.

42 Robert Graham. (June 2016). How Terrorists Use Encryption. The Combatting Terrorism Center Sentinel, Westpoint. Accessible at https://ctc.usma.edu/how-terrorists-use-encryption/

43 It should be noted that references to technology in this section should be read in terms of the broader understanding of the term which includes tools, machines, techniques, crafts, systems, and methods of organization, and not only digital technologies.

44 Renske van der Veer. (2019). Terrorism in the age of technology in Strategic Monitor 2019-2020. The Hague Centre for Strategic Studies and the Clingendael Institute. Accessible at https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terrorism-in-the-age-of-technology/

45 Kuljit Kaur. (Apr. 1, 2007). High Technology Terrorism: A Threat to Global Security. India Quarterly, 63(2), 81-95 Accessible at https://www.jstor.org/stable/26295959

Broadly speaking, the interplay of technology and terrorism is evident in three primary ways. First, terrorists rely upon technology as a weapon to carry out attacks. In fact, the nature of terrorist attacks has changed significantly over time in connection with technological developments.[46] Terrorist arsenals have expanded significantly, from the use of knives and guns to aircraft hijackings and other vehicle-based attacks, with some groups even displaying degrees of intent to acquire and use chemical, biological or radiological materials. The adoption of the automatic rifle is arguably one of the most significant adaptations to technological development. Due to its low cost and lethality, the automatic rifle has become the weapon of choice for terrorist groups in many parts of the world.[47] Second, technological developments in transportation and logistics have transformed the abilities of terrorist and criminal groups in general, allowing them to increase the speed, reach and magnitude of their operations and rendering them global, rather than local, threats.[48] Lastly, developments in information and communication technologies have allowed terrorist groups and individuals to communicate more quickly and covertly over expanding distances and spread viral videos and information to foster terror faster and at a larger scale. In doing so, they have been able to both increase the efficiency and effectiveness of their attacks and reach out to potential recruits.[49, 50, 51] Mobile communication devices, the Internet and, more recently, social media and the dark web are the leading examples of this.

Photo by Niclas Lundin on Unsplash

46 Herbert K. Tillema. (2002). A Brief Theory of Terrorism and Technology. In T. K. Ghosh (Ed.), Science and Technology of Terrorism and Counterterrorism.

47 T. X. Hammes. (Sept. 4, 2020). Terror and Technology From Dynamite To Drones. War on the Rocks. Accessible at https://warontherocks.com/2020/09/terror-and-technology-from-dynamite-to-drones/

48 Herbert K. Tillema. (2002). A Brief Theory of Terrorism and Technology. In T. K. Ghosh (Ed.), Science and Technology of Terrorism and Counterterrorism.

49 ibid.

50 Karlheinz Steinmüller. (2017) The World in 2040. Framework Conditions for New Kinds of Terrorism. In T.J. Gordon et al. (Eds.), Identification of Potential Terrorists and Adversary Planning: Emerging Technologies and New Counter-Terror Strategies.

51 Truls Hallberg Tønnessen. (2017). Islamic State and Technology – A Literature Review. Perspectives on Terrorism, 11(6), 101-111. Accessible at https://www.jstor.org/stable/26295959

Recent examples of the terrorist use of technology include an array of advanced devices. For instance, global positioning system (GPS) devices, mobile phones, and the Internet were utilized by the perpetrators of the Mumbai attacks in 2008 to plan, coordinate and carry-out their mission.[52] Although from a contemporary setting this may no longer appear particularly ground-breaking, at that time it marked an innovative use of the latest technological advancements. More recently, Blockchain-based virtual assets, such as "Bitcoin", as well as mobile banking, and crowdfunding have been used by terrorists for fundraising purposes or to move funds,[53] while the dark web serves as a marketplace for materials, weapons and fake documents.[54]

Evidence does however suggest that, in their use of technology, lone terrorist actors in particular tend towards readily and publicly available technologies for communications, weapons and transportation purposes.[55] Equipment requiring low levels of technicality that can be acquired from do-it-yourself stores appear to be preferred. In such forms of "low-tech terrorism", terrorists, especially lone actors, seek ways to transform everyday tools and vehicles, such as kitchen knives, cars, and trucks into weapons.[56]

Notwithstanding this, with each passing day, cutting edge technology, once limited to specialized communities, becomes increasingly accessible to the general public.[57] A prime example is that, whereas a decade ago, a terrorist group manoeuvring a fleet or "swarm" of drones carrying an explosive payload would have been deemed unrealistic, such a scenario is a feasible threat today.[58]

This expansion in what is technologically possible can result in law enforcement, counter-terrorism units and other security forces being caught "off-guard" by innovative terrorist groups and individuals that have identified new and unforeseen ways and means, using affordable and commercial technologies for malicious purposes.[59]

Considering this, and reflecting on recent trends, developments and the potential in the field of AI, including computer vision, NLP and so on, the question, therefore is whether, – or perhaps better, *when* – AI will become another instrument in the toolbox of terrorism.

52 Jeremy Kahn. (Dec. 8, 2008). Mumbai Terrorists Relied on New Technology for Attacks. The New York Times. Accessible at https://www.nytimes.com/2008/12/09/world/asia/09mumbai.html

53 Financial Action Task Force (FATF). (2015). Emerging Terrorist Financing Risks. Accessible at www.fatf-gafi.org/publications/methodsandtrends/documents/emerging-terrorist-financing-risks.html

54 Abeer ElBahrawy, Laura Alessandretti, Leonid Rusnac, Daniel Goldsmith, Alexander Teytelboym & Andrea Baronchelli. (Nov. 2020). Collective dynamics of dark web marketplaces. Scientific Reports, 10, Article 18827. Accessible at https://doi.org/10.1038/s41598-020-74416-y

55 Herbert K. Tillema. (2002). A Brief Theory of Terrorism and Technology. In T. K. Ghosh (Ed.), Science and Technology of Terrorism and Counterterrorism.

56 Truls Hallberg Tønnessen. (2017). Islamic State and Technology – A Literature Review. Perspectives on Terrorism, 11(6), 101-111. Terrorism Research Initiative. Accessible at https://www.jstor.org/stable/26295959

57 Karlheinz Steinmüller. (2017). The World in 2040. Framework Conditions for New Kinds of Terrorism. In T.J. Gordon et al. (Eds.), Identification of Potential Terrorists and Adversary Planning: Emerging Technologies and New Counter-Terror Strategies.

58 Truls Hallberg Tønnessen. (2017). Islamic State and Technology – A Literature Review. Perspectives on Terrorism, 11(6), 101-111. Terrorism Research Initiative. Accessible at https://www.jstor.org/stable/26295959

59 T. X. Hammes. (Sept. 4, 2020). Terror and Technology From Dynamite To Drones. War on the Rocks. Accessible at https://warontherocks.com/2020/09/terror-and-technology-from-dynamite-to-drones/

IV. CLASSIFYING AI THREAT TYPES

AI can pose numerous novel challenges in different contexts for individuals, organizations and states. These challenges arise at the different stages of AI's lifecycle – from design to deployment –[60] and can stem from both intended and unintended actions.

Chief among the concerns around the use of AI by legitimate actors is the very real and serious potential for this technology to infringe upon human rights. When AI technology is not used appropriately, it can threaten, for instance, the rights to privacy, equality, including gender equality, and non-discrimination. The violation of rights can result from an unjustifiable or disproportionate use of AI, or it can be unintentional, for instance, through the use of unconsciously biased data to train machine learning algorithms, resulting in unfair decisions discriminating against individuals, groups or communities on prohibited grounds.[61, 62]

The term "malicious use of AI" is generally reserved to acts that engender harmful consequences by intention.[63] In 2018, a group of eminent authors from diverse disciplines and organizations – Including Oxford University's Future of Humanity Institute, Cambridge University's Centre for the Study of Existential Risk, OpenAI, the Electronic Frontier Foundation, and the Center for a New American Security –, examined the malicious use of AI by states, criminals, and terrorists. Their report, titled "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation", estimated a rapid growth in malicious use of the technology over the coming decade. The authors considered that the malicious use of AI posed threats in terms of cybersecurity, physical security, and political security:[64]

Cyber threats: Cyber threats are a growing area of concern considering the inherent vulnerabilities in cyberspace and the asymmetrical nature of the threats posed by cyber-attacks. From a terrorism perspective, threats include phishing, man-in-the-middle, ransomware and DDoS attacks, as well as defacement of websites. Additionally, there is a growing concern over the misuse of information and communication technologies by terrorists, particularly the Internet and social media, to commit, incite, recruit for, fund or plan terrorist acts. As will be explained in detail in the following chapters, terrorists can leverage AI systems, for instance, to increase the potency and effectiveness of the conventional cyber-attacks or compromise the security of information by infringing its confidentiality or attacking its integrity and availability.

Physical threats: Over the past decade, everyday life has become increasingly interconnected through technology. This interconnectedness is reflected in the emergence of the concept of the Internet of Things (IoT) – an ecosystem of connected digital devices and physical objects that transfer data through the Internet. In this connected world, drones have started making deliveries and autonomous vehicles are already taking to the roads. At the same time, with the integration of these technologies and connected devices into daily life, novel challenges for humans and infrastructures arise. Interconnectivity and increasingly autonomous devices and robots in smart cities or home environments expand the opportunities for and scale of possible attacks.

60 Patrick Bradley. (2020). Risk management standards and the active management of malicious intent in artificial superintelligence. AI & Society, 35(2), 319-328. Accessible at https://doi.org/10.1007/s00146-019-00890-2

61 Nancy G. Leveson, Clark S. Turner. (1993). An investigation of the Therac-25 accidents. Computer, 26(7), 18–41. Accessible at https://doi.org/10.1109/MC.1993.274940

62 Ben Shneiderman. (2016). Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight. Proceedings of the National Academy of Sciences, 113(48),13538-13540. Accessible at https://doi.org/10.1073/pnas.1618211113

63 Patrick Bradley. (2020). Risk management standards and the active management of malicious intent in artificial superintelligence. AI & Society, 35(2), 319-328. Accessible at https://doi.org/10.1007/s00146-019-00890-2

64 Miles Brundage, Shahar Avin et al. (Feb. 2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Accessible at https://maliciousaireport.com/

Political threats: With advancements in information and communication technologies and global prominence of social media, how, when and why individuals communicate and find news sources, are inevitably undergoing an unprecedented change. This transformation can be seen all around the world and has influenced the outcome of elections, encouraged popular protests, and empowered people to exercise their fundamental rights. At the same time, the prominence of social media can equally make people vulnerable to manipulation through misinformation and disinformation and has increased the capabilities of both public and private entities to conduct profiling and surveillance operations. The integration of AI into this equation, for instance through the proliferation of deepfakes, will greatly enhance the nature of this threat.

As the authors of the 2018 report noted, these categories are not necessarily mutually exclusive. For example, AI-enabled hacking can be directed at cyber-physical systems resulting in physical harm, and physical or digital attacks could be carried out for political purposes. Moreover, "political" is a complex categorization, particularly in the context of terrorism, for which a political motivation is often very much linked to the general understanding of the concept of terrorism, along with social, ideological, religious and economic factors.

In this regard, for the purposes of this report, the malicious use of AI for terrorist purposes will consider two primary types of threats, namely *cyber threats* and *physical threats*, as well as add to the discussion other relevant activities connected with the actions of terrorist groups and individuals, including financing methods, propaganda and disinformation strategies and other operational tactics.

V. FACT OR SCIENCE FICTION?

Having examined some categories of threats posed by the malicious use of AI, this chapter will address whether there is any credible substance to such threats or if the malicious use of AI for terrorist purposes is little more than science fiction.

From the outset, it is important to clarify that no clear evidence of the actual use of AI by terrorist organizations has been identified to date. In fact, in its most recent report the ISIL/Al-Qaida Monitoring Team observed that "Notwithstanding continuing Member State concerns about abuse of technology by terrorists, especially in the fields of finance, weaponry and social media, neither ISIL nor Al-Qaida is assessed to have made significant progress in this regard in late 2020."[65]

There are, however, important caveats to this. AI, as has already been seen, is very much already part of daily life and is used by many individuals, often unbeknownst to them. For instance, NLP is the basis of smart assistants such as Apple's Siri and Amazon's Alexa, and is used to correct typos in text messages, emails and Word documents. Facial recognition is used to unlock smartphones and object recognition helps to classify images and improve the results of Google searches. In this regard, the above statement does not presume to exclude the possibility of terrorist groups and individuals having used AI *indirectly* – for instance, passively, or even unwittingly, as described above. Rather, it is intended to imply that AI has not been used *directly*, for instance, to specifically improve or amplify an attack.

65 Analytical Support and Sanctions Monitoring Team submitted pursuant to resolution 2368 (2017) concerning ISIL (Da'esh), Al-Qaida and associated individuals and entities, Twenty-seventh report, S/2021/68 (3 February 2021).

Photo by Nathan Dumlao on Unsplash

The lack of evidence of the direct use of AI in terrorism should also not be interpreted as indicating that terrorists are indifferent or disinterested in the technology. Although concrete evidence of terrorists' interest or intention to use AI has not been found, it is prudent to assume that these groups and individuals are aware of this technology, which has been so often lauded by many for its revolutionary potential. It is, for instance, noteworthy that, in 2016, a video seemingly prepared by ISIL in Syria showed the group experimenting with a rudimentary version of a self-driving car, which in this case was remote-controlled.[66] The car in question had mannequins placed inside the vehicle to fool observers into thinking a person was driving. It is believed that ISIL also had plans to attempt to replicate the heat signature of a human in an attempt to further fool security systems into believing that someone was actually inside the vehicle.[67] Not long thereafter, the Chief Research Officer at F-Secure indicated that there was evidence that ISIL was indeed working to develop self-driving cars to use in place of suicide bombers[68] and, in 2018, British prosecutors revealed that two ISIL supporters were planning to involve the use of a driverless car in the perpetration of terrorist attacks.[69, 70]

More recently, in March 2020, an ISIL supporter circulated a video on Rocket.chat – a decentralized social media platform that has been used by ISIL to spread terrorist content and facilitate online collaboration and coordination – explaining how facial recognition software could be used.[71] The video in question purported that facial recognition could identify individuals on the basis of their facial features, even if they had sought to obscure their identity by using a facial covering or digitally blurring their face. The video further claimed that this capability would most certainly aid authorities in foiling terrorist plots and apprehending perpetrators. While the current capabilities of this technology

66 Ronan Glon. (Jan. 13, 2016). ISIS is testing a deadly remote-controlled car bomb that fools infrared sensors. Digital Trends. Accessible at https://www.digitaltrends.com/cars/isis-remote-contrlled-car-bomb-news-demonstration/

67 Stephen Edelstein. (Mar. 16, 2016). ISIS progressing with work on driverless car bombs, security analyst says. Digital Trends. Accessible at https://www.digitaltrends.com/cars/isis-autonomous-car-bombs/

68 Pete Bigelow. (Mar. 15, 2016). ISIS could use a self-driving car to deliver a bomb. Autoblog. Accessible at https://www.autoblog.com/2016/03/15/isis-terrorists-bomb-self-driving-cars-sxsw/?guccounter=2

69 Telegraph Reporters. (Sept. 4, 2018). "Isil-supporters" accused of plotting terror attack using driverless car bomb to spare their own lives. The Telegraph. Accessible at https://www.telegraph.co.uk/news/2018/09/04/isil-supporters-accused-plotting-terror-attack-using-driverless/

70 Kelly Lin. (May 2, 2016). ISIS Working on Weaponizing Self-Driving Cars, NATO Expert Warns. MotorTrend. Accessible at https://www.motortrend.com/news/isis-working-on-weaponizing-self-driving-cars-nato-expert-warns/

71 Memri – Cyber & Jihad Lab. (Mar. 31, 2020). ISIS Supporter Shares Video on Rocket.Chat Demonstrating Abilities Of Facial Recognition Software. Memri. Accessible at https://www.memri.org/cjlab/isis-supporter-shares-video-rocketchat-demonstrating-abilities-facial-recognition-software#_ednref1

might have been exaggerated in the video, the acknowledgment of its existence and the fact that the video was quickly shared on several other channels, confirm that terrorist groups are aware of the potential of AI and are following its trends and developments, at least at a superficial level.

While terrorist groups have not been seen to use AI directly, there is considerable evidence of the exploitation of AI-related technologies by these groups. This is observed, in particular, in the exploitation of unmanned aerial systems, also known as "drones". Drones are considered an AI-related technology for the purposes of this report in that, even if they are manually operated, they can have varying degrees of autonomy. For instance, drones can already be equipped with Global Navigation Satellite System (GNSS)-supported flight stabilization and "sense and avoid" features, and AI could be leveraged to provide even greater degrees of autonomy.[72]

The use of such technology by terrorist groups, as well as other non-State actors, is not a novel development. Such groups have been seen to have experimented with drones, or remotely-controlled aircrafts, for several years, dating as far back as unused plans from Aum Shinrikyo – the Japanese cult behind the Tokyo subway sarin attack in 1995.[73]

The nature of drone usage by such groups has been varied and includes actual and attempted attacks, disruption, surveillance and propaganda.[74] Additionally, it has further been suggested that drones could be used to carry out intelligence, surveillance and reconnaissance missions; monitor targets, security protocols, and patterns of behaviour; increase the accuracy of indirect fire; collect footage to use in propaganda materials; disrupt law enforcement operations; disrupt, interfere with or paralyze key infrastructures, air traffic and economic assets; smuggle illicit goods across borders or into sensitive areas; intimidate and harass; and incite panic in mass gatherings.[75]

Photo by Shutterbouy Photography on Unsplash

72 For more on AI-enable drones, please refer to Chapter eight below.

73 In the early 1990s, Aum Shinkrikyo purchased two remote control drones with spray attachments as part of an unused plot: Amy E. Smithson. (2000). Chapter 3: Rethinking the Lessons of Tokyo. In Amy E. Smithson
And Leslie-Anne Levy, (Eds.), Ataxia: The Chemical and Biological Terrorism Threat and the US Response. Accessible at https://www.stimson.org/wp-content/files/file-attachments/atxchapter3_1.pdf

74 United Nations Counter-Terrorism Committee Executive Directorate (May 2019). Greater Efforts Needed to Address The Potential Risks Posed By Terrorist Use Of Unmanned Aircraft System. CTED Trends Alert. Accessible at https://www.un.org/sc/ctc/wp-content/uploads/2019/05/CTED-UAS-Trends-Alert-Final_17_May_2019.pdf

75 Global Counterterrorism Forum (GCTF) Initiative to Counter Unmanned Aerial System Threats. (Sept. 2019). Berlin Memorandum on Good Practices for Countering Terrorist Use of Unmanned Aerial Systems. Tenth GCTF Ministerial Plenary Meeting. Accessible at https://www.thegctf.org/Portals/1/Documents/Framework%20Documents/2019/Berlin%20Memorandum%20EN.pdf?ver=2020-01-13-143548-187

Notably, there is evidence that ISIL has been employing drones since approximately 2016. ISIL has reportedly even formed an "Unmanned Aircraft of the Mujahedeen" unit, which is responsible for the development and usage of drones.[76] In what is believed to be their first use of the technology, ISIL members deployed drones loaded with explosives in attacks in northern Iraq, killing two Kurdish peshmerga fighters and wounding two troops from the French Special Operations.[77] In 2017, ISIL boasted the success of its drone initiatives, alleging that their drone attacks had killed or wounded 39 soldiers in one week.[78] The terrorist group has also distributed guidance online to its supporters on the use of drones and released propaganda materials, calling for attacks involving drones.[79]

Other non-State actors are also known to have used drones.[80] Two high-profile drone attacks are worth mentioning: the August 2018 assassination attempt of the President of Venezuela Nicolas Maduro and the September 2019 attacks on Saudi Aramco oil processing facilities in Abqaiq and Khurais in Saudi Arabia.[81] Both incidents involved drones with explosive payloads. The latter is particularly noteworthy in that it involved a swarm of as many as 25 drones operating in unison.

The key factors behind the growing interest in the use of this technology for malicious purposes are the drones' commercial availability, affordability and convenience, coupled with the challenges of countering their use. The potential dramatic effect of the use of drones in a terrorist attack is also a further factor that should be considered when seeking to understand their appeal to terrorist groups and individuals.

There has accordingly been heightened concern regarding the use of drones by terrorist groups and other non-State actors, in particular in terms of the possible use of swarms.[82] In urging Member States to take greater collective effort to prevent terrorists from acquiring weapons, the United Nations Security Council through Resolution 2370 (2017), strongly condemned flow of drones to and between ISIL, Al-Qaida, their affiliates, and associated groups, illegal armed groups and criminals.[83] The Global Counterterrorism Forum (GCTF) has also acknowledged that drones constitute a growing area of concern and, in this regard, released the Berlin Memorandum on Good Practices for Countering Terrorist Use of Unmanned Aerial Systems in 2019. The memorandum provides guidance for the identification, development and refinement of policies, practices, guidelines, regulations, programmes, and approaches for countering the terrorist use of drones.[84]

Notwithstanding this, the current use of drone technology by terrorist organizations remains a rarity, is rather unsophisticated in nature and is heavily dependent on human control. While drones may leverage AI for increased autonomy, there is limited evidence at this stage that any terrorist or other non-State actor have used or sought to use AI-enabled drones.

76 Joby Warrick and Jason Aldag. (Feb. 21, 2017). Use of Weaponized Drones by ISIS Spurs Terrorism Fears. Washington Post. Accessible at https://www.washingtonpost.com/world/national-security/use-of-weaponized-drones-by-isis-spurs-terrorism-fears/2017/02/21/9d-83d51e-f382-11e6-8d72-263470bf0401_story.html

77 Thomas Gibbons-Neff. (Oct. 11, 2016). ISIS Used an Armed Drone to Kill Two Kurdish Fighters and Wound French Troops Report Says. Washington Post. Accessible at https://www.washingtonpost.com/news/checkpoint/wp/2016/10/11/isis-used-an-armed-drone-to-kill-two-kurdish-fighters-and-wound-french-troops-report-says/

78 Joby Warrick and Jason Aldag. (Feb. 21, 2017). Use of Weaponized Drones by ISIS Spurs Terrorism Fears. Washington Post. Accessible at https://www.washingtonpost.com/world/national-security/use-of-weaponized-drones-by-isis-spurs-terrorism-fears/2017/02/21/9d-83d51e-f382-11e6-8d72-263470bf0401_story.html

79 Steven Stalinsky and R. Sosnow. (Feb. 21, 2017). A Decade of Jihadi Organizations' Use Of Drones – From Early Experiments By Hizbullah, Hamas, And Al-Qaeda To Emerging National Security Crisis For The West As ISIS Launches First Attack Drones. Memri. Accessible at https://www.memri.org/reports/decade-jihadi-organizations-use-drones-%E2%80%93-early-experiments-hizbullah-hamas-and-al-qaeda#ISIS%20Anchor

80 Robert J. Bunker. (Aug. 2015). Terrorist and Insurgent Unmanned Aerial Vehicles: Use, Potentials, And Military Implications. Strategic Studies Institute and U.S. Army War College Press. Accessible at https://www.hsdl.org/?view&did=786817

81 Jacob Ware. (Sept. 24, 2019). Terrorist Groups, Artificial Intelligence, And Killer Drones. War on the Rocks. Accessible at https://warontherocks.com/2019/09/terrorist-groups-artificial-intelligence-and-killer-drones/

82 Renske van der Veer. (2019). Terrorism in the age of technology in Strategic Monitor 2019-2020. The Hague Centre for Strategic Studies and the Clingendael Institute. Accessible at https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terrorism-in-the-age-of-technology/

83 United Nations Security Council. (Aug. 2, 2017). Resolution 2370 (2017) Adopted by the Security Council at its 8017th meeting, on 2 August 2017. Accessible at https://undocs.org/S/RES/2370(2017)

84 GCTF Initiative to Counter Unmanned Aerial System Threats. (Sept. 2019). Berlin Memorandum on Good Practices for Countering Terrorist Use of Unmanned Aerial Systems. Tenth GCTF Ministerial Plenary Meeting, New Accessible at https://www.thegctf.org/Portals/1/Documents/Framework%20Documents/2019/Berlin%20Memorandum%20EN.pdf?ver=2020-01-13-143548-187

VI. AI-ENABLED TERRORISM THROUGH THE LOOKING GLASS

The previous chapters indicated that terrorist groups and individuals have repeatedly demonstrated the ability to innovate and adapt to new and emerging technologies, such as GPS, mobile phones and, more recently, drones. Indeed, understanding and accepting that terrorism is an evolving threat is vital to ensure the ability of the international community to prevent and counter terrorism. A failure of imagination can have deadly consequences.

With this in mind, this chapter will extrapolate some potential threats involving terrorism and AI that may be on, or just over, the horizon by presenting a selection of potential malicious uses of AI by terrorist organizations. These malicious uses are inspired by trends and developments in the field of AI, the existing *modus operandi* of terrorist groups and individuals, and currently known criminal uses of AI.[85]

The list of potential malicious uses of AI included in this chapter is not an exhaustive summary of how terrorist organizations might employ AI, nor is it intended in any way to indicate the likelihood of any such scenario occurring. Instead, it aims to stimulate thought and inform conversation on how terrorist groups and individuals might further innovate using AI technology, therefore building knowledge and improving the understanding of the issue by national and international entities responsible for countering terrorism. In the absence of evidence, only through speculation can adequate levels of preparedness be ensured.

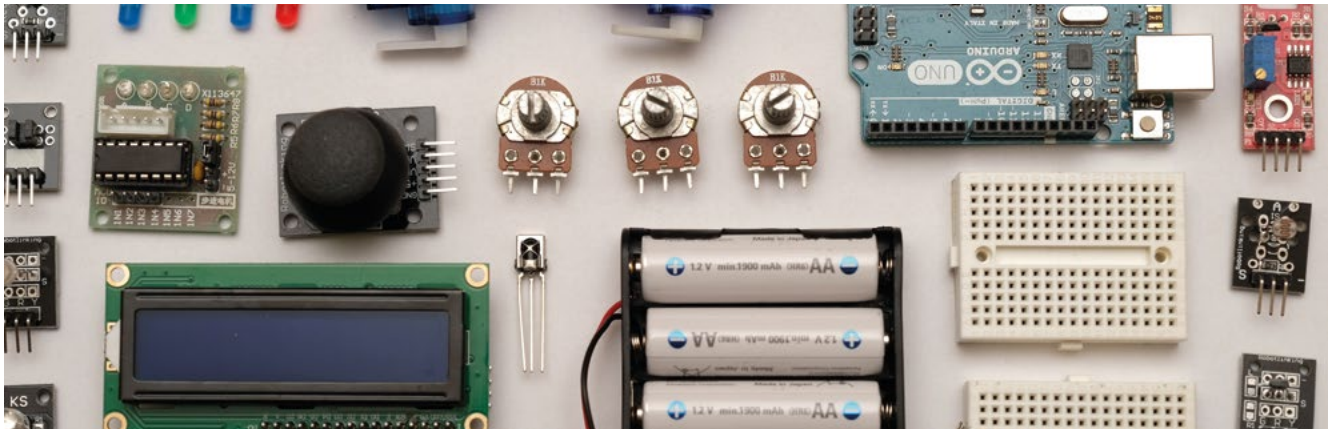


Photo by Robin Glauser on Unsplash

These malicious uses have been characterized and grouped according to their purposes, namely: enhancing cyber capabilities, enabling physical attacks, facilitating the financing of terrorism, spreading propaganda and disinformation, and other operational tactics. Some of the malicious uses presented below may, however, have a dual purpose or function. For instance, ransomware may be part of a cyber-attack and also enable terrorist financing. For this reason, the categorization used in this chapter should be interpreted fluidly.

Finally, before proceeding further with the review of these potential malicious uses, it merits noting that, while advancements in AI are likely to enhance certain capabilities of malicious actors, the national authorities and private entities tasked with ensuring the security of platforms or systems against attack are not stagnant. Such entities are equally evolving and adapting to the latest technological trends, developments and breakthroughs. Thus, while AI may amplify existing threats or present novel threats from the perspective of terrorism, it is important to bear in mind that it will also improve or provide new capabilities to prevent or mitigate potential threats. This is particularly the case with respect to cybersecurity, where AI is already playing a considerable role.[86]

85 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philipp Amann, Aglika Klayn, Odhran McCarthy and Maria Eira. (Nov. 19, 2020). Malicious Uses and Abuses of Artificial Intelligence. Trend Micro Research. Accessible at http://unicri.it/sites/default/files/2020-11/AI%20MLC.pdf

86 Kaspersky. AI and Machine Learning in Cybersecurity — How They Will Shape the Future. Kaspersky. Accessible at https://www.kaspersky.com/resource-center/definitions/ai-cybersecurity

i. Enhancing Cyber Capabilities

a. Denial-of-Service Attacks

The denial-of-service (DoS), or distributed denial-of-service (DDoS), attacks have been among the most popular cyber-attacks for decades.[87] The ultimate purpose of these attacks is to render a computer system connected to the Internet temporarily unavailable for its users by completely exhausting its memory through multiple connection requests.[88] With DDoS attacks, the attackers use more than one and often thousands of machines, in what are known as "botnets", to direct the requests at the target system.[89]

It is believed that between late 2016 and early 2017, ISIL launched its first-ever successful series of DDoS attacks, following discussion among its members on the possibility of conducting such attacks in a top-tier ISIL Dark Web forum.[90] These attacks used a DDoS tool named "Caliphate Cannon" and targeted, primarily, military, economic and education infrastructure clearly demonstrating the seriousness of this threat. Since then, the ISIL hacking division has also claimed responsibility for similar attacks disrupting online services.

Part of what makes DoS or DDoS attacks appealing to cybercriminals, terrorists and other malicious actors is that they can be launched with very little effort and their performance is relatively straightforward.[91] Launching such an attack does not require the attacker to target specific vulnerabilities – the fact that the targeted system is connected to the Internet is generally sufficient.[92] Machine learning is poised, however, to take the ease and simplicity of DoS or DDoS attacks to the next level by automating processes that are traditionally performed by the attacker. For instance, machine learning algorithms can be used to control the botnets behind the attack or enable them to identify vulnerable systems through sophisticated network reconnaissance.

The potential of leveraging machine learning in DDoS attacks is already being explored by malicious actors. For instance, in 2018, TaskRabbit – an online marketplace for freelance laborers – was the target of a DDoS attack carried out by a hacker using a botnet controlled by an AI software. This attack affected 3.75 million users of the website, which suffered a significant data breach.[93]

Interestingly, in 2019, Link11 reported that nearly half of DDoS attacks are now carried out by using cloud services such as Amazon Web Services, Microsoft Azure and Google Cloud.[94] Leveraging the computing power provided by such services, malicious virtual machines can be created using machine learning, which are then used as part of a botnet to launch the DDoS attack.[95]

87 History of DDoS Attacks. (Mar. 13, 2017). Radware. Accessible at https://security.radware.com/ddos-knowledge-center/ddos-chronicles/ddos-attacks-history/

88 Christoph L. Schuba, et al. (1997). Analysis of a Denial of Service Attack on TCP in Proceedings of the 1997 IEEE Symposium on Security and Privacy (Cat. No. 97CB36097). Accessible at https://doi.org/10.1109/SECPRI.1997.601338

89 John Ioannidis and Steven M. Bellovin. (2002). Implementing Pushback: Router-based Defense Against DDoS Attacks. Accessible at https://www.cs.columbia.edu/~smb/papers/pushback-impl.pdf

90 Cyber Jihadists Dabble in DDoS: Assessing the Threat. (Jul. 13, 2017). Flash Point. Accessible at https://www.flashpoint-intel.com/blog/cyber-jihadists-ddos/

91 Christoph L. Schuba, et al. (1997). Analysis of a Denial of Service Attack on TCP, in Proceedings, IEEE Symposium on Security and Privacy (Cat. No. 97CB36097). Accessible at https://doi.org/10.1109/SECPRI.1997.601338

92 John Ioannidis and Steven M. Bellovin. (2002). Implementing Pushback: Router-based Defense Against DDoS Attacks.

93 Sam Bocetta. (Mar 10, 2020). Has an AI Cyber Attack Happened Yet? InfoQ. Accessible at https://www.infoq.com/articles/ai-cyber-attacks/

94 *AI vs AI: Artificial Intelligence and the DDoS Attack.* (n.d.) Verdict-AI. Accessible at https://verdict-ai.nridigital.com/verdict_ai_winter19/artificial_intelligence_ddos_attack

95 AI vs AI: Artificial Intelligence and the DDoS Attack. (n.d.) Verdict-AI. Accessible at https://verdict-ai.nridigital.com/verdict_ai_winter19/artificial_intelligence_ddos_attack

b. Malware

Malware, or *mal*-icious soft-*ware* [emphasis added], refers to a wide range of software that intrudes into a computer system or network, bringing it down and harming, exploiting or disrupting the target. Some examples of malware are spyware, ransomware, viruses, worms, trojan horses, and adware. Malware has long been utilized by vandals, swindlers, blackmailers, and other criminals, and is also an obvious instrument for use by terrorist groups and individuals.[96] Malware can be used, for instance, to enable the access of malicious actors to a website, server or network to obtain credit card or other confidential information, or damage the cyber-infrastructure of public or private institutions.[97]

Advancements in AI, and particularly in machine learning, are finding enormous applications in combatting cybersecurity threats such as malware and enabling specialists to analyze data from past attacks and use it to detect anomalies and fend off potential threats.

However, at the same time, AI can also be exploited by malware developers. For instance, AI could be used to automate attack processes, improve the efficacy of malware attacks or even create entirely new forms of malware. It could hypothetically even be used to write code for entirely novel forms of malware. In fact, cybercriminals have already used AI to create polymorphic malware – a type of smart malware that adapts and changes in order to avoid detection,[98] and AI could further be used to enhance this type of malware, speeding up the rate at which it can adapt.

It is also noteworthy that AI can play a major role in enhancing and even automating the distribution of malware. Phishing campaigns are, for instance, one of the primary ways of distributing malware. In a recent experiment involving AI, called "SNAP_R", spear-phishing tweets were delivered to 819 users at a rate of 6.75 tweets per minute, with 275 being successful. The human counterpart in the experiment sent tweets to 129 users at a rate of 1.075 per minute, with only 49 being successful.[99] Using machine learning, those planning an attack can also scan social media to identify the vulnerable targets for a phishing campaign. They can also use machine learning algorithms to analyze emails and the responses received in previous phishing attacks to create content that is more sophisticated and seems more authentic and can, therefore, avoid detection by spam filters and better trick victims into installing the malware.[100]

Another increasingly used way to deliver malware is through Structured Query Language (SQL) injection attacks. SQL injection attacks can be also facilitated by AI. For instance, the AI-enabled DeepHack tool learns how to break into web applications using a neural network.[101] Tools like this can be used to create fully operational automated hacking systems that could function and deliver malware even without the prior knowledge of the target system.

Although a 2019 study by Malwarebytes reported that there was no real-word evidence of AI-enabled malware yet at the time,[102] it is worth noting that IBM researchers presented a new malware that they had developed known as "DeepLocker" at the 2018 Black Hat USA Conference.[103] The DeepLocker malware was developed to demonstrate precisely how AI could enhance malware attacks.[104] DeepLocker masquerades as a video conferencing software and lies in hiding until it identifies its intended victim through facial and voice recognition and then deploys its payload. Instruments like this in the hands of terrorist groups would certainly enhance the severity of the threat of cyber-terrorism.

96 J. P. I. A. G. Charvat. (2009). Cyber Terrorism: A New Dimension in Battlespace. In Christian Czosseck and Kenneth Geers (Eds.), The Virtual Battlefield: Perspectives on Cyber Warfare.

97 UNODC in collaboration with the United Nations Counter-Terrorism Implementation Task Force. The use of the Internet for terrorist purposes. United Nations. Accessible at https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_Internet_for_terrorist_purposes.pdf

98 Adam Kujawa. (Feb. 11, 2020). Real-world AI threats in cybersecurity aren't science fiction. Venture Beat. Accessible at https://venturebeat.com/2020/02/11/real-world-ai-threats-in-cybersecurity-arent-science-fiction/

99 Thomas Brewster. (Jul. 25, 2016). Who's Better at Phishing Twitter, Me or Artificial Intelligence? Forbes. Accessible at https://www.forbes.com/sites/thomasbrewster/2016/07/25/artificial-intelligence-phishing-twitter-bots/#79c3442976e6

100 Adam Kujawa. (Feb. 11, 2020). Real-world AI threats in cybersecurity aren't science fiction. Venture Beat. Accessible at https://venturebeat.com/2020/02/11/real-world-ai-threats-in-cybersecurity-arent-science-fiction/

101 Dan Petro and Ben Morris. (2017). Weaponizing Machine Learning: Humanity was Overrated Anyway. DefCon. Accessible at https://www.defcon.org/html/defcon-25/dc-25-speakers.html#Petro

102 Pieter Arntz, Wendy Zamora, JérÔme Segura and Adam Kujawa. (Jun. 2019). When artificial intelligence goes awry: separating science fiction from fact. Malwarebytes Labs. Accessible at https://resources.malwarebytes.com/files/2019/06/Labs-Report-AI-gone-awry.pdf

103 Dhilung Kirat, Jiyong Jang, and Marc Ph. Stoecklin. (Aug. 9, 2018). DeepLocker — Concealing Targeted Attacks with AI Locksmithing. Black Hat USA. Accessible at https://i.blackhat.com/us-18/Thu-August-9/us-18-Kirat-DeepLocker-Concealing-Targeted-Attacks-with-AI-Locksmithing.pdf

104 Dan Patterson. (Aug. 16, 2018). How weaponized AI creates a new breed of cyber-attacks. TechRepublic. Accessible at https://www.techrepublic.com/article/how-weaponized-ai-creates-a-new-breed-of-cyber-attacks/

c. Ransomware

Ransomware has been repeatedly recognized as one of the top cybersecurity threats globally.[105] A subset of malware, ransomware is a malicious software that encrypts the victims' files and demands the payment of ransom for the files to be decrypted. The threat of ransomware attacks is amplified by the fact that the malware can easily spread to thousands of devices due to its self-replanting functionality. This was exemplified by the 2017 WannaCry ransomware attack, which affected more than 200,000 computers across 150 countries.[106] In 2019, it was estimated that by 2021 there would be a ransomware attack every 11 seconds with a combined annual turnover of approximately $20 billion.[107] More recently, hospitals in the United States reported a 71% increase in the number of ransomware attacks between September and October 2020, threatening an infrastructure already heavily strained by the COVID-19 pandemic.[108]

Photo by engin akyurt on Unsplash

Integrating AI into ransomware could vastly enhance the effects of these attacks. Machine learning models could be used to proliferate the existing ransomware ecosystem by creating new types of attacks or amplifying the effects of the existing ones via intelligent targeting methods. Machine learning algorithms could also be used to improve the efficacy of the phishing campaign to deliver the ransomware as described above. By using AI to perpetrate sophisticated ransomware attacks, an already profitable method of attack could become considerably more so. Such attacks could in turn be used to generate income supporting the infrastructure or activities of terrorist groups and individuals.

105 Internet Organised Crime Threat Assessment (IOCTA). (Oct. 9, 2019). Europol. Accessible at https://www.europol.europa.eu/activities-services/main-reports/internet-organised-crime-threat-assessment-iocta-2019

106 Reuters Staff. (May 14, 2017). Cyber attack hits 200,000 in at least 150 countries: Europol. Reuters. Accessible at https://www.reuters.com/article/us-cyber-attack-europol/cyber-attack-hits-200000-in-at-least-150-countries-europol-idUSKCN18A0FX

107 Luka Arežina. (Nov. 13, 2019). Ransomware statistics in 2020: From random barrages to targeted hits. DataProt. Accessible at https://dataprot.net/statistics/ransomware-statistics/#:~:text=By%20that%20time%2C%20the%20global,billion%20in%20revenue%20for%20cybercriminals.

108 Patrick Howell O"Neill. (Oct. 29, 2020). A wave of ransomware hits US hospitals as coronavirus spikes. Accessible at https://www.technologyreview.com/2020/10/29/1011436/a-wave-of-ransomware-hits-us-hospitals-as-coronavirus-spikes/

Recalling the predilection of terrorist and violent extremist groups for kidnapping for ransom,[109] AI-enabled ransomware attacks that are effective and profitable would appear to be a natural fit in their repertoire as a form of "kidnapping for ransom 2.0".

At the same time, while the main goal of ransomware has traditionally been to extort money from victims, the 2017 NotPetya attack demonstrated that attacks can also be used with destructive or disruptive purposes. In the NotPetya attack, the ransomware was altered so that it was not possible to undo the encryption and revert the system to the status prior to the attack.[110] In this regard, an AI-supported ransomware attack could be used by terrorist groups and individuals for destructive and not just fundraising purposes.

d. Password Guessing

Passwords are the first line of defence against hacking and are essential in all cyber-defence strategies, from large companies to households. Obtaining a password to access protected websites can enable a malicious actor to enter systems or networks to, for instance, disrupt essential services, create disruption, steal valuable data or information, manipulate data or processes, or install malicious software. Supporters of terrorist groups such as ISIL have a long history of hacking websites and social media accounts for the purposes of defacement and disseminating propaganda materials.[111]

In order to enhance security, platforms and websites have introduced numerous measures to protect against password guessing, including requiring longer passwords with a minimum of eight characters, or a combination of alphanumeric and upper- and lower-case characters. Nonetheless, people tend to follow certain patterns when choosing passwords, like combining first names, last names and dates of birth. They also tend to use simple and predictable passwords and to reuse passwords across multiple services. This greatly facilitates the work of a hacker.[112]

Entire databases of passwords stolen from different platforms can be found online for hackers to trawl through, learn from and later use in their attempt to hack websites. Avast, for instance, reported that it has identified 30,160,455,237 stolen passwords online.[113] Password guessing tools, such as "John the Ripper", utilize these databases to guess passwords, but they require extensive work in manual coding to create an attack plan.

Advancements in AI can, however, be leveraged to greatly expedite, enhance and automate the process of password guessing. Malicious actors could train neural networks with these enormous online databases of passwords, which can in turn generate more sophisticated variations of passwords than humans could ever imagine. These neural networks could then run multiple attempts back-to-back until a solution is determined, thereby alleviating the need for the direct involvement of the hacker. In a 2017 study, researchers fed tens of millions of leaked passwords into a neural network tasked with generating new passwords. These passwords were then cross-referenced with leaked passwords from sites such as LinkedIn in order to measure how successful the neural network would be at cracking users' passwords. The study found that they were able to crack 27% of passwords in the LinkedIn set.[114] A subsequent study found that AI can guess a password by detecting what keys are being typed based on shoulder movements analyzed during video calls. The results of the study showed that the AI software in question had an alarming accuracy rate of 75% to 93%.[115]

---

109 United Nations Security Council. (Oct. 29, 2014). Sixteenth report of the Analytical Support and Sanctions Monitoring Team submitted pursuant to resolution 2161 (2014) concerning Al-Qaida and associated individuals and entities (S/2014/770). Accessible at https://www.securitycouncilreport.org/atf/cf/{65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9}/s_2014_770.pdf

110 Andy Greenberg. (Aug. 22, 2018). The Untold Story of NotPetya, the Most Devastating Cyberattack in History. Wired. Accessible at https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/

111 ISIS Sympathizers Defacing and Exploiting WordPress Sites, FBI Warns. (Apr. 13, 2015). TrendMicro. Accessible at https://www.trendmicro.com/vinfo/tr/security/news/cyber-attacks/isis-sympathizers-defacing-and-exploiting-wordpress-sites-fbi-warns

112 J.M. Porup. (Mar. 28, 2018). 1.4B stolen passwords are free for the taking: What we know now. CSO. Accessible at https://www.csoonline.com/article/3266607/1-4b-stolen-passwords-are-free-for-the-taking-what-we-know-now.html

113 Has my password been stolen? (n.d.) Avast. Accessible at https://www.avast.com/hackcheck

114 Matthew Hutson. (Sep. 15, 2017). Artificial intelligence just made guessing your password a whole lot easier. Science. Accessible at https://www.sciencemag.org/news/2017/09/artificial-intelligence-just-made-guessing-your-password-whole-lot-easier

115 Conor Cawley. (Nov. 11, 2020). AI Can Now Guess Your Password by Looking at Your Shoulders. Tech.co. Accessible at https://tech.co/news/ai-guess-password-shoulders

Research has also shown that sophisticated password generation can be performed using GANs that analyze a large dataset of passwords and generate variations similar to the original examples, meaning that it is able to produce billions of guessed passwords beforehand, enabling more targeted and effective password guesses.[116]

Image by TheDigitalWay from Pixabay

While such developments place increased emphasis on having unique and strong passwords, as well as deploying two-factor or multi-factor authentication through, for instance, a mobile device, as an additional layer of defence, it is important to recall that not even the latter provides a complete defence as this has its own vulnerabilities in, particular in terms of social engineering – another area, which, as will be described below, machine learning also stands to enhance.

e. CAPTCHA Breaking

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is another important security measure designed to protect networks and websites from attacks. As the name suggests, CAPTCHA is intended to distinguish real users from spam robots, permitting humans to access and block robots.[117] Websites use CAPTCHA as automated non-human access to web-mail accounts may, for instance, lead to the inflation of spam e-mail messages, and "blog-spammers" may benefit from artificially inflated number of clicks to gain economic advantages.[118] The system works by applying a challenge-response authentication in order to understand whether the access request comes from a human or a computer.[119]

Efforts to break CAPTCHA systems date back to the early days of its introduction, but advancements in machine learning have allowed methods for breaking CAPTCHA systems to become more sophisticated. In 2013, an AI start-up claimed that they succeeded to defeat CAPTCHA systems with its brain-mimicking software with a success rate of more than 90%, without requiring large datasets to train the system or a lot of computing power. The algorithm had been trained to recognize numbers and letters and showed particular efficiency on the CAPTCHAs consisting of letters appearing as if they are made out of fingerprints.[120] Since then, interest in machine learning and deep learning techniques to break CAPTCHA systems are on increase.[121]

116 Briland Hitaj, Paolo Gasti, Giuseppe Ateniese, Fernando Perez-Cruz. (2017). PassGAN: A Deep Learning Approach for Password Guessing. Arxiv. Accessible at https://arxiv.org/pdf/1709.00440.pdf

117 Luis von Ahn, Manuel Blum, Nicholas J. Hopper & John Langford. (2003). CAPTCHA: Using Hard AI Problems for Security. International Conference on the Theory and Applications of Cryptographic Techniques, Berlin, Heidelberg.

118 M. Motoyama et al. (Aug. 2010). Re: CAPTCHAs-Understanding CAPTCHA-Solving Services in an Economic Context, USENIX Security"10: Proceedings of the 19th USENIX conference on Security, 10. Accessible at https://dl.acm.org/doi/10.5555/1929820.1929858

119 Google Workspace Admin Help. What is CAPTCH. Google. Accessible at https://support.google.com/a/answer/1217728?hl=en

120 Rachel Metzarchive. (Oct. 28, 2013). AI Startup Says It Has Defeated Captchas. MIT Technology Review. Accessible at https://www.technologyreview.com/2013/10/28/175597/ai-startup-says-it-has-defeated-captchas/

121 G. Ye et al. (2018). Yet Another Text Captcha Solver: A Generative Adversarial Network Based Approach, in Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security.

Overcoming CAPTCHA systems could significantly enable terrorist groups and individuals to carry out cyberattacks. For instance, it could allow the distribution of automated and large-scale spam e-mails containing malware, terrorist content or other disruptive or propaganda materials.

f. Encryption and Decryption

In a society where essential features of everyday life have become increasingly digitized, encryption is crucial for government agencies, businesses and the general public to safeguard the confidentiality, integrity and accessibility of their communications and stored information. Encryption can be understood as the process of converting data, such as messages or pieces of information, in a way that prevents unauthorized access. Decryption, in turn, is the process of reversing encrypted data back to its initial stage.[122]

As a strong and practical tool to protect against unauthorized access, encryption is also used by those with more nefarious intentions, including criminals and terrorists, as it enables them to communicate and share information securely while preserving their anonymity.[123] Decryption has an equal appeal for such groups or individuals as it can enable them, for instance, to gain access to otherwise confidential information.

In one notable instance in 2012, a French national was sentenced to five years of imprisonment in a case where dozens of encrypted e-mails exchanged between members of terrorist organizations were presented to the court.[124] It was claimed that the terrorist organization was using the encryption software called "Mujahedeen Secrets" in order to facilitate covert online communications among its members.[125] It was also seen that some terrorist organizations benefited from software such as Camouflage and WinZip to mask distributed and shared information through steganography and encryption.[126]

AI-powered encryption tools are being explored at present. In 2016, researchers from Google Brain successfully trained two neural networks to communicate to each other without allowing a third neural network to intercept their messages. The investigation showed that the first two neural networks succeeded to autonomously create their own form of sending encrypted messages back and forth and decrypt these messages in turn.[127] Interestingly, researchers from Columbia University also developed a deep learning technique in 2018 that effectively enabled people to embed sensitive information in seemingly ordinary-looking text. In this way, information such as text, images or Quick Response (QR) codes could be kept hidden from the naked eye in plain sight.[128]

With advancements in AI, encryption and decryption techniques may become even stronger. Relying on AI-powered sophisticated encryption techniques, members of terrorist organizations would be able to communicate among themselves with greater ease and without the integrity of the information being breached. AI-powered decryption techniques would, in turn, allow terrorist organizations to more readily access sensitive encrypted intelligence being communicated by counter-terrorism entities.

---

122 First Report of the Observatory Function on Encryption (2019). EUROPOL & EUROJUST. Accessible at https://www.europol.europa.eu/publications-documents/first-report-of-observatory-function-encryption

123 Robert Graham. (Jun. 2016). How Terrorists Use Encryption. 9, Issue 6. CTC Sentinel, 9(6). Accessible at https://www.ctc.usma.edu/how-terrorists-use-encryption/

124 UNODC in collaboration with the United Nations Counter-Terrorism Implementation Task Force. (Sept. 2012). The use of the Internet for terrorist purposes, United Nations. Accessible at https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_internet_for_terrorist_purposes.pdf

125 UNODC in collaboration with the United Nations Counter-Terrorism Implementation Task Force. (Sept. 2012). The use of the Internet for terrorist purposes. United Nations. Accessible at https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_internet_for_terrorist_purposes.pdf

126 GCTF Initiative to Counter Unmanned Aerial System (UAS) Threats. (Sept. 2019). Berlin Memorandum on Good Practices for Countering Terrorist Use of Unmanned Aerial Systems. Tenth GCTF Ministerial Plenary Meeting, New York. Accessible at https://www.thegctf.org/Portals/1/Documents/Framework%20Documents/2019/Berlin%20Memorandum%20EN.pdf?ver=2020-01-13-143548-187

127 M. Abadi & D. G. Andersen. (2016). Learning to Protect Communications with Adversarial Neural Cryptography, arXiv preprint arXiv:1610.06918.

128 Nvidia. (April 12, 2018). Using AI to Encrypt Messages in Plain Sight. Nvidia. Accessible at https://news.developer.nvidia.com/using-ai-to-encrypt-messages-in-plain-sight/

ii. Enabling Physical Attacks

a. Autonomous Vehicles

Vehicles, particularly cars, vans and trucks, have long been used in terrorist attacks. There are countless high-profile examples of their use. Vehicles have been used, for instance, in deliberate ramming attacks, as seen in the Berlin Christmas market attack in December 2016[129] and in the attack in Barcelona in August 2017.[130] Vehicles have also been used in car-bomb attacks, such as the Kabul ambulance bombing in 2018 that killed 103 people and injured 235 others.[131]

One of the most popularly-known applications of AI is autonomous vehicles – also referred to as self-driving or driverless cars. Autonomous vehicles are considered by many to be a safer, more convenient and more efficient means of transportation for our future. In essence, AI embedded in the vehicle's onboard computer employs deep learning techniques to mimic the decision-making processes of the driver in controlling the actions of the vehicle, steering, acceleration, operation brake, etc.[132] Companies such as Tesla and Google have long advocated for the practical application of this technology, leading the efforts in research, testing and development of autonomous cars. In recent years, countless major car companies have joined them in their efforts to put self-driving cars on the road. In November 2019, Waymo – a subsidiary of Alphabet Inc, the parent company of Google – reached a major milestone, commencing an autonomous taxi service in Phoenix, Arizona, in the United States, that did not involve a safety backup driver.[133]

With the rapid advancement of the technology, combined with significant commercial investment into the autonomous vehicles industry and the accomplishment of numerous milestones in addressing legal and policy challenges, it appears inevitable that AI will indeed eventually transform the driving experience, although it is still unknown precisely when this will happen.[134]

Reflecting on the extensive history of terrorism and vehicles, increased autonomy in cars could very well be an amenable development for terrorist groups, allowing them to effectively carry out one of their most traditional types of attacks remotely, without the need for a follower to sacrifice his or her life or risk being apprehended.[135] Aside from facilitating attacks with fully autonomous vehicle-borne improvised explosive devices, it has also been suggested that self-driving cars could be used in order to cause serious accidents, blocking the roads or cause self-driving carnage.[136] Notwithstanding this, there are grounds to believe that safety features enabling them to detect and avoid a situation such as a collision with pedestrians, engaging the breaking system or setting the vehicle on an alternative course, would frustrate terrorist plots to use such vehicles in this manner. Indeed, as was described above, there have already been some developments surrounding self-driving cars and terrorism, including rudimentary experiments and testimony regarding plans by ISIL supporters that did not materialize.

It is pertinent to note of course that the term "vehicles" does not necessarily imply only wheeled motor vehicles but also includes subsurface vehicles, like submarines, and flying vehicles, such as unmanned aerial systems, commonly referred to as "drones". As previously noted, drones are largely remote-controlled and possess limited degrees of au-

129 Jason Hanna. (Dec. 23, 2016). Berlin Christmas market attack: The victims. CNN. Accessible at https://edition.cnn.com/2016/12/23/europe/berlin-christmas-market-attack-victims/

130 CNN Editorial Research. (Sept. 6, 2020). Terrorist Attacks by Vehicle Fast Facts. CNN. Accessible at https://edition.cnn.com/2017/05/03/world/terrorist-attacks-by-vehicle-fast-facts/index.html

131 James Doubek and Amy Held. (Jan. 27, 2018). At Least 103 Killed, 235 Wounded In Taliban Car Bombing In Kabul. NPR. Accessible at https://www.npr.org/sections/thetwo-way/2018/01/27/581265342/dozens-killed-more-than-100-wounded-in-taliban-car-bombing-in-kabul?t=1598955985489

132 Katie Burke. (May. 7, 2019). How Do Self-Driving Cars Make Decisions? Nvidia. Accessible at https://blogs.nvidia.com/blog/2019/05/07/self-driving-cars-make-decisions/

133 Andrew J. Hawkins. (Dec. 9, 2019). Waymo's Driverless Car: Ghost-Riding In The Back Seat Of A Robot Taxi. The Verge. Accessible at https://www.theverge.com/2019/12/9/21000085/waymo-fully-driverless-car-self-driving-ride-hail-service-phoenix-arizona

134 CB Insights. (Dec. 16, 2020). 40+ Corporations Working On Autonomous Vehicles. CB Insights. Accessible at https://www.cbinsights.com/research/autonomous-driverless-vehicles-corporations-list/

135 Jeffrey W. Lewis. (Sept. 28, 2015). A Smart Bomb in Every Garage? Driverless Cars and the Future of Terrorist Attacks. Smart. Accessible at https://www.start.umd.edu/news/smart-bomb-every-garage-driverless-cars-and-future-terrorist-attacks

136 A. Lima et al. (2016). Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems, Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy.

tonomy, but AI equally presents the possibility for drones to become more or even fully autonomous.[137] Researchers exploring the application of AI in drones have been able to develop autonomous control systems for drone swarms, allowing them to navigate and even carry out acrobatic manoeuvres, such as barrel rolls and flips, without the intervention of the human controller on the ground.[138] Here too, technical advancements open the door to swarms of autonomous drones. From a technical perspective, it is arguable that the development of autonomous drones, or even submersibles, may be even more attainable in the near future than driverless cars, considering the reduced number of variables that drone programmers need to accommodate and the more simplified legal frameworks applicable to drones.

Photo by John Rodenn Castillo on Unsplash

b. Drones with Facial Recognition

The adoption of facial recognition technology has increased dramatically over the past few years, fuelled by the rapid improvement of machine learning. Its commoditization has created new opportunities, from improving authentication to access electronic devices to accelerating airplane boarding and security controls in airports. Moving forward, the adoption of facial recognition may reach new services and eventually even become the preferred authentication means to access services.

In 2017, the Future of Life Institute, a United States-based non-profit research institute and outreach organization, released a video titled "Slaughterbots", in which a swarm of micro drones loaded with several grams of explosives use facial recognition to identify and attack their targets in a kamikaze fashion. Facial recognition technology enabled the controller to programme the drone to autonomously acquire, identify, and engage a selected target by cross-referencing images collected by the drone with images uploaded into an embedded facial recognition database.[139] Although highly dramatized, the video quickly went viral, gaining as many as three million views online, and transforming the possible combination of these technologies into a hot topic. Fortunately, this technology does not exist in an "off-the-shelf" format, although it is not an entirely novel notion nor is mere science fiction. Several commercial drone products do already include limited facial recognition capabilities, although this is limited to specific features such as unlocking flight capabilities and enabling "follow-me" modes. Currently drones to not include facial recognition technology to identify and target individuals during flight.[140] In this regard, the use of facial recognition in drones is much more limited.

137 Vikram Singh Bisen. (Feb. 5, 2020). How AI Based Drone Works: Artificial Intelligence Drone Use Cases. Medium. Accessible at https://medium.com/vsinghbisen/how-ai-based-drone-works-artificial-intelligence-drone-use-cases-7f3d44b8abe3

138 Nick Lavars. (Jun. 23, 2020). AI algorithm enables autonomous drones to do barrel rolls and flips. New Atlas. Accessible at https://newatlas.com/drones/ai-algorithm-autonomous-drones-barrel-rolls-flips/

139 Slaughterbots. (Nov. 13, 2017). Stop Autonomous Weapons. YouTube. Accessible at https://www.youtube.com/watch?v=9CO6M2HsoIA

140 Larry Haller. (Feb. 13, 2020). These 4 Drones Can Recognize Your Face. Drones Globe. Accessible at https://www.dronesglobe.com/guide/face-recognition/

Several law enforcement agencies around the world have already begun experimenting with the combination of these technologies, for instance, to aid in the search for missing and vulnerable individuals or to identify persons of interest in crowded spaces.[141, 142] The ability to locate, track and identify targets in such an automated manner is naturally attractive for law enforcement agencies, but its use has given rise to concerns about mass surveillance and human rights violations, particularly in terms of the right to privacy. In light of several controversial developments surrounding the use of facial recognition by law enforcement, the future of this AI technology is uncertain.[143]

The malicious use of drones by terrorist groups for attacks can be considered a growing, although not yet a major threat. The integration of facial recognition technology would certainly be a game-changer in this regard though, significantly enhancing the level of the threat of drones and enabling very targeted attacks. Although the technology is not readily available, individuals with the right know-how could develop "Slaughterbot" capabilities by combining the different elements themselves in a do-it-yourself fashion.[144]

c. Genetically Targeted Bio-weapons

The COVID-19 pandemic has had a wide range of detrimental effects, from the individual loss of lives to a massive global economic downturn. In July 2020, the United Nations Secretary-General, António Guterres observed that "the pandemic has also highlighted vulnerabilities to new and emerging forms of terrorism, such as misuse of digital technology, cyberattacks and bioterrorism".[145] Indeed, new and emerging technologies, particularly biotechnology combined with AI, may present the opportunity for the development of deadly new strains of pathogens specifically targeted for certain genetic groups, although the technical hurdles for doing it would be substantial.[146] In his report on the activities of the United Nations system in implementing the United Nations Global Counter-Terrorism Strategy, the Secretary-General furthermore specifically identified synthetic biology as an example of a new and emerging technology that can present a risk from the perspective of terrorism[147].

Developments in genetic sequencing technologies allowed researchers from various fields to process more genetic data and extract more genetic information. Although these developments foster scientific development and can significantly improve our quality of life, they raise security concerns. By applying machine learning models, researchers are able to further advance their practices of genetic-based diagnostics and therapeutics.[148]

The majority of the collected genetic materials are stored in genetic databases or biobanks and are circulated among relevant researchers. In other words, the collected genetic material, as well as the information extracted from it are increasingly accessible by what is a largely unregulated research stakeholder community.

141 Ken Macdonald. (Nov. 4, 2019). Police to use AI recognition drones to help find the missing. BBC. Accessible at https://www.bbc.com/news/uk-scotland-50262650

142 Faine Greenwood. (Jul. 8, 2020). Can a Police Drone Recognize Your Face? Slate. Accessible at https://slate.com/technology/2020/07/police-drone-facial-recognition.html

143 Nila Bala, Caleb Watney. (Jun. 20, 2019). What are the proper limits on police use of facial recognition?. Brookings. Accessible at https://www.brookings.edu/blog/techtank/2019/06/20/what-are-the-proper-limits-on-police-use-of-facial-recognition/

144 Renske van der Veer. (2019). Terrorism in the age of technology in Strategic Monitor 2019-2020. The Hague Centre for Strategic Studies and the Clingendael Institute. Accessible at https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terrorism-in-the-age-of-technology/

145 António Guterres. (Jul. 6, 2020). Secretary-General's remarks at the opening of the Virtual Counter-Terrorism Week. United Nations. Accessible at https://www.un.org/sg/en/content/sg/statement/2020-07-06/secretary-generals-remarks-the-opening-of-the-virtual-counter-terrorism-week-united-nations-delivered

146 Vivek Wadhwa (Sept. 11, 2020). The Genetic Engineering Genie Is Out of the Bottle. Foreign Policy. Accessible at https://foreignpolicy.com/2020/09/11/crispr-pandemic-gene-editing-virus/

147 United Nations General Assembly. (February 7, 2020). Report of the Secretary-General on the Activities of the United Nations System in Implementing the United Nations Global Counter-Terrorism Strategy. Accessible at https://undocs.org/pdf?symbol=en/A/74/677

148 S. Sawaya, E. Kennally, D. Nelson & G. Schumacher. (2020). Artificial Intelligence and the Weaponization of Genetic Data. SSRN Electronic Journal.

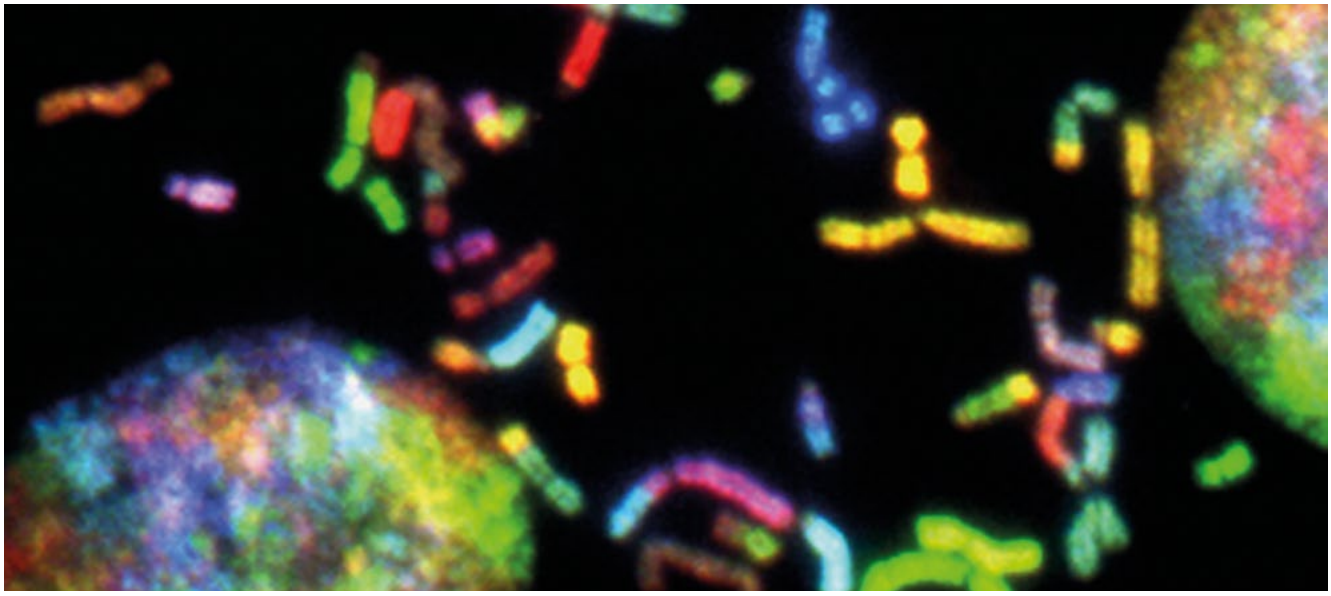


Photo by National Cancer Institute on Unsplash

Despite the advantages of this open science model, lowering or even eliminating the barriers to access and use genetic material and derived information can lead to undesired consequences.[149] Should malicious actors gain access to this genetic data, it is increasingly possible to use it to train machine learning models in order to produce dangerous pathogens. If identifiable genetic markers, for example, single nucleotide polymorphisms (SNPs), can reliably be assigned to differentiate ethnic groups, it would theoretically be possible to target individuals of a specific ethnicity.[150] It is important to note, however, that the genetic knowledge to target such specific groups remains to date elusive for biotechnology researchers and that such malicious initiatives would require advanced technical skills and specialized equipment.

iii. Providing Means for Financing Terrorism

a. Audio Deepfakes[151]

Beyond improving the efficacy of robocalls – the largescale unsolicited automated calls used to deliver pre-recorded messages, often by malicious companies, organized criminal groups and individual scammers –, machine learning can play a significant role in malicious telephonic schemes. In particular, the introduction of "deepfaked" audio content can be used to convince individuals that they are communicating with a person that they know.[152] As will be described in more detail below, deepfakes involve the use of AI techniques to manipulate or generate visual and audio content that is difficult for humans or even technological solutions to immediately distinguish from authentic ones. To create a deepfake audio, machine learning engines are trained using conference calls, YouTube, social media updates and even TED talks, to copy the voice patterns of some target individuals and then generate new audios with the same voice characteristics.

149 ibid.

150 Tao Huang, Yang Shu and Yu-Dong Cai. (2015). Genetic differences among ethnic groups. BMC Genomics 16, 1093. https://doi.org/10.1186/s12864-015-2328-0

151 This section should be read in connection with or with reference to the section on social engineering section below, given the inherent potential for audio deepfakes to feed into social engineering efforts. See Chapter VI Section v(d) below.

152 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philipp Amann, Aglika Klayn, Odhran McCarthy and Maria Eira. (Nov. 19, 2020). Malicious Uses and Abuses of Artificial Intelligence. Trend Micro, EUROPOL and UNICRI. Accessible at http://unicri.it/sites/default/files/2020-11/AI%20MLC.pdf

Precisely this technique was used in 2019, when the CEO of a UK-based firm received a call from who he believed was the chief executive of his firm's parent company in which he was asked to send $244,000 to a Hungarian supplier in an urgent manner. It later became apparent that the caller had used an AI-based software to mimic the voice of the chief executive.[153] The same tactic was used in July 2020 in an attempt to defraud a tech company based in the United Sates.[154]

Should these malicious schemes prove to be a lucrative enterprise, terrorist groups could possibly seek to leverage them to raise funds from people by tricking or threatening them. Besides using these schemes for financial opportunities, they could equally be leveraged by such groups to spy and obtain information by impersonating persons who are in critical positions or tricking the people who are in crucial positions with the deepfake embedded robocalling systems.

b. Crypto-trading

In 2009, a white paper on a peer-to-peer electronic payment system, referred to as Bitcoin, was published by an enigmatic group or individual known only as Satoshi Nakamoto.[155] The paper laid the foundations for a decentralized convertible virtual currency protected by cryptography, making it nearly impossible to counterfeit or double-spend. Shortly after the release of the white paper, the genesis block of Bitcoin was "mined" by Nakamoto,[156] sparking a wild storm of interest and investment in digital assets. New forms of virtual assets, or cryptocurrencies as they are also often called, soon began to appear, such Litecoin, Ripple, Ethereum, Monero, Libra, the meme-inspired Dogecoin and many more.[157]

Photo by Pierre Borthiry on Unsplash

153 Catherine Stupp. (Aug. 30, 2019). Fraudsters Used AI to Mimic CEO"s Voice in Unusual Cybercrime Case. The Wall Street Journal. Accessible at https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

154 Lorenzo Franceschi-Bicchierai. (Jul. 23, 2020). Listen to This Deepfake Audio Impersonating a CEO in Brazen Fraud Attempt. Vice. Accessible at https://www.vice.com/en_us/article/pkyqvb/deepfake-audio-impersonating-ceo-fraud-attempt

155 S. Nakamoto. (2009). Bitcoin: A Peer-to-peer Electronic Payment System. Bitcoin.org.

156 Usman W. Chohan. (2017). A History of Bitcoin. UNSW Business School.

157 European Central Bank. (May 24, 2015). Virtual Currency Schemes – A Further Analysis, European Central Bank.

Due to their nature, cryptocurrencies have found widespread use globally, with even UNICEF beginning to receive, hold and distribute donations in a virtual format in 2019.[158] At the same time, their nature provides cryptocurrencies with an anonymity that has made them an appealing medium for malicious actors to, for instance, illicitly sell drugs, firearms and explosives, smuggle people, launder money and facilitate cybercrime.[159]

In addition to their use as a form of currency, cryptocurrencies have also become a popular asset to be traded as a result of their high market volatility, having created an entire generation of crypto-millionaires and billionaires in a remarkably short period of time.[160] It is in this specific context that AI can play a major role with respect to cryptocurrencies, making market speculation around this virtual currency a valuable avenue for fundraising, rather than merely making appeals to sympathizers for donations.

For instance, on famous underground forums such as "blackhatworld.com", the development and use of AI-powered bots dedicated to cryptocurrency trading has been discussed. As with other applications of machine learning, such a system would rely on the training of machine learning systems with historic data in order to receive more accurate and sophisticated predictions for more profitable cryptocurrency trading.[161] Other forms of the use of AI have also been identified on these forums, such as scanning hundreds of cryptocurrencies in order to find patterns for optimizing the trade of cryptocurrencies, and the use of AI to create cryptocurrency-trading-bots.[162, 163] Several groups in the trading and exchanges world have been developing AI stock trading bots for some time now, with no major success,[164] but the fact that a lot of underground blogs refer to this topic makes it relevant to mention it nevertheless. Besides the AI use to manipulate the cryptocurrency space for financial profit, terrorists could make use of AI to facilitate the theft of cryptocurrencies from "hot wallets",[165] or to facilitate more anonymous transactions on the blockchain. Ensuring more anonymous trading practices reduces the overall risk of them being exposed and losing funds.

Although the systematic usage of cryptocurrencies by terrorist groups and individuals has not yet been seen,[166] there is a growing concern around the use of cryptocurrencies by terrorist organizations.[167] Even if documented examples are limited, there are several notable instances.[168] For example, investigators in the 2019 Sri Lankan bombings that killed more than 250 people observed that the number of transactions in Bitcoin wallets used by ISIL to raise funds increased notably before the bombings, leading to the belief that these Bitcoins played a role in financing the attacks.[169, 170] Similar suspicions exist with respect to the Paris attacks in 2015 by ISIL, although evidence confirming these sus-

158 UNICEF. Blockchain. United Nations. Accessible at https://www.unicef.org/innovation/blockchain

159 INTERPOL. Darknet and Cryptocurrencies. INTERPOL. Accessible at https://www.interpol.int/en/How-we-work/Innovation/Darknet-and-Cryptocurrencies

160 Financial Times. (Mar 7, 2018). The rise - and fall - of the cryptocurrency millionaires. Financial Times. Accessible at https://www.ft.com/content/d0df4322-1559-11e8-9c33-02f893d608c2

161 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philipp Amann, Aglika Klayn, Odhran McCarthy and Maria Eira. (Nov. 19, 2020). Malicious Uses and Abuses of Artificial Intelligence. Trend Micro, EUROPOL and UNICRI. Accessible at http://unicri.it/sites/default/files/2020-11/AI%20MLC.pdf

162 Janny. (Aug. 16, 2019). A comprehensive introduction to Crypto Trading with Bots. Hackernoon. Accessible at https://hackernoon.com/a-comprehensive-introduction-to-crypto-trading-with-bots-ti23930ly

163 Brian Hart (Jun. 1, 2020). AI Trading Platform Tickeron Unveils Cryptocurrency Market Forecasting And Pattern Analysis. PR Newswire. Accessible at https://www.prnewswire.com/news-releases/ai-trading-platform-tickeron-unveils-cryptocurrency-market-forecasting-and-pattern-analysis-301068369.html

164 Victor Hogrefe. (Jun 7, 2018). How Effective Are Trading Bots Really? Victor Hogrefe. Accessible at https://victorhogrefe.medium.com/how-effective-are-trading-bots-really-1684acc1f496

165 Zack Whittaker. (Mar. 16, 2021). $5.7M stolen in Roll crypto heist after hot wallet hacked. TechCrunch. Accessible at https://techcrunch.com/2021/03/16/5-7m-stolen-in-roll-crypto-heist-after-hot-wallet-hacked/

166 Cynthia Dion-Schwarz, David Manheim, Patrick B. Johnston. (2020). Terrorist Use of Cryptocurrencies: Technical and Organizational Barriers and Future Threats. RAND Corporation

167 FATF. (2015). Emerging Terrorist Financing Risks. FATF. Paris. Accessible at www.fatf-gafi.org/publications/methodsandtrends/documents/emerging-terrorist-financing-risks.html

168 The Soufan Center. (Dec. 10, 2020). The Soufan Center. IntelBrief: Terrorists' Use of Cryptocurrency. Accessible at https://thesoufancenter.org/intelbrief-2020-december-10/

169 Roy Katsiri. (2 May, 2019). Bitcoin donations to ISIS soared day before Sri Lanka bombings. Globes. Accessible at https://en.globes.co.il/en/article-exclusive-isis-funded-sri-lanka-bombings-with-bitcoin-donations-1001284276

170 Chainalysis Team. (May 20, 2020). Fact Checking Recent Cryptocurrency Terrorism Financing Reports. Chainalysis. Accessible at https://blog.chainalysis.com/reports/cryptocurrency-terrorism-financing-fact-check

picions is lacking.[171] In a major breakthrough in early 2020, more than 1 million USD in cryptocurrencies was seized from accounts linked to ISIL and Al-Qaida by authorities in the United States.[172] It is also noteworthy that the Financial Action Task Force noted an increased misuse of online financial services and virtual assets to move and conceal illicit funds during the COVID-19 pandemic, describing it as an emerging money laundering and terrorist financing risk.[173] In light of these developments, the possibility of the use of AI-empowered crypto-trading by terrorist groups for the purposes of fundraising should also be considered, although the volatility of the market is certainly likely to diminish the appeal of such tactics at a grander scale.

iv. Spreading Propaganda and Disinformation

a. Deepfakes and Other Manipulated Content

The term deepfakes refers to a type of fake audio and/or visual content that has been manipulated or generated using GANs. As a result of the difficulty they present for both humans and machines to distinguish the real from the fake, deepfakes have arguably become one of the most visible misuses of AI today and garnered significant media attention.

Deepfakes and the technology behind them can be a powerful weapon in today's disinformation wars. Moreover, coupled with the reach and speed of the Internet, social media and messaging applications, deepfakes can quickly reach millions of people in an extremely short period of time. In this regard, deepfakes present considerable potential for a range of malicious and criminal purposes which include: destroying the image and credibility of an individual; harassing or humiliating individuals online, including through the use of sexual deepfakes; perpetrating blackmail, extortion and fraud; disrupting financial markets; and stoking social unrest and political polarization.

The use of deepfakes for disinformation purposes will likely hamper people's trust in traditionally authoritative media. Flooded with increasingly AI-generated fake news that builds on bigoted text, fake videos, and a plethora of conspiracy theories, people may feel that online information, including video, simply cannot be trusted, thereby resulting in a phenomenon termed as "information apocalypse" or "reality apathy".[174] With the rise in fake content, disinformation and misinformation, the world is witnessing a "truth decay" and while many actors play a role in this process, terrorists can most certainly contribute to and exploit it for their own purposes. Indeed, disinformation itself per se is not the most damaging aspect of deepfakes, but rather it is the idea that any information can be fake. The difficulty in authenticating videos will, moreover, allow any compromising information to be denied because any audio-visual content could be fabricated. As a result, even though such video content may in fact not been compromised, it could be claimed as a fake allowing individuals to shirk accountability for their actions.

171 Luis Buenaventura. (Dec. 13, 2015). Did ISIL Really Use Bitcoin to Fund the Paris Attacks? Medium. Accessible at https://medium.com/cryptonight/did-isil-really-use-bitcoin-to-fund-the-paris-attacks-1287cea605e4

172 Andy Greenberg. (Aug.13, 2020). ISIS Allegedly Ran a Covid-19 PPE Scam Site. Wired. Accessible at https://www.wired.com/story/isis-allegedly-ran-a-covid-19-ppe-scam-site/?utm_medium=social&utm_source=twitter&utm_brand=wired&utm_social-type=owned&mbid=social_twitter

173 FATF. (May 2020). COVID-19-related Money Laundering and Terrorist Financing - Risks and Policy Responses. FATF. Accessible at https://www.fatf-gafi.org/media/fatf/documents/COVID-19-AML-CFT.pdf

174 Jennifer Kavanagh and Michael D. Rich (2018). Truth Decay: A Threat to Policymaking and Democracy. The RAND Corporation. Accessible at https://www.rand.org/pubs/research_briefs/RB10002.html

Face swap using Neural Net - Source: Wikipedia commons

Currently, deepfakes are overwhelmingly used to create pornographic content, combining the faces of female celebrities with bodies of pornographic actors.[175] Notwithstanding, deepfakes have the potential of seriously affecting democracies and national security. Considering that social media has become one of the main sources of information for the public, deepfakes pose a significant threat in terms of spreading disinformation, as the information is consumed and reproduced rapidly with users spending little, if any time, authenticating the content.[176,177] Although deepfake videos generally have a quite short lifespan online, they can create momentary panic and confusion, especially when they go viral. Indeed, the inability of individuals to distinguish faked content and the confusion over whether a video is a deepfake or not may even be enough to create significant problems.

Considering the adverse effects of deepfakes, it is conceivable that terrorist groups or individuals can seek to leverage the technology behind deepfakes to run disinformation campaigns on social media to manipulate public opinion or undermine people's confidence in state institutions.[178]

Such technology could also be used as an effective instrument for propaganda, radicalization or as a call for action. For instance, this could be achieved through the creation of "deepfaked" content in which a targeted political figure makes offensive remarks against a specific community in an effort to increase outrage within it and increase the number of sympathizers.

In addition to producing audio-visual deepfakes, AI can also be used to generate customized radicalization narratives. New advanced techniques in NLP, including OpenAI's much publicized GPT-3,[179] has raised concerns with respect to the potential use of the technology in micro-profiling and micro-targeting, generating automatic text for recruitment purposes or spreading customized fake news and terrorism related conspiracy theories, such as assertions by ISIL

175 Giorgio Patrini. (Oct. 7, 2019). Mapping the Deepfake Landscape. DeepTrace.

176 Marie-Helen Maras and Alex Alexandrou. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. International Journal of Evidence & Proof, 23(3): 255–262.

177 Daniel Thomas. (Jan. 23, 2020). Deepfakes: A threat to democracy or just a bit of fun? BBC. Accessible at https://www.bbc.com/news/business-51204954

178 Mika Westerlund. (2019). The Emergence of Deepfake Technology: A Review, Technology Innovation Management Review, V. 9/11.

179 OpenAI. (Mar. 25, 2021). GPT-3 Powers the Next Generation of Apps. OpenAI. Accessible at https://openai.com/blog/gpt-3-apps/

and Al-Qaida that the COVID-19 pandemic is "God's wrath on the West".[180] The use of AI-powered fake news media sites could prove damaging given increasing online readership trends to share articles based on the title or to quick "skim read" an article without conducting substantive due diligence on the website(s) in question.[181] As such, it remains within the realm of possibility that terrorist entities could one day disseminate AI systems which could automatically read real news headlines and create truncated, fake messages to spread on social media and other channels in furtherance of their cause.

Foto di memyselfaneye da Pixabay

Finally, AI could be used in mining for easily-recruitable or easily radicalized men and women, allowing for a targeted distribution of terrorist content or messages. In this case, terrorists could make use of AI as "algorithmic amplifiers" and "recommenders" for spreading propaganda, for instance by directing targeted messages at individuals that have repeatedly searched for violent content online or streamed films portraying alienated and angry antiheroes.

v. Other Operational Tactics

a. Surveillance

Significant developments in computer vision – the methods for acquiring, processing, analyzing and extracting information from digital images or videos – attributable to advancements in machine learning have possibly been one of the major recent developments in the field of AI. Deep learning has revolutionized image and video processing, particularly object recognition, enabling machines to conduct face detection and recognition and to recognize facial

180 UNICRI. (Nov. 2020). Stop the Virus of Disinformation. United Nations. Accessible at http://www.unicri.it/sites/default/files/2020-11/SM%20misuse.pdf

181 Caitlin Dewey. (Jun. 16, 2016). 6 in 10 of you will share this link without reading it, a new, depressing study says. Washington Post. Accessible at https://www.washingtonpost.com/news/the-intersect/wp/2016/06/16/six-in-10-of-you-will-share-this-link-without-reading-it-according-to-a-new-and-depressing-study/

expressions.[182] Beyond the hotly contested realm of facial recognition,[183] the developments in computer vision have also allowed for improvements in human body detection, person identification, attribute recognition, human behaviour recognition, and body movement (gait) recognition.[184] Deep learning has also improved object detection, recognition and tracking, including, for instance, vehicle identification and re-identification and license plate recognition.[185] At the same time, these advancements have allowed substantial drops in error rates of misidentification.[186]

Law enforcement, being a community that has long embraced the use of a wide-range of surveillance technologies such as closed-circuit television (CCTV), body-worn cameras ("bodycams") and patrol drones, has been quick to recognize the potential of deep learning-enabled computer vision to facilitate the identification of victims, perpetrators or other persons of interest. In recent years, there has been a significant growth in law enforcement's interest in AI-based surveillance technologies. An AI Global Surveillance Index compiled by the Carnegie Endowment for International Peace found that 75 of 176 analyzed countries are actively using AI technologies for surveillance purposes, including in smart city/safe city platforms, facial recognition systems and smart policing, demonstrating that the adoption of AI surveillance is increasing rapidly across the globe.[187, 188] The COVID-19 pandemic has also played a significant role in terms of heightening the interest in AI-based surveillance technologies, with several national authorities demonstrating the controversial potential of the technology to support their digital contact tracing effort or to facilitate the enforcement of quarantine measures.[189]

Photo by Dmitry Ratushny on Unsplash

182 Shan Li and Weihong Deng. (2018). Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing. PP. 10.1109/TAFFC.2020.2981446.

183 Unite Nations Human Rights Office of the High Commissioner. (Jun. 25, 2020). New technologies must serve, not hinder, right to peaceful protest, Bachelet tells States. United Nations. Accessible at https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25996&LangID=E

184 Honghua Xu, Li Li, Ming Fang and Fengrong Zhang. (2018). Movement Human Actions Recognition Based on Machine Learning. International Journal of Online Engineering (iJOE). 14. 193. 10.3991/ijoe.v14i04.8513.

185 Jianzong Wang, Xinhui Liu, Aozhi Liu and Jing Xiao. (2019). A deep learning-based method for vehicle licenseplate recognition in natural scene. APSIPA Transactions on Signal and Information Processing, 8, E16. doi:10.1017/ATSIP.2019.8.

186 Douglas Heaven. (Oct. 9, 2019). Why deep-learning AIs are so easy to fool Artificial-intelligence researchers are trying to fix the flaws of neural networks. Nature. Accessible at https://www.nature.com/articles/d41586-019-03013-5

187 Steven Feldstein. (Sept. 2019). The Global Expansion of AI Surveillance. Carnegie. Accessible at https://carnegieendowment.org/files/WP-Feldstein-AISurveillance_final1.pdf

188 It should be noted, as the author repeatedly observes in his assessment, that the use of AI-based surveillance technologies does not necessarily imply these systems are being abused or used in a manner that contravenes human rights.

189 Yann Sweeney. (2020). Tracking the debate on COVID-19 surveillance tools. Nature Machine Intelligence 2, 301–304. https://doi.org/10.1038/s42256-020-0194-1

Flipping this use case on its head, however, the potential for the malicious use of AI-based surveillance technologies becomes a possibility. Particularly for large-scale terrorist attacks, extensive periods of surveillance are often required for planning and preparation. Terrorist groups surveil both places and people to identify and detect a target, determine its suitability for attack and identify weaknesses that can be exploited to facilitate the attack. Traditionally, this is done on foot, in a parked car or online via social networks and may span weeks, months or even years. Advances in AI-enabled surveillance capabilities could theoretically eliminate a considerable portion of the time-consuming aspects of surveillance. With the help of technology, terrorists would be able, for instance, to monitor places and track the movements of people, identify targeted individuals and assets and assess physical security measures at a target location automatically and remotely.

b. Fake Online Identities and Human Impersonation on Social Networking Platforms

The Internet inherently provides a certain degree of anonymity to users. This has been one of the factors that enables some of the Internet's most base and harmful uses, including online trolling, cyberbullying, and the grooming and sexual exploitation of children. It is estimated that there are more than 750 million fake accounts on Facebook that are used to actively spread predetermined and compromised content online. Facebook has acknowledged the scope of this challenge, indicating that it took down as many as 2.9 billion fake accounts in 2019 alone.[190]

The United Nations Secretary-General António Guterres observed in September 2019 that the use of social media and the dark web to coordinate attacks, spread propaganda and recruit new followers had become the new frontier of cyberterrorism. As the main target of the terrorist organizations for recruiting purposes are young persons between 17 and 27 years old,[191] the effective usage of social media becomes even more important for terrorist organizations given the popularity of such platforms within these age groups. In the past, ISIL's usage of social media platforms has fuelled an unprecedented growth in the number of foreign terrorist fighters travelling to the conflict zones. Moreover, the exploitation of the social media platforms has also enabled terrorist organizations to carry out attacks, identify potential recruits, transmit propaganda, disseminate training materials, engage in illicit trading, and raise funds.[192] As terrorist organizations are already using social media platforms with such efficiency, the application of AI would inevitably only further improve their success rate.

In fact, advancements in AI promise to bring a whole new dimension to this phenomenon. GANs, the technology behind deepfakes, are particularly effective at synthesizing highly realistic fake images of faces. For instance, the website "ThisPersonDoesNotExist.com" uses GANs to generate a new and entirely fabricated image of a human face every time the page is accessed or refreshed.

The potential for the malicious use of such technology has already been seen. In 2019, an AI-generated profile picture was used on a LinkedIn account under the name Katie Jones. A young professional in her 30's, Katie profile indicated she worked at a Washington-based think tank and was connected with a number of United States governmental officials. Experts examining Katie's profile flagged it as fictitious, concluding that the fake profile was most probably an attempt to lure persons of interest and collect information from them, possibly as part of an intelligence-gathering operation.[193]

Moreover, the creation of AI-made fake accounts and chatbots for social media platforms is on the rise in criminal forums.[194] Research showed that these fake accounts and bots are increasingly sophisticated and can impersonate

190 Karen Hao. (Mar. 4, 2020). How Facebook uses machine learning to detect fake accounts. MIT Technology Review. Accessible at https://www.technologyreview.com/2020/03/04/905551/how-facebook-uses-machine-learning-to-detect-fake-accounts/

191 António Guterres. (Sept. 25, 2019). Secretary-General Calls Cyberterrorism Using Social Media, Dark Web, "New Frontier" in Security Council Ministerial Debate. United Nations. Accessible at https://www.un.org/press/en/2019/sgsm19768.doc.htm

192 United Nations. (Dec. 1, 2016). Concept Note: Special Meeting of the Counter-Terrorism Committee with Member States and relevant international and regional organisations, civil society and the private sector on "Preventing the Exploitation of Information and Communications Technologies for Terrorist Purposes, while Respecting Human Rights and Fundamental Freedoms". Accessible at https://www.un.org/sc/ctc/wp-content/uploads/2016/11/Concept-Note.pdf

193 Raphael Satter. (Jun. 13, 2019). Experts: Spy used AI-generated face to connect with targets. ABC news. Accessible at https://abcnews.go.com/Technology/wireStory/experts-spy-ai-generated-face-connect-targets-63674174

194 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philipp Amann, Aglika Klayn, Odhran McCarthy and Maria Eira. (Nov. 19, 2020). Malicious Uses and Abuses of Artificial Intelligence. Trend Micro, EUROPOL and UNICRI. Accessible at http://unicri.it/sites/default/files/2020-11/AI%20MLC.pdf

average users of the relevant social media platforms, allowing them to avoid being flagged by users or being detected and banned by the platforms.[195] These fake accounts and bots are being used for several purposes, such as increasing hits, followers or "likes" of certain pages or content.

Terrorist organizations can benefit from such AI-powered developments to increase the efficiency of their use of social media platforms. AI-generated fake accounts and bots that can impersonate average users can help terrorist organizations to spread their messages with greater ease on social media platforms and with less risk of the real individuals been banned or facilitate social engineering efforts in an effort to solicit sought after information or to support radicalization efforts.

c. Morphed Passports

Improperly obtaining, altering or counterfeiting travel documentation is essential for modern terrorist groups. Falsified documentation is regularly used by terrorists, for instance, to facilitate international travel – as was the case with several of the 9/11 hijackers.[196] It is also often used as proof of identity for other administrative purposes. For instance, the individuals behind the 2015 attacks in Paris used forged passports to obtain a loan in advance of the attacks.[197] The use of forged passports for terrorism purposes is so widespread that investigators even believed that terrorist organizations such as Al-Qaida have their own specialized members in various countries whose sole task is to supply other members of the organization passports and other relevant documents upon request.[198] More recently, ISIL is considered to have "industrialized" the production of fake passports.[199]

Photo by CardMapr.nl on Unsplash

A dangerous and new AI-based method for the creation of fake passports may soon come to the fore, with what can be referred to as "morphed" passports. Using the MorGAN (*Mor*phing through *G*enerative *A*dversarial *N*etworks) method, it has been seen that criminals can create passport photos that can be matched with more than one individual.[200] In

195 ibid.

196 National Commission on Terrorist Attacks upon the United States. (2004). Entry of the 9/11 Hijackers into the United States: Staff Statement No. 1. USA Government. Accessible at https://govinfo.library.unt.edu/911/staff_statements/staff_statement_1.pdf.

197 Europol. (n.d.) Forgery Of Administrative Documents And Trafficking Therein. Europol. Accessible at https://www.europol.europa.eu/crime-areas-and-trends/crime-areas/forgery-of-administrative-documents-and-trafficking-therein

198 Oriana Zill. Crossing Borders: How Terrorists Use Fake Passports, Visas, and Other Identity Documents. PBS. Accessible at https://www.pbs.org/wgbh/pages/frontline/shows/trail/etc/fake.html

199 Brian Ross, Michele Mcphee and Lee Ferran. (Jan. 25, 2016). ISIS Has Whole Fake Passport "Industry," Official Says. ABC. Accessible at https://abcnews.go.com/International/isis-fake-passport-industry-official/story?id=36505984

200 Naser Damer, PhD. (n.d.). Face morphing: a new threat? Fraunhofer IGD. Accessible at https://www.igd.fraunhofer.de/en/press/annual-reports/2018/face-morphing-a-new-threat

other words, morphed passports refer to a single passport that can be used by two or more individuals.[201] Leveraging this, malicious actors can trick both human and machine-based recognition systems, easily bringing down a traditionally strong border security.[202] In order to address some of the security concerns surrounding morphed passports, authorities in some countries have begun to adopt additional measures, such as obliging individuals to take passport photos at passport offices, rather than to supply their own – possibly fake – photo.[203, 204]

Whether these additional measures prove sufficient remains to be seen, but what is already clear is that unchecked, morphed passports may greatly enhance the ability of terrorists to move undetected through border controls and security checks at airports or even just in public spaces.

d. Online Social Engineering

Social engineering is a well-established attack vector that relies on human interaction to exploit weaknesses, frequently involving manipulation. It is often used by criminals and other malicious actors in scams to obtain money or confidential information or to convince the victims to do something that they would not otherwise do.[205] This attack vector can be applied both online, primarily through social media, or offline, in-person.

Chatbots are one of the most visible and discernible uses of AI in modern-day society. As AI keeps advancing, bots will likely play an increasing role in online deception, including through social engineering schemes.

Needless to say, terrorist organizations exploit social engineering tactics online, primarily to help them identify and recruit new members and sympathizers. Indeed, these groups already have considerable experience in using bot accounts.[206] Following the 2015 Paris attacks, the hacktivist group Anonymous launched an online campaign against ISIL where it claimed to have removed as many as 25,000 ISIL bots online.[207]

Currently, chatbots excel in very narrow contexts with repetitive elements, such as e-commerce support and customer service. With advancements in NLP, bots can learn over time on the basis of their interactions with humans, enabling them to respond in a manner that better resembles a human. As the ability to distinguish a bot from a human becomes more challenging, the potential for the use of bots in social engineering attacks increases.[208]

At the same time, the success of social engineering tactics depends on their cogency and, in this regard, obtaining detailed and accurate information on the target plays an essential role. AI can play a role here too. For instance, new AI-powered account detection tools, using facial recognition algorithms, are being explored in online forums that would enable a user to match multiple separate accounts of the same person on different social media platforms, even if the profile picture is not the same. With this technology, it could be possible for a malicious actor to quickly identify several of their target's social media profiles.[209] By analyzing these profiles, the malicious actors can then develop a more complete understanding of the individual in question and gather the necessary information to better manipulate the individual in question and, for instance, make him or her share confidential information either through deceit or coercion.

201 David J. Robertson, Andrew Mungall, Derrick G. Watson, Kimberley A. Wade, Sophie J. Nightingale, Stephen Butler. (2018). Detecting Morphed Passport Photos: A Training and Individual Differences Approach. Cognitive Research: Principles and Implications, V. 3/1.

202 David J. Robertson, Andrew Mungall, Derrick G. Watson, Kimberley A. Wade, Sophie J. Nightingale, Stephen Butler. (2018). Detecting morphed passport photos: a training and individual differences approach. Cognitive Research: Principles and Implications.

203 Reuters Staff. (Jun. 3, 2020). Germany bans digital doppelganger passport photos. Reuters. Accessible at https://www.reuters.com/article/us-germany-tech-morphing/germany-bans-digital-doppelganger-passport-photos-idUSKBN23A1YM

204 Luana Pascu. (Jun 17, 2020). Germany bans passport photo morphing to prevent biometric spoofs at border checks. Biometric Update. Accessible at https://www.biometricupdate.com/202006/germany-bans-passport-photo-morphing-to-prevent-biometric-spoofs-at-border-checks

205 INTERPOL. Social engineering scams. INTERPOL. Accessible at https://www.interpol.int/en/Crimes/Financial-crime/Social-engineering-scams

206 Steven Stalinsky & R. Sosnow. (Aug. 5, 2020). Jihadi Use of Bots on the Encrypted Messaging Platform Telegram. Memri. Accessible at https://www.memri.org/reports/jihadi-use-bots-encrypted-messaging-platform-telegram

207 Leanna Garfield. (Dec 14, 2015). ISIS has created thousands of political bots — and hacktivists want you to destroy them. Business Insider. Accessible at https://www.businessinsider.com/anonymous-battles-isis-political-bots-2015-12

208 Simon Chandler. (Dec. 21, 2018). The evolution of evil chatbots is just around the corner. Daily Dot. Accessible at https://www.dailydot.com/debug/evil-chatbot-hackers-ai/

209 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philipp Amann, Aglika Klayn, Odhran McCarthy and Maria Eira. (Nov. 19, 2020). Malicious Uses and Abuses of Artificial Intelligence. Trend Micro, EUROPOL and UNICRI. Accessible at http://unicri.it/sites/default/files/2020-11/AI%20MLC.pdf

VII. UNFOLDING THE TERRORIST USE OF AI

The malicious use of AI is very much a new and not yet fully understood domain. The potential impact of the use of AI for terrorist purposes equally remains to be seen. It may, accordingly, be difficult for law enforcement agencies, security forces and counter-terrorism bodies, as well as policymakers, industry and academia, to come to terms with the malicious use of AI. Consequently, it may be difficult to carry out a thorough and informed assessment of the risk.

In order to support overcoming these difficulties, a series of fictional scenarios have been developed. These scenarios are described below with the intention of aiding readers to visualize how AI could be integrated into the modus operandi of fictional terrorist groups. The likelihood of such a scenario – as perceived by participants of the UNCCT/UNOCT-UNICRI Expert Group Meeting – is additionally provided after each scenario.

Scenario 1:

Pandemic Problems – Lantium Government and Vaccine Suffer Deadly Blow in Complex Cyber Attack

Following a period of economic downturn in the Republic of Lantium, a terrorist group has risen in prominence. The group, known as "The Redeemers", has launched several cyber and physical attacks on the government and its institutions in an attempt to weaken and overthrow it. The group is also known to have associations with several like-minded groups throughout the region, to which they have provided their extensive cyber-capabilities in pursuit of their common objectives. These capabilities include the creation of a social media application with end-to-end user encryption, known as "The Hub", that allows both the exchange of information between these groups and the dissemination of propaganda to the public to bolster the Redeemers' appeal and recruit new members.

Now, deep in the midst of a global pandemic, Lantium is struggling to launch its vaccination programme to recover from the outbreak and keep its citizens safe from the virus. More than 3% of its population has been infected since the beginning of the pandemic, with a mortality rate 3.4 percent higher than that registered in most countries. In these early days of the vaccination programme, Lantium has secured 40,000 doses a week of the FuturePharm vaccine. All eyes are on the implementation of the government's vaccination programme during these difficult times. Aware of this, the Redeemers decide to exploit these circumstances and begin to plan a new attack to deliver a deadly blow to the government's credibility as a whole.

Relying on their extensive network of sympathizers, the Redeemers succeed in gaining access to Lantium's CCTV surveillance system operated by the National Police's Biometrics Unit. Within a few hours of accessing the system and monitoring the movements of individuals that work at the Lantium's National Hospital, the Redeemers identify the key hospital staff based on their movements in and out of its administrative wing. Using a facial recognition software programme, they are able to search for their identity by using face screenshots of the CCTV cameras' recordings. Cross-referencing the names of the identified individuals with user profiles of the professional networking site "connected.com", the Redeemers select an appropriate target in the hospital – Ms. Alison Apple, one of the hospital's senior programme leads for the vaccination programme rollout.

The following night, the Redeemers start their extensive cyber-attack, targeting Ms. Apple as the point of entry. They start by launching an AI-password guessing attack against her, employing a neural network that was trained using online databases of stolen passwords obtained on the dark web. Once engaged, the neural network generates high-quality password guesses and ultimately gains access to Ms. Apple's official account on the hospital network. Once in, the Redeemers search for vaccine programme related documentation. They soon discover that these sensitive documents have been encrypted to protect their integrity. In response, the Redeemers deploy an AI-powered decryption tool, which unlocks the documents' security features, revealing essential data and information concerning the vaccine programme, including the primary storage hub for the vaccine in the hospital and the information on the protocols and systems for its storage. Having access to the hospital network, the group targets the refrigeration system for the vaccine, increasing the temperature in the freezers above the recommended minus 40 degrees Celsius for five hours overnight. Being kept outside of the temperature requirements necessary to preserve its integrity, the active components are neutralized, and the vaccine's effectiveness decreases from 95% to just 5% without any change in its external appearance. The group's strong computer science background allows them to also disable the temperature sensors and the notification system employed to alert the hospital staff of any temperature changes. Having done this, the group scrubs the records before exiting the system, leaving no evidence about any intervention. By the time the hospital staff returns to work the following morning, all freezers are back to functioning at the recommended storage temperature. Unaware of the Redeemer's intervention, the hospital staff continue to deliver the vaccine.

After several weeks of vaccinations, major media outlets begin to report growing numbers of cases in which people contract the virus, and some even die, despite having been fully vaccinated. As a frenzy develops around these reports and concern surrounding the vaccine grows, the Redeemers, released a deepfake audio clip on The Hub of the Prime Minister, Ms. Cristina Caballero, who was a steadfast champion and public face of the government's response to the virus. In the fake call, Ms. Caballero suggests to members of her cabinet team that the vaccine does not work and that it was all part of an elaborate ruse to collect the genetic material from citizens in order to enhance the government's control over the population. The deepfake was generated by the Redeemers using a deepfake app available for smartphones and several audio samples of public statements made by the Prime Minister. Although Ms. Caballero promptly denies the authenticity of the audio clip, it quickly goes viral, prompting public outrage and forcing her resignation. Although the Deputy Prime Minister takes her place, the confidence in the government hits an all-time low and, without the strong leadership of Ms. Caballero, the government slowly begins to crumble under internal conflicts, leading the public to call for general elections. At the same time, market analysts report that "hundreds of users per minute" joined The Hub in the aftermath of the audio clip's release.

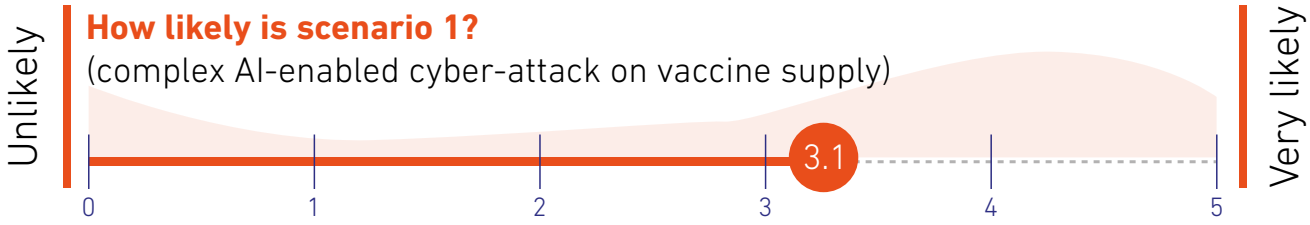


Figure 4: Expert Group Meeting participants perceived likelihood of scenario 1.

Scenario 2:

No Place to Hide – Drone Swarm Kills Famous Activist

A sunny afternoon in New Vineland is disturbed as a fleet of five drones flying in unison whizz through the city's crowded central square. The crowds' interest in seeing the unusual display, mere metres from their heads, quickly turns to panic and confusion as a small blast suddenly rocks the square. Rushing to the scene, police officers discover that one man has died in the explosion and several others suffered serious injuries as a result of their proximity to the explosion. The deceased man is later identified as Mr. Benjamin Brown.

Mr. Brown was an outspoken member of his community who had begun to receive considerable media attention for opinion pieces in local media and public speeches in which he delivered intense and scathing remarks undermining the ideology of the terrorist group known as "the Chosen". Over time, the passion and power of his statements contributed to him becoming an internationally recognized figure, and he had even been recently delivered a TED talk, in which he again vigorously attacked methods and motivations of the Chosen. On the morning of the attack, Mr. Brown had been speaking at a local book shop on the Central Square in an event related to the release of his latest book "Chosen to be nothing more than liars and petty criminals".

The Chosen quickly claimed responsibility for the attack in a video message shared on a social media channel linked to the group. In the video, the group celebrates the death of Mr. Brown, flaunting that "vengeance will find their enemies wherever they hide".

As the investigation into the attack progresses, the authorities determine that the drones seen on the Central Square just before the attack played a central role and that the Chosen had weaponized the drone swarm with improvised explosive devices (IED). Each of the drones had been equipped with cameras and facial recognition technology, scanning the target from different angles, and therefore having a different perspective of his face. To avoid false positives, the software would automatically compare the results from the facial recognition application installed in each drone and release the explosives only when at least two of the drones would detect the target. With this method, the accuracy of the drones' target detection was increased to 90%, ensuring that the opportunity for the Chosen to eliminate this thorn in its side would not be wasted. The fleet scanned for the target coming out of the book shop and as soon as Mr. Brown's identity had been detected, the drones entered a downward spiral, deploying their explosive payload and killing the target in the blast.

After several weeks, the police conclude their extensive investigation and arrest three members of the Chosen for their involvement in the attack. In a press release, the police indicate that evidence suggests that the drones had been purchased online through different commercial vendors for drone enthusiasts. They also communicate that the Chosen had used the same website to acquire the multi-flyer controller smartphone app that allowed to pilot the drone fleet and perfectly coordinate the drones' flight. With respect to the facial recognition software, the police indicated that, although the Chosen were not known to have a very strong background in computer science, it was believed that the group had also acquired online a commercially available facial recognition software, having chosen the most suitable option based on online reviews and descriptions. The software itself came with instructions that described in detail the installation process. Its specifications also noted that the programme simply required multiple pictures of the targets to be operationalized – something easily obtainable for a public figure such as Mr. Brown.

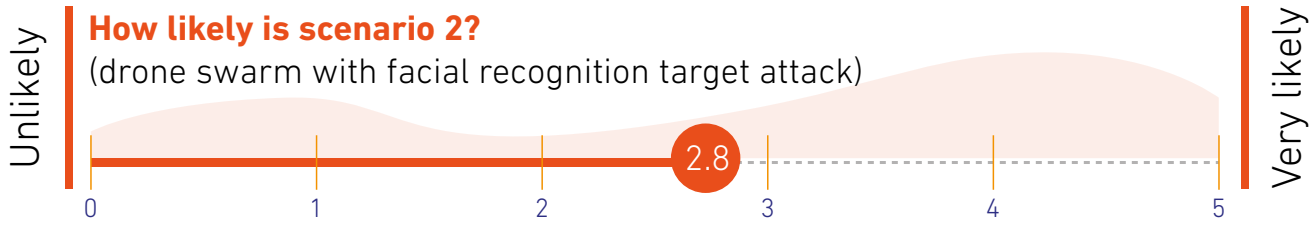


Figure 5: Expert Group Meeting participants perceived likelihood of scenario 2.

Scenario 3:

Borders No More – Fake Passports Facilitate Bomb Attack in Capital City

Early in the year, the popular underground hacking community forum, "Broke.it", began to feature advertisements from the hacker group "Hacking For You" for "AI-as-a-service". The ads in the forum's channels included a wide range of services, including the generation of morphed passports. The ads indicate that interested customers need only to send pictures of the intended users' faces and one photograph of a local passport and, upon the payment of a fee, the hacker group would produce and issue the fake passport.

This service quickly became a bestseller for Hacking For You, given that the production of morphed services was relatively easy and that the hacker group charged low fees for the service. The hacker group's work and reputation quickly grew throughout the criminal underworld, with transnational criminal groups increasingly using these fake passports to pass undetected through border controls and carry out their activities internationally with greater ease.

As the popularity of these morphed passports continued to grow, word began to also spread to other malicious actors, including terrorist and violent extremist groups. In July, Hacking For You received a request for morphed passports from individuals associated with the violent extremist group Brimstone. The group, which is physically located in and operates out of Terre North, is famed for its violent and high-profile attacks on soft targets throughout the region, including in neighbouring Southany, Westlandia, and Eastopolis. Having presented the fee upfront, Hacking For You accepts the request from Brimstone and begins to produce several morphed passports for them.

In a matter of weeks, Brimstone receives its shipment of morphed passports, which combine the faces of members of the group with illegally acquired images of faces of nationals from Terre North, Southany, Westlandia and Eastopolis. This allows members of the group to travel back and forth across the borders undetected, as the border control officers tend to quickly check the name, ID number and profile picture of the traveller. Initially, Brimstone uses this capability to facilitate some of their illicit activities that finance the group, including kidnapping and ransom, illegal mining, extortion, and the production and distribution of illegal drugs.

However, they soon realize that these passports have even greater potential. After several weeks of discussion, Brimstone settles on their next high-profile attack: a series of sequential bombings on key targets in the Capital City in Westlandia, including subway stations, a public square, a sporting venue, a hospital and the city's High Court. Brimstone considers that amid the confusion that will ensue from the sequential attacks, the attackers will be able to quickly and quietly slip across the border, which lies just 100 kilometres to the east of Capital City and return to Terre North unnoticed using their new morphed passports. Confident that they can successfully evade capture, Brimstone opts to proceed with the plan, settling on mid-October for the attack.

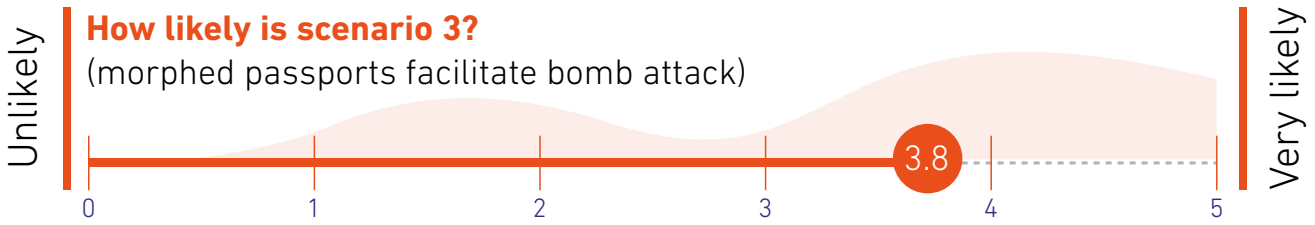


Figure 6: Expert Group Meeting participants perceived likelihood of scenario 3.

VIII. ASSESSING THE THREAT

Having looked at the current state of play in terms of the terrorist use of AI and identified several hypothetical yet conceivable examples that demonstrate how terrorists could leverage this technology, one critical question remains: is there cause for concern regarding terrorist groups or individuals directly employing AI in ways similar to those described in fictional scenarios depicted in the previous chapter?

Before addressing this, it is prudent to note, as Van der Veer observes, that there are "stakes and agendas" in discussions about the terrorist use of technology.[210] She observes that consultants, advisors and other private entities might be interested in maintaining an alarmist narrative around issues for which they sell their services. Non-expert audiences participating in the debate might consequently have a hard time differentiating between biased or influenced narratives and sustained or neutral claims, particularly when these debates focus on highly technical issues.

Taking this into account, this chapter endeavours to objectively reflect on the aforementioned question by analyzing the level of the threat of the terrorist use of AI in an attempt to reach a conclusion. The term "threat" is commonly understood as a combination of *intention* and *capability*. Both terms will be addressed in the sections that follow.

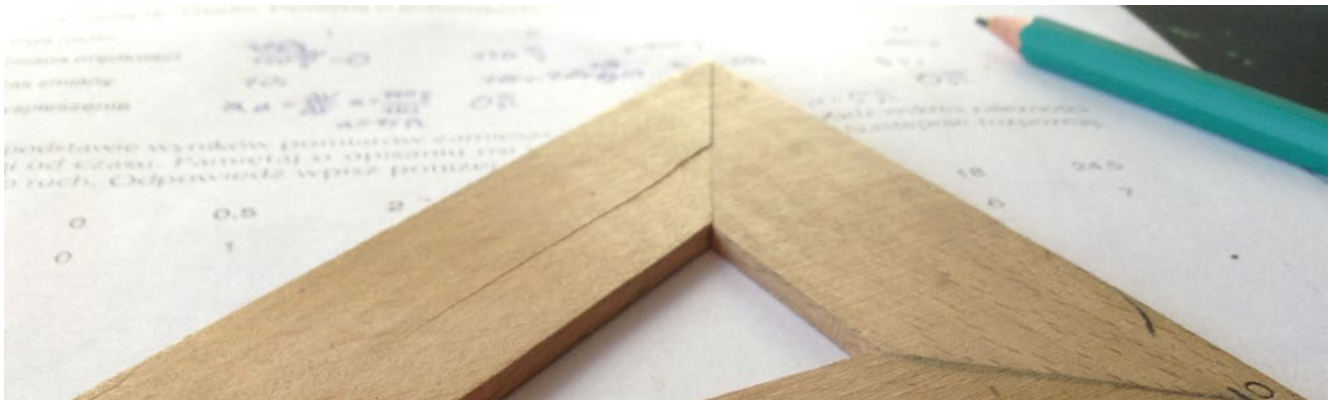


Photo by Dawid Małecki on Unsplash

210 Renske van der Veer. (2019). Terrorism in the age of technology in Strategic Monitor 2019-2020. The Hague Centre for Strategic Studies and the Clingendael Institute. Accessible at https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terrorism-in-the-age-of-technology/

a. Intention

Given the nature of terrorism, it is difficult to ascertain the intentions of any terrorist groups or individuals. Nevertheless, in order to assess terrorists' intention to use AI, there is merit in looking at the suitability of this technology to terrorism. In her analysis of how open technological innovation is arming the terrorists of tomorrow, Cronin lists the characteristics that she considers non-State actors seek out in innovative weapons.[211] Cronin considers that innovative weapons need to be accessible, cheap, simple to use, transportable, concealable, and effective – characteristics that AI does not necessarily possess. The reality is that AI is not perfect. Contrary to its frequent representation in popular culture and the media, AI is not a silver bullet. It can fail and very often it does fail.[212] According to VentureBeat, an estimated 87% of data science projects never make it to production.[213] TechRepublic reports that 56% of global CEOs do not expect to make any return on investment for 3-5 years.[214] Successfully developing and implementing AI in an effective and dependable manner requires considerable time, money, and effort. As Hoffman implies, there are good reasons for terrorist groups having stuck to two primary weapons systems – firearms and explosives – for more than a century: they are effective and dependable.[215]

On the other hand, Cronin also notes that the innovative weapons must be useful in a wide range of contexts to be attractive to terrorists.[216] They should be part of a cluster of technologies that can magnify their effects, are symbolically resonant, and can be given to unexpected uses. Unlike the prior set of characteristics, it can be argued that AI does in many ways fit this description and lends credence to the possibility of terrorists having interest in the technology.

Notwithstanding this analysis, as has already been noted, there have been some early signs of interest from terrorist groups or individuals in AI and related technologies. Drones, for instance, are increasingly being integrated into the modus operandi of groups like ISIL. Moreover, reflecting once again on the extensive history of how such groups have innovated and embraced new technology in the past, it is perhaps prudent to at least consider the possibility that terrorist organizations have, to some degree, the intention of exploring or seeking to understand how AI can be leveraged for malicious purposes.

b. Capability

The capability of terrorist groups or individuals to develop or deploy AI could very well be a "make-or-break" aspect in this analysis.

It is undeniable that AI capabilities are rapidly growing across the globe and, generally speaking, AI technologies, as well as the means to develop and deploy these technologies, can be acquired commercially, and some are even open-source. For example, TensorFlow, an open-source library for large-scale machine learning and numerical computation, allows users to easily build a neural network with simple object detection or even a facial recognition model without the need of sophisticated skills or computers.[217] Github is another open-source platform that could lower the threshold of use and access, increasing the possibility of malicious actors such as terrorists to make use of AI.[218] Yet, the accessibility of technology alone is insufficient if the capacity required to leverage it does not exist.

Examining the level of technical capabilities, experts have tended to suggest that terrorist groups such as ISIL have not carried out effective and sophisticated cyber and technology-based attacks because they lack the necessary capabili-

211 Audrey Kurth Cronin. (Jan. 2020). Power to the People: How Open Technological Innovation is Arming Tomorrow's Terrorists.

212 Terence Tse, Mark Esposito, Takaaki Mizuno, and Danny Goh. (Jun. 8, 2020) The Dumb Reason Your AI Project Will Fail. Harvard Business Review. Accessible at https://hbr.org/2020/06/the-dumb-reason-your-ai-project-will-fail

213 Venture Beat Staff. (July 19, 2019) Why do 87% of data science projects never make it into production? Accessible at https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/

214 Alison DeNisco Rayome. (July 9, 2018). ROI on AI investments could take up to 5 years, 56% of manufacturing CEOs say. Accessible at https://www.techrepublic.com/article/roi-on-ai-investments-could-take-up-to-5-years-56-of-manufacturing-ceos-say/

215 United Nation Office on Drugs & Crime. Terrorism and Conventional Weapons. Accessible at https://www.unodc.org/images/odccp/terrorism_weapons_conventional.html

216 Audrey Kurth Cronin. (Jan. 2020). Power to the People: How Open Technological Innovation is Arming Tomorrow's Terrorists.

217 TensorFlow. (n.d.). Introduction to TensorFlow. TensorFlow. Accessible at https://www.tensorflow.org/learn

218 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philipp Amann, Aglika Klayn, Odhran McCarthy and Maria Eira. (Nov. 19, 2020). Malicious Uses and Abuses of Artificial Intelligence. Trend Micro, EUROPOL and UNICRI. Accessible at http://unicri.it/sites/default/files/2020-11/AI%20MLC.pdf

ties or funding or are simply not sufficiently organized to do so.[219] Even if the ISIL's creation of a United Cyber Caliphate is a troubling development, the group is nevertheless considered to be very much in its infancy.[220] In fact, the group employs techniques used by so-called "script kiddies" – low-skilled hackers that use scripts or programmes developed by others to carry out their attacks without really understanding how they work. Notwithstanding this, Flashpoint researchers have suggested that "willingness to adapt and evolve in order to be more effective and garner more support indicates that while these actors are still unsophisticated, their ability to learn, pivot, and reorganize represents a growing threat".[221] Even if groups such as the United Cyber Caliphate may not have the technical capabilities required within their ranks at present, the actions of such groups are likely to capture the imagination of others and may, in time, stimulate the next generation of more capable cyberterrorists.

In lieu of having their own team of experts or in-house technical capabilities to carry out more sophisticated cyber-attacks and theoretically launch an AI-enable attack, terrorist groups and individuals may alternatively look to outsource or offshoot opportunities.[222] For instance, it has been seen that for cyber-terrorism purposes, some Al Qaida/ISIL-inspired organizations have joined forces even with groups that are not identified with their agenda. In light of the evolving and multifaceted nexus between organized crime and terrorism and the advent of the "crime-as-a-service" model, whereby the tools of the cybercrime trade are sold for a monetary value, it may be possible for such low-skilled terrorist groups to simply purchase pre-built or custom-made algorithms "off the shelf", ready for use, or avail themselves of such systems a service. By way of example of the potential of this "crime-as-a-service" model, in November 2020 a digital rights activist reportedly purchased access to Moscow's facial recognition system for 16,000 roubles (approximately 200 USD) from an ad posted on Telegram.[223] Shortly after paying this fee, the activist was able to obtain a detailed report of her movements recorded in the facial recognition system over the previous months. In this sense, terrorist organizations do not necessarily need to be capable of developing such technology themselves, but can outsource attacks via the black market to "hackers for hire" or other nefarious criminal groups, providing a greater direct correlation between criminal accessibility and terrorist capability.

Photo by Charles Deluvio on Unsplash

Additionally, the risk of already developed sophisticated technologies ending up in the wrong hands is ever-present. With the growing interest in drones and their use in conflict zones, as well as in the development and deployment of increasingly autonomous weapons systems in combat environments, there is a concern that such weaponry could

219 Grant Gross. (Apr. 28, 2016). ISIS cyberattack capabilities are unorganized, underfunded—for now. PC World from IDG News Service. Accessible at https://www.pcworld.com/article/3062980/isis-cyberattack-capabilities-are-unorganized-underfunded-for-now.html

220 C.S. Liang. (2018). Unveiling the United Cyber Caliphate and the Birth of the e-Terrorist, Georgetown University Press.

221 L. Alkhouri, A. Kassirer and A. Nixon. (Apr. 2016). Hacking for ISIS: The Emergent Cyber Threat Landscape. Flashpoint.
L. Alkhouri, A. Kassirer and A. Nixon, 2016. ""Hacking for ISIS: The Emergent Cyber Threat Landscape" , Flashpoint, April, 201
L. Alkhouri, A. Kassirer and A. Nixon, 2016. ""Hacking for ISIS: The Emergent Cyber Threat Landscape" , Flashpoint, April, 201

222 This may be understood as part of a broader tendency of warfare in the 21st century being increasingly waged by state and non-states actors using either or both human and technological surrogates: Krieg, Andreas and Rickli, Jean-Marc (2019). Surrogate Warfare: The Transformation of War in the Twenty-first Century. Georgetown: Georgetown University Press. Accessible at http://press.georgetown.edu/book/georgetown/surrogate-warfare.

223 Russell Brandom. (Nov. 11, 2020). Moscow's facial recognition system can be hijacked for just $200, report shows. The Verge. Accessible at https://www.theverge.com/2020/11/11/21561018/moscows-facial-recognition-system-crime-bribe-stalking

be seized or illegally purchased or acquired by non-State actors, like terrorist groups.[224] The possibility of this is in fact one of the arguments often used by experts calling for prohibitions on the development of autonomous weapon systems.[225]

Ultimately, while it may appear that groups such as ISIL lack the capabilities to design, develop and implement AI themselves, the possibility of such groups or individuals acquiring the capabilities to deploy these technologies cannot be ruled out. It is prudent to note that, historically, significant advancements in terms of capabilities have taken place within relatively short timespans. It is demonstrative that it took less than a year for ISIL to successfully use drones in their operations after evidencing their initial interest to use this technology as part of their repertoire.[226] Even if for now, the most concerning technologies are those that have low barriers to entry, malicious actors are likely to build-up their skills for more advanced attacks over time.

c. Cause for Concern?

In 2004, the United States' 9/11 Commission released its report on the events leading up to the 11 September 2001 terrorist attacks in the United States. In this report, the Commission underscored the dangers in failures of imagination and went as far as to encourage the institutionalization of imagination in assessing terrorist threats going forward.[227] Reflecting on these hard-learned lessons, it is perhaps prudent to thus consider the threat of the terrorist use of AI as a possibility. Even if the assessments of the terrorist intent or capability are not entirely conclusive at present, it is, in the interest of not being caught unprepared, advisable to err on the side of caution in terms of the preceding assessment of intent and capability.

As has been seen on several occasions and touched upon in the present report, technology plays a role in shaping the forms of terrorism.[228] Considering the rapid integration of AI into daily life, it could be said that it can no longer remain a low probability that terrorist groups and individuals will not use AI-based technologies in the not too distant future[229] – whether it is in one or more of the malicious uses described in this report or in some other still unimaginable way.[230] In this regard, progress and development in AI and the increasing interest of terrorist groups and individuals in these and related technologies should not be overlooked.

Be that as it may, there is another side to the coin that merits consideration. Andy Patel, a researcher at F-Secure, has suggested that today's AI systems have more to fear from humans than humans have to fear from AI.[231] In this regard, he notes that it is more likely for terrorist groups and individuals to abuse the AI systems, rather than using them as a part of their attacks.

There is, accordingly, merit in noting one final distinction before concluding this examination of the intersection of AI and terrorism: the distinction between the *use* and *abuse* of AI. Whereas the malicious *use* of AI concerns malicious actors using AI to, for instance, enhance the efficacy of an attack, the *abuse* of AI concerns attacks directed to the

224 Colin P. Clarke. (Aug. 20, 2018). Drone terrorism is now a reality, and we need a plan to counter the threat. World Economic Forum. Published in collaboration with the RAND Corporation. Accessible at https://www.weforum.org/agenda/2018/08/drone-terrorism-is-now-a-reality-and-we-need-a-plan-to-counter-the-threat/

225 P. Chertoff. (Oct. 2018) Perils of Lethal Autonomous Weapons Systems Proliferation: Preventing Non-State Acquisition, Strategic Security Analysis, Geneva Centre for Security Policy.

226 T. H. Tønnessen. (2017). Islamic State and Technology – A Literature Review, Perspectives on Terrorism, V. 11/6.

227 The 9/11 Commission. (Jul. 22, 2004). The 9/11 Commission Report. Thomas H Kean. Accessible at https://www.9-11commission.gov/report/911Report.pdf

228 H. K. Tillema. (2002). A Brief Theory of Terrorism and Technology. In T. K. Ghosh (ed.), Science and Technology of Terrorism and Counter-terrorism.

229 K. Steinmüller, (2017) The World in 2040. Framework Conditions for New Kinds of Terrorism, in Identification of Potential Terrorists and Adversary Planning: Emerging Technologies and New Counter-Terror Strategies, T.J. Gordon et al. (eds.).

230 Adam Pilkey. (Jul. 11, 2019). Artificial intelligence attacks. F-Secure. Accessible at https://blog.f-secure.com/artificial-intelligence-attacks/

231 Help Net Security. (Jul. 16, 2019). How can attackers abuse artificial intelligence? Help Net Security. Accessible at https://www.helpnetsecurity.com/2019/07/16/abuse-artificial-intelligence/

operation or functionality of AI, by manipulating them with physical and/or cyber capabilities.[232] Such use/abuse dynamics can be seen with respect to other technologies, including for instance virtual assets. While such assets are certainly used to facilitate cybercrime, crypto-assets and enterprises increasingly serve as a target for hackers and fraudsters.[233] The abuse of AI would essentially entail situations such as a terrorist group instigating an attack on an AI system or seeking to obstruct such a system. As AI is increasingly integrated into systems in both the public and private sector, new vulnerabilities arise – particularly when such systems are in embedded in critical infrastructure. A recent attempt by a hacker to poison a water treatment plant in Florida presents the magnitude that a traditional cyber-attack on critical infrastructure could have on an entire city population.[234]

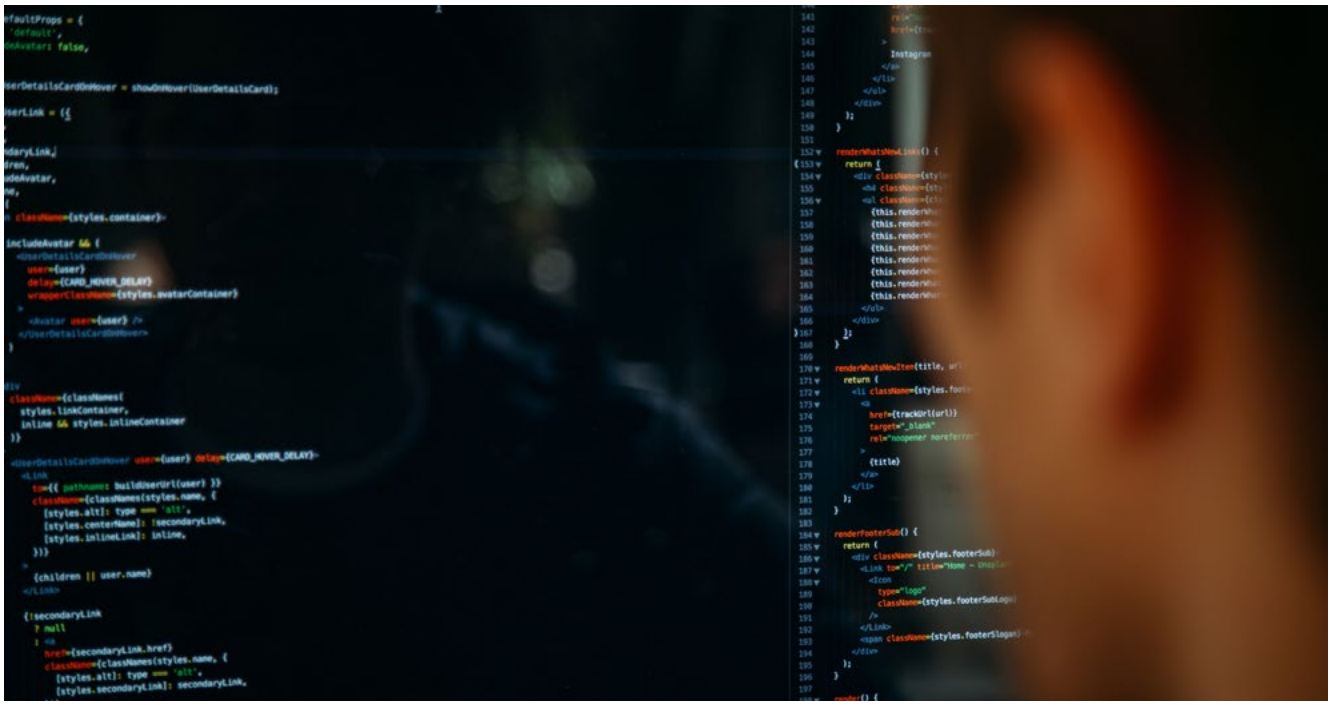


Photo by Charles Deluvio on Unsplash

While the realm of possibilities of the abuse of AI is quite broad, several scenarios are conceivable. One such possibility is the hacking of AI technology in use by national authorities or organizations, leaking information which can then be used for terrorist purposes. As society is becoming increasingly data-driven and sensitive data is collected by both governments and private actors alike, the probability of compromising and damaging data breaches occurring increases in a manner proportionately.

Another very prominent possibility is the hacking of autonomous vehicles. As far back as the early 2000s, the potential to hack cars has been apparent, with researchers demonstrating the ability to hack into a Ford Escape and disable its brakes in 2013 and bring a Jeep Cherokee to a standstill in the middle of a US Interstate Highway in 2015.[235] As AI continues to make its way into the automobile industry and cars become increasingly autonomous and connected, the opportunities for malicious actors to hack one or more of the hundreds of millions of lines of code that go into these vehicles significantly increase.

232 Vincenzo Ciancaglini, Craig Gibson, David Sancho, Philipp Amann, Aglika Klayn, Odhran McCarthy and Maria Eira. (Nov. 19, 2020). Malicious Uses and Abuses of Artificial Intelligence. Trend Micro, EUROPOL and UNICRI. Accessible at http://unicri.it/sites/default/files/2020-11/AI%20MLC.pdf.

233 Europol. (Oct. 9, 2019). Internet Organised Crime Threat Assessment (IOCTA). Europol. Accessible at https://www.europol.europa.eu/activities-services/main-reports/Internet-organised-crime-threat-assessment-iocta-2019

234 Alex Marquardt, Eric Levenson and Amir Tal. (Feb. 10, 2021). Florida water treatment facility hack used a dormant remote access software, sheriff says. CNN. Accessible at https://edition.cnn.com/2021/02/10/us/florida-water-poison-cyber/index.html

235 Andy Greenberg (Jul. 21, 2015). Hackers Remotely Kill a Jeep on the Highway—With Me in It. Wired. Accessible at https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/

Still within the domain of autonomous vehicles, further abuse of AI could include abusing image recognition systems. As noted, the machine learning models on which autonomous vehicles rely are dependent on the accuracy of the information they receive. If this information is compromised, so too is the vehicle itself. In early 2020, hackers tricked the autopilot system in two different models of Tesla into accelerating up to 85 mph instead of 35 mph by sticking a tape on a speed limit sign.[236] Similarly, by placing small stickers on the road, researchers also managed to careen another Tesla into the oncoming traffic lane.[237] The malicious potential for terrorist groups or individuals to create chaos with such kind of abuses is self-evident.

Similar techniques could also be leveraged in connection with an attack in order to lead people into targeted areas or to delay the arrival of security forces and/or emergency services to a venue in the aftermath of an attack, thereby amplifying its effects. One recent study demonstrated that even a relatively small-scale hack of self-driving cars would be sufficient to cause collisions and gridlock.[238] The study found that by hacking 10-20% of vehicles at rush hour in Manhattan, it would be possible to render half the city virtually inaccessible.

Photo by Sajjad Ahmadi on Unsplash

A further abuse worth mentioning is the possibility of malicious actors interfering with services or applications that employ AI by modifying the parameters that the system employs or poisoning the data sets used to train the system by feeding them incorrect data. In doing so, malicious actors could succeed to steer an AI system in a desired direction or, for instance, generate erroneous or biased outputs. For example, in 2016, a new Microsoft machine learning chatbot known as "Tay" was prematurely shut down after it began tweeting inflammatory and offensive tweets shortly after its release.[239] The machine learning function of the chatbot had been maliciously targeted, with Tay being intentionally fed racist, misogynistic, and anti-Semitic language in a coordinated manner in order to influence how the chatbot would vocalize itself publicly on Twitter. A further simple, yet effective, example of the potential of interfering with the

236 Isobel Asher Hamilton. (Feb. 19, 2020). Hackers stuck a 2-inch strip of tape on a 35-mph speed sign and successfully tricked 2 Teslas into accelerating to 85 mph. Business Insider. Accessible at https://www.businessinsider.nl/hackers-trick-tesla-accelerating-85mph-using-tape-2020-2?international=true&r=US

237 Karen Hao. (Apr. 1, 2019). Hackers trick a Tesla into veering into the wrong lane. MIT Technology Review. Accessible at https://www.technologyreview.com/2019/04/01/65915/hackers-trick-teslas-autopilot-into-veering-towards-oncoming-traffic/

238 Jamie Carter. (Mar 5, 2019). Hacked Driverless Cars Could Cause Collisions and Gridlock in Cities, Say Researchers. Forbes. Accessible at https://www.forbes.com/sites/jamiecartereurope/2019/03/05/hacked-driverless-cars-could-cause-collisions-and-gridlock-in-cities-say-researchers/?sh=5fe14fbd2a09

239 Oscar Schwartz. (Nov. 25, 2019). In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation. IEEE Spectrum. Accessible at https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation

proper functioning of AI-based systems took place in February 2020, when a German artist tricked Google Maps into believing that the traffic on the streets of Berlin was higher than it was in reality by feeding incorrect data to Google Map's machine learning models.[240] The artist succeeded to do so by carrying 99 mobile phones around with him as he walked the streets, which Google Maps incorrectly understood as presenting people in their cars, causing the system to run awry. Again here, the malicious potential is evident. Such tricks or targeted exploitations could be powerful tools to create chaos and confusion when leveraged maliciously. For example, in urban areas where people are increasingly dependent on such map applications to generate their routes, terrorists creating fake traffic congestion data could lead crowds into designated areas before executing their attacks or once again prevent the early arrival of security forces and/or emergency services to the venue of their attacks.[241]

Hence, when considering the malicious use of AI for terrorist purposes, it is also relevant to consider the other side of the coin: the malicious abuse of AI. In contrast to the malicious use of AI by terrorists, the threat of abuse is perhaps more within the existing capabilities of terrorist groups and individuals and could be extremely effective in augmenting existing attack methods and contributing to "terrorism as theatre".

IX. FROM ASSESSMENTS TO ACTION

While the use of AI for terrorist purposes is certainly not a developed threat, terrorism is far from stagnant. Currently, the technical capability of terrorist groups and individuals to deploy technologies such as AI may be considered low, but it is important to not underestimate their intention to avail of the latest technological trends, as well as their increasing capabilities to do so. As AI and related technologies become ever-more accessible to the public, it becomes incumbent upon those responsible for counter-terrorism to stay ahead of the curve. At the same time, even if AI-enabled terrorism may not be an imminent threat, it is imperative to remain vigilant of the potential abuse of AI systems by terrorist groups and individuals. This is an increasingly concerning angle, given the rate at which AI is being integrated into processes in both the public and private sector, including critical infrastructure.

In view of this, the following recommendations are provided for counter-terrorism bodies and law enforcement agencies, as well as policy-makers, industry and academia to consider for the future, and to guide follow-up actions for capacity-building to prepare for the possible future of AI-enabled terrorism. These recommendations were compiled and categorized on the basis of the feedback collected from participants at the UNCCT/UNOCT-UNICRI Expert Group Meeting. The order of recommendations should not be interpreted as indicating any particular priority.

Photo by Andrew Neel on Unsplash

240 Brian Barrett. (Mar. 2, 2020). An Artist Used 99 Phones to Fake a Google Maps Traffic Jam. Wired. Accessible at https://www.wired.com/story/99-phones-fake-google-maps-traffic-jam/

241 Simone Raponi, Savio Sciancalepore, Gabriele Oligeri, Roberto Di Pietro. (2020). Fridges on the Highway: Road Traffic Poisoning of Navigation Apps, arXiv preprint arXiv:2002.05051.

Further Research

- The evolution of the adoption of AI by terrorist groups and individuals should be monitored.
- Additional consultations within the research community on the basis of this report should be carried out in order to obtain further evidence and feedback so as to guide the prioritization of future research.
- The potential threat of cyber-attacks by terrorist groups or individuals on AI systems or on the integrity of data used in AI systems, particularly in the context of critical infrastructure, should be further assessed.
- The legal aspects surrounding the malicious use or abuse of AI should be reviewed and analyzed.
- The convergence of AI with other technological advancements, including biotechnology, brain-computer interface and the extraction and manipulation of data should be further explored.

Multi-stakeholder Cooperation

- The range of stakeholders involved in discussions regarding the use of AI for terrorist purposes should be expanded to all levels and all regions.
- Dialogue and cooperation between technical and non-technical experts should be promoted and sustained.
  - National authorities and policymakers should engage closely with the AI research community, including to help inform the drafting and implementation of protective measures and policies and to improve foresight activities.
  - Discussions amongst AI practitioners related to the pertinent challenges presented by AI should be taken beyond just the technology industry and involve all stakeholders, including civil society organizations, human rights experts and gender advisors.

Awareness-raising and Knowledge-building

- Efforts to raise awareness of governments and industry partners about the potential malicious uses and abuses of AI-enabled tools for terrorist purposes should be enhanced, as this awareness is key to ensure a timely response to possible threats.
  - There should be close engagement with the AI research community in raising awareness of the possible malicious uses of technologies being developed.
  - Knowledge and awareness within the research community should be built from the earliest stages of tech development, targeting, for instance, the student community and researchers seeking grants.
- Literacy of policymakers on AI technology, including the potential malicious uses and abuses should be developed.
  - Care should be taken to promote caution, while avoiding over-hyping the level of threat and the nature of threat scenarios.

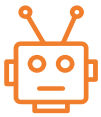
Capacity-building

- The capacity of all stakeholders to identify and respond to the threat of the malicious use and abuse of AI for terrorist purposes should be improved.
- Activities to streamline cooperation and coordination and the sharing of experiences between stakeholders should be organized.

Policies and Guidance

- Clear policies and practical guidelines on how to respond to AI-enabled attacks should be considered and developed by states and organizations, to ensure appropriate and adequate responses measures for such attacks, which are in line with the values enshrined in the Charter of the United Nations, the Universal Declaration of Human Rights and the norms and standards of international law.

- Regulation and certification processes to ensure AI systems are secured against adversarial use and provide accountability in case they are misused should be explored.



Leveraging AI in Countering Terrorism

- The use of AI and related emerging technologies to counter AI-enabled terrorist threats should be explored, in particular to counter terrorist radicalization and spread positive narratives.

- A comprehensive and in-depth mapping of the intersection between AI and counter-terrorism should be undertaken.

- Human rights should be at the centre of any such use of AI to counter terrorism, including countering the malicious use and abuse of AI for terrorist purposes.