CCDCOE
NATO COOPERATIVE
CYBER DEFENCE
CENTRE OF EXCELLENCE

Generative Adversarial Networks from a Cyber Intelligence perspective

Fabio BIONDI
NATO CCDCOE, Researcher

Giuseppe BUONOCORE
ITA JCNO, Researcher

Richard MATTHEWS
RHEM LABS, Researcher

Tallinn 2021

About the authors

Fabio BIONDI

Fabio Biondi is a Lieutenant Colonel (OF-4) in the ITA Air Force, currently on duty at CCD COE Tallinn as researcher and Director of the Operational Cyber Int. Course. He enlisted in the AF in 1988 and from the beginning his specialty was IT. He was a programmer on IBM mainframe platforms and a systems analyst. He was in charge of the Service Desk for the AF for several years, being after that Group CO. Joined the NATO Programming Centre in Glons (BE), becoming Project Manager of the NATO Integrated Command and Control System. Returning to the ITA AF, he was appointed Section Head of NATO C2 Systems in the ITA AF Logistic Command. He took part in the NATO Unified Protector and EU Sophia operations. His last assignment in Italy was in Joint Ops Command as CIS coordinator for the joint exercises of the ITA Armed Forces. He is an advisor for the IHL-International Humanitarian Law, Red Cross certified. He has a Bachelor's degree in Administration and Organisation Science and a master's degree in Management and Corporate Communication. Married to Paola, has one son, Lorenzo.

Giuseppe BUONOCORE

Giuseppe Buonocore is a Chief (OR-8) in the ITA Navy, currently on duty at the ITA Joint Command for Network Operations as a cybersecurity expert and researcher in the Cyber Operations Department. After computer science training focused on software design and development, he worked for several years on intelligent combat systems in submarine and cyber warfare. He is specialised in open-source intelligence, social media intelligence and cyber operations. He is currently engaged in the exploitation, penetration testing and reverse engineering of IT systems, and digital forensic acquisition and analysis. He has been studying the use of neural systems by cyber threats, paying particular attention to the development of cyber weapons. He has been involved in the production and management of the command and control of digital systems. He took part in NATO's Operations Unified Protector and Ocean Shields and Exercises Locked Shields and Crossed Swords and EU MilCERT exercises to test and improve the capabilities of ITA Cyber Command. He has Bachelor's degrees in political science and international relations and the science and management of maritime activities and a Master's degree in cyber defence governance. Cohabiting with Filomena, has one son, Diego.

Richard MATTHEWS

Dr. Richard Matthews is a systems engineer practising digital image forensics, cyber and information risk. He holds a PhD in image forensics and a Bachelor of Engineer (Electrical and Electronic) from the University of Adelaide. Dr. Matthews also holds several certificates in cybersecurity and intelligence from the Tallinn University of Technology. In private practice, Dr. Matthews is the Managing Director of RHEM Labs. He provides expert witness testimony on most topics related to technology. He has worked in diverse fields including the military, governance and academia. His adaptive leadership style brings diverse, dissenting views together on problems that have no well-defined solutions. In his TEDx 'How safe is the Internet?' Dr. Matthews set a bleak look at our use of technology and how businesses must consider what we post online. His research focuses on the previously disjointed interdisciplinary fields of electronic engineering, computer science and forensic science. He has an extensive publication record including professional journals, science communication, radio, television and print media. His work has been translated into Arabic, Indonesian, German and Estonian. Dr. Matthews holds professional memberships of the Australian Institute of Company Directors, the Institute of Electrical and Electronics Engineers, the Australian & New Zealand Forensic Science Society and the Institute of Engineers Australia.

The CCDCOE

The NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) is a NATO-accredited cyber defence hub focusing on research, training and exercises. It represents a community of most NATO nations and partners of the Alliance across the globe. The Centre provides a comprehensive cyber defence capability, with expertise in the areas of technology, strategy, operations, and law. The heart of the Centre is a diverse group of international experts from military, government, academia and industry backgrounds.

The CCDCOE is home to the Tallinn Manual. For a decade, the Tallinn Manuals have served as an essential academic tool for policy and legal experts on how international law applies to cyber operations. In 2021, the CCDCOE launched the Tallinn Manual 3.0 Project that will revise and expand the current 2017 edition, Tallinn Manual 2.0, to include recent State practice and policy statements. In addition to cyber defence exercises, the Centre hosts CyCon, the International Conference on Cyber Conflict. Every spring, this unique event joins key experts and decision-makers from the global cyber defence community. CCDCOE also offers a wide range of courses on a variety of subjects related to cyber defence that attract attendees from a wide range of nationalities and backgrounds.

The Centre is staffed and financed by its 29 member nations – Austria, Belgium, Bulgaria, Croatia, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Montenegro, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States.www.ccdcoe.org
publications@ccdcoe.org

Disclaimer

This publication is a product of the NATO Cooperative Cyber Defence Centre of Excellence (the Centre). It does not necessarily reflect the policy or the opinion of the Centre or NATO. The Centre may not be held responsible for any loss or harm arising from the use of the information contained in this publication and is not responsible for the content of the external sources, including external websites referenced in this publication.

Digital or hard copies of this publication may be produced for internal use within NATO and for personal or educational use when for non-profit and non-commercial purpose, provided that copies bear a full citation.

Table of Contents

1. Abstract

Generative adversarial networks (GAN) are a hot topic in cyber intelligence, as they begin to demonstrate abilities that will assist the public intelligence analyst to play a more active role in global security. Not only can they help you sort through the vast OSINT sources of material to identify potential threats, but additional features are also now being demonstrated which will allow links to be drawn in real time between potential threats.

This is an incremental report produced for the NATO Cooperative Cyber Defence Centre of Excellence, Cyber Intelligence division, under agreement. It focuses on preliminary work addressing a wider project to produce a paper on the cyber intelligence uses of GAN.

The work has two main focuses:

- to investigate different applications to assess the most interesting and promising in respect of cyber intelligence; and

- to determine a framework to assess current examples of GAN that offer solutions relevant to cyber intelligence.

To facilitate this, some questions need to be addressed:

(1) what is a GAN?

(2) what is cyber intelligence?

(3) how do the two interact?

The focus of the project is to assess the potential use of GAN in the context of cyber intelligence, the unified kill chain model and how ready such technology is for deployment; both for legal uses and illegal. This raises certain questions:

(1) How might GAN be used to assist in the collection, validation, exploitation or reporting of intelligence information?

(2) How might GAN be used to assist in gaining initial footholds in systems, propagate or pivot and finally achieve actions on an objective?

(3) How has GAN has been used traditionally, what opportunities does this present to the intelligence community (IC) and are these uses legal?

This focus is driven by two approaches, those that are collection-orientated and those that are offensive or defensive towards the cyber threat. The framework we will design will incorporate an assessment for both.

2. Introduction

Generative adversarial networks (GANs) are a deep-learning model first described by Ian Goodfellow in 2014.[1] They use two neural networks – one that creates content and one that analyses it – in a pseudo-game-like adversarial process. To understand what a GAN is, first we must understand some fundamental principles of machine learning. Machine learning is 'a subset of artificial intelligence (AI) where an artificial intelligence platform is trained to make predictions based on data'. The goal is to build algorithms that can 'learn' to make predictions based on new data to which it has access. This data forms training sets from which algorithms can correlate future examples based on prior inputs.

To showcase the power of artificial intelligence, the entire prior paragraph was not written by the authors which is why the definition of machine learning contains no citation. Instead, a neural network[2] was used to complete the passage based on a few seed words (the training set) provided to the network and then manually edited for proof and clarity. Indeed, the first paragraph of the previous section was written in the same manner. This is one power of GAN which has yet to be fully realised; the widespread automation of report writing using GAN tools as the first pass.

In this work, we use the definition of machine learning from Shalev-Shwartz and Ben-David as follows: 'machine learning refers to the automated detection of meaningful patterns in data'.[3] The common models used in machine learning can be divided into two categories: supervised and unsupervised. In supervised learning, models are trained on well-labelled data sets and then tested against the same. This is best suited to regression or classification problems. 'In the predictive or supervised learning approach, the goal is to learn a mapping from inputs x to outputs y, given a labelled set of input-output pairs'.[4]

Unsupervised learning is useful when perfectly labelled data is not available to the machine. A deep learning model is used in conjunction with a dataset without predefined labels for what the machine is supposed to accomplish:

> 'Here we are only given inputs, and the goal is to find 'interesting patterns' in the data. […] This is a much less well-defined problem, since we are not told what kinds of patterns to look for, and there is no obvious error metric to use (unlike supervised learning, where we can compare our prediction of y for a given x to the observed value)'.[5]

GANs change the definition given above just slightly by using both supervised and unsupervised components. Instead of relying upon training sets of known data, the network is based on a game theory scenario where two neural networks compete with one another. The first neural network, a generator, will compete against a discriminator. The generator is responsible for creating samples (unsupervised) to intertwine within the training set (supervised). The discriminator is responsible for determining if the example it is presented with is from the training set of real data or if it has been given a sample from the generator. This output of this network is a sample that has been created by the generator and which passes the test of the discriminator. This layout is shown in Figure 1. The generator attempts to fool the classifier into believing its samples are real, indistinguishable from the real learning data.[6]

[1] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial networks. arXiv preprint arXiv:1406.2661.

[2] InferKit, 2020. Talk to Transformer. [Online] Available at: https://app.inferkit.com/demo [Accessed 30 09 2020].

[3] Shalev-Shwartz, S. and Ben-David, S., 2014. Understanding machine learning: From theory to algorithms. Cambridge university press.

[4] Robert, C., 2014. *Machine learning, a probabilistic perspective*. Vancouver, p.2.

[5] Ibid.

[6] Goodfellow, I., Bengio, Y. & Courville, A., 2016. Deep Learning. Online ed. s.l.: MIT Press.

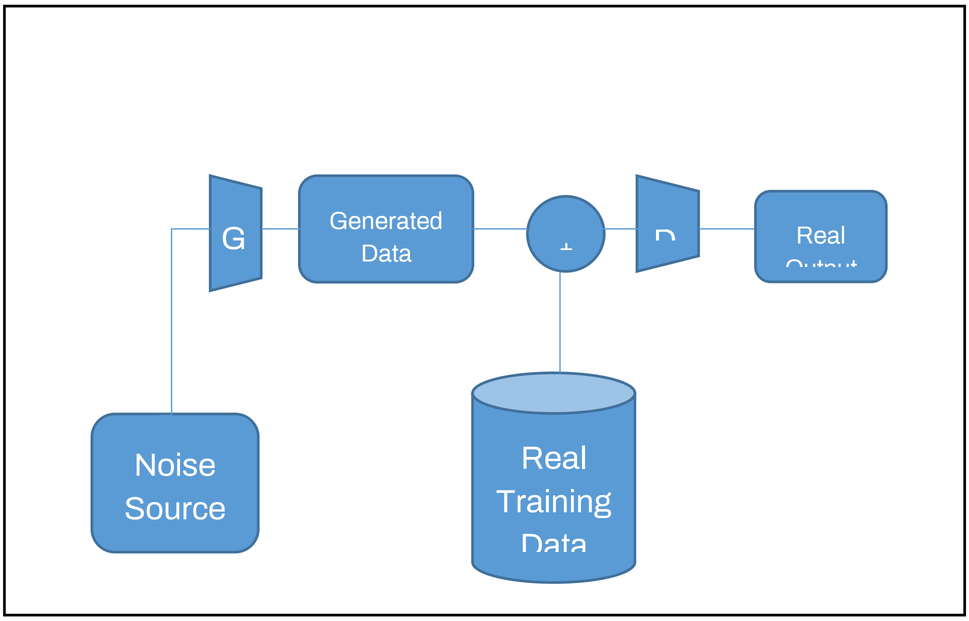


FIGURE 1: THE LAYOUT OF A GAN. THE GENERATOR (G) AND THE DISCRIMINATOR (D) ARE IMPLEMENTED USING NEURAL NETWORKS.

This approach is best explained using Goodfellow's counterfeiter analogy:

> The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect counterfeit currency. Competition in this game drives both teams to improve their methods until the counterfeits are indistinguishable from the genuine articles.[7]

GANs work well with image synthesis. The literature suggests that they are most suited to deep fakes which is a type of multimedia where the file is manipulated to appear as real as possible. This is not the only application of GAN technology, however. Other applications have been postulated to include image-to-image translation, text-to-image translation, artificial face ageing and enhanced super-resolution.[8] Most often, applications of GAN are within the multimedia space as image and video data is well suited to deep learning methods. Applications within the intelligence space use not just GAN but other machine learning algorithms. In this report, we consider machine learning applications within the cyber intelligence domain, but restrict our assessments to those which use GAN. Further work should be conducted to expand this study into cyber intelligence applications using any machine learning method, not just GAN.

[7] Goodfellow, 2014, p.1.

[8] Brownlee, J., 2019. 18 Impressive Applications of Generative Adversarial Networks (GANs). [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/impressive-applications-of-generative-adversarial-networks/> [Accessed 3 February 2021].

2.1 Intelligence

There are many definitions of intelligence. Clark (1955) defined it as the discipline which deals with all the things which should be known in advance of initiating a course of action.[9] The US Army defines intelligence using a three-way definition:

> 'Intelligence is: (1) the product resulting from the collection, processing, integration, evaluation, analysis, and interpretation of available information concerning foreign nations, hostile or potentially hostile forces or elements, or areas of actual or potential operations; (2) the activities that result in the product; and (3) the organisations engaged in such activities'.[10]

We see here that intelligence must focus on a specific process regardless of discipline. Based on this definition we see that the process can be broken down into five stages: planning and direction, collection, processing, analysis and production, and dissemination. The mere act of collecting information within a specific collection discipline does not immediately make it intelligence. It is this process through which the information is collected and then analysed which creates an actionable intelligence item. We use these phases to assist in defining our framework below.

Fundamentally the intelligence community (IC) – those organisations which are engaged in the collection of intelligence – define intelligence by five core disciplines: open-source intelligence (OSINT); human intelligence (HUMINT); signals intelligence (SIGINT) geospatial intelligence (GEOINT) also referred to as image intelligence (IMINT); and measurements and signature intelligence (MASINT).[11] Cyber intelligence is not defined as a primary collection discipline.

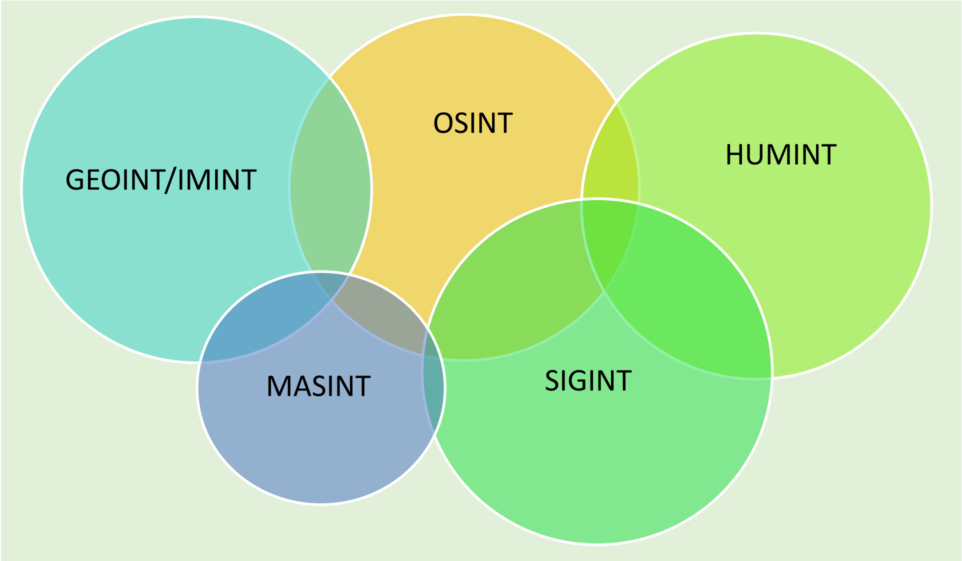


Figure 2: The five primary collection disciplines of the intelligence community. The overlapping nature of the disciplines is shown to demonstrate an example of the blurring between domains only.[12]

[9] Clark, M (1955), 'Intelligence activities'. Interim technical report to Congress. Commission on Organisation of the Executive Branch of the Government [the Hoover Commission].

[10] United States of America, 2019. ADP 2-0 Intelligence, Department of the Army. 1-1.

[11] Lowenthal, M. & Clark, R., 2016. The Five Disciplines of Intelligence Collection. London: SAGE.

[12] Williams, H. J. & Blum, I., 2018. Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise, Santa Monica: RAND National Defense Research Institute.

OSINT

Open-source intelligence, or OSINT, is often described as the source of first resort due to its ease and low cost of gathering, collecting and analysing. It includes 'information that is publicly available to anyone through legal means, including request, observation, or purchase, that is subsequently acquired, vetted and analysed to fulfil an intelligence requirement'.[13] It is often used as a pivot point for other closed intelligence sources where the information obtained can be merged to create all-source intelligence reports. Many misbelieve that OSINT is just the use of social media on the internet in the modern age, however, its use is much broader and is not confined to just this one source. Here, the most use can be obtained, especially when it is vetted for misinformation and disinformation. Disinformation is a problem with OSINT sources as many nation-states now use this channel to increase the signal-to-noise ratio with useless information. As a result, information from OSINT must be vetted against other sources, most often HUMINT sources. The key to OSINT is that all information must be legally obtained. While this seems straightforward, it is fraught with legal issues due to growing privacy concerns including GDPR and the 14th Amendment to the US Constitution. As a result, any seizure of OSINT must be specifically defined in an agency's requirement.

HUMINT

When we think of intelligence collection, HUMINT is what often comes to mind. HUMINT is the collection discipline that involves the use of human sources as a primary collection means.[14] It is the source most often portrayed in Hollywood dramatisations. Unlike OSINT, HUMINT may involve the use of both overt and covert collection methods when searching for information.[15] However, unlike OSINT, information is not the primary goal of HUMINT; secrets are.[16]

SIGINT

The collection discipline most like cyber intelligence is SIGINT. This is the discipline that deals with information derived from electronically transmitted data and information.[17] This data includes that derived from electronic intelligence (ELINT), radio and wireless transmission and at times telemetry and other instrumentation data included in foreign instrumentation signal intelligence (FISINT).[18] As a matter of process, SIGINT also includes encrypted communications; the code making and code breaking agencies.[19] Along with ELINT, communications intelligence (COMINT) is another subdiscipline of SIGINT. COMINT is the focus on communications within the SIGINT domain.[20] Due to the historical use in monitoring foreign radio broadcasts, COMINT also encompasses the translation of foreign language broadcasts. As technology progressed in the early 20th century from radio transmissions to telephone services and now the internet, we see the basis for our sub-discipline of cyber intelligence, but we do not see a unique need for a complete field in and of itself. Cyber intelligence is just another tool within which SIGINT lives and grows, as technology itself grows. Only the future will determine if the resource allocation to cyber is sufficient and distinctive enough to make it a discipline in its own right.

[13] Lowenthal, M. & Clark, R., 2016.
[14] Ibid.
[15] Ibid.
[16] Ibid.
[17] Ibid.
[18] Ibid.
[19] Ibid.
[20] Ibid.

IMINT/GEOINT

GEOINT has its roots in mapmaking, cartography and the use of aerial photography during World War I.[21] The official definition of GEOINT originates in 2003 from the US Code, Title 10, section 467:

> The term 'geospatial intelligence' means the exploitation and analysis of imagery of geospatial information to describe, assess, and visually depict physical features and geographically reference activities of the earth. Geospatial intelligence consists of imagery, imagery intelligence, and geospatial information.[22]

Subdomains of GEOINT include location-based intelligence (LBI) and the use of global positioning system (GPS) technology. The use of GPS has recently been seen within the literature as part of offensive and defensive[23] cyber capabilities due to the ability to affect the reliability of the signal.[24] As such, we see that even in this basic map-based intelligence field, cyber has applications. It is unclear how image intelligence that is unrelated to geospatial information remains to be extracted from the GEOINT definition. It is for this reason that image analysts will often be employed from GEOINT domains to assess raw images that are not necessarily in the geospatial domain due to the overlapping skills required in signal processing.

MASINT

There are two approaches to defining MASINT; one by what it is and one by what it does. In the mid-1990s the first definition emerged:

> Measurement and Signature Intelligence (MASINT) is technically-derived intelligence that enables detection, location, tracking, identification, and description of unique characteristics of fixed and dynamic target sources. MASINT embodies a set of sub-disciplines that operate across the electromagnetic, acoustic and seismic spectrums, and material sciences. MASINT capabilities include radar, laser, optical, infrared, acoustic, nuclear radiation, radio frequency, spectro-radiometric, and seismic sensing systems as well as gas, liquid, and solid materials sampling and analysis. MASINT is an integral part of the all-source collection environment and contributes both unique and complementary information on a wide range of intelligence requirements. MASINT is highly reliable since it is derived from the performance data and characteristics of actual targets.[25]

A second definition from the US Department of Defence is based around what MASINT does:

> Information produced by quantitative and qualitative analysis of physical attributes of targets and events to characterise, locate, and identify them. MASINT exploits a variety of phenomenologies to support signature development and analysis, to perform technical analysis, and to detect, characterise, locate, and identify targets and events. MASINT is derived from specialised, technically-derived measurements of physical phenomena intrinsic to an object or event and it includes the use of quantitative signatures to interpret the data.[26]

We see here that MASINT heavily relies on signal processing and overlaps SIGINT with the difference being in the signatures obtained rather than the information itself. MASINT can be further broken down into subdisciplines, but these create confusion as we see significant overlap with disciplines such as GEOINT and SIGINT when the inclusion of radar, radio frequency and geophysical data is included in

[21] Ibid.

[22] 10 U.S.C. §467, 2003.

[23] Crossing, I., Heading, B., Hilliard, J., Page, L., Sorell, M. And Mathews, R., Techniques For GPS Spoofing Detection On Android Devices. In Cyber Research Conference 2020, p.13.

[24] Norman, S., Shelby-James, L., Matthews, R. and Sorell, M., 2019. Reliability and Trust in Global Navigation Satellite Systems (Honours Thesis, University of Adelaide).

[25] John Morris, 'MASINT', American Intelligence Journal 17, no. 1 & 2 (1996): 24–27.

[26] DoD Instruction number 5105.58, April 22, 2009, http://www.dtic.mil/whs/directives/corres/pdf/510558p.pdf.

the MASINT definition. The key feature remains the information obtained from this source. Where SIGINT can be likened to the ears and GEOINT the eyes, MASINT can be thought of as the other senses in the human body used to obtain the full picture.[27]

Looking towards the future, it is MASINT that would likely stand to benefit the most from the expanded use of machine learning (including the use of GAN) due to the vast data matching involved in identifying signatures from sensors. It is this type of processing that machine learning is acutely matched to performing.

2.2 Cyber Intelligence

Cyber intelligence is not a clearly defined discipline within the five. While cyber is an emerging threat, the majority of intelligence activity in this space can be categorised as an overlap of the other fields. This does not mean there is no room to define cyber intelligence as a new field. It just means it has not yet been done. Cyber is a battlespace in which the intelligence activities defined above can play. No definition of cyber intelligence is to be found in the literature. This is an indication that the community has struggled either to define this area well or is relying on the implicit link between the current definition of intelligence and the assumed cyber threat.

The Intelligence and National Security Alliance has developed the landscape for cyber intelligence in a series of papers. No definition of cyber intelligence has been provided; instead references to the implicit link between the IC and the use of the cyber threat have been made with a better definition of cyber intelligence as a new discipline in the IC, a key conclusion.[28] This suggestion has not been quickly adopted. By 2016, there were still only five recognised disciplines of intelligence.[29] The establishment of cyber intelligence as a sixth is an ongoing challenge.

Attempts have been made at defining cyber intelligence in the literature. Eom[30] defined it as:

> the product resulting from the collection, processing, analysis, integration, evaluation, and interpretation of available data concerning hostile cyber organisation, cyber forces capabilities, network system, hardware, software, data, threats, vulnerabilities, and so on.

Mattern et al.[31] contradict the product-based definition by stating that:

> Cyber Intelligence is not a collection discipline like signals intelligence (SIGINT) or open-source intelligence (OSINT). Instead, similar to 'medical intelligence,' it is more of an analytic discipline relying on information collected from traditional intelligence sources intended to inform decision-makers on issues pertaining to operations at all levels in the cyber domain.[32]

Mandt (2017)[33] notes that the balance of power within the cyber domain has largely been left unchanged since the Morris worm in 1988 with a 'continuous state of insecurity' existing in the cyber domain. They

[27] Lowenthal, M. & Clark, R., 2016.

[28] Fast, B., Johnson, M. & Schaeffer, D., 2011. Cyber Intelligence: setting the landscape for an emerging discipline, s.l.: Intelligence and National Security Alliance.

[29] Lowenthal, M. & Clark, R., 2016.

[30] Daehakro, D.G., 2014. Roles and responsibilities of cyber intelligence for cyber operations in cyberspace. International Journal of Software Engineering and Its Applications, 8(9), pp.137-146.

[31] Mattern, T., Felker, J., Borum, R. & Bamford, G. 2014. Operational Levels of Cyber Intelligence, International Journal of Intelligence and Counterintelligence, 27:4, 702-719, DOI: 10.1080/08850607.2014.924811.

[32] Ibid. 704.

[33] Mandt, E., 2017. Integrating Cyber-Intelligence Analysis and Active Cyber-Defence Operations. Journal of Information Warfare, 16(1), pp.31-48.

remark that finding such a definition for cyber intelligence is still a continuous challenge of the community as we struggle to transition to an 'intelligence-driven active cyber-defence posture'.[34]

It appears that the work of Fast et al. has not been widely adopted. Current references still refer only to the main categories of intelligence as HUMINT, SIGINT, IMINT and MASINT with OSINT as the foundation for all-source analysis.[35] This failure raises several important issues, including whether cyber intelligence is required in the current disciplines as recognised by the IC or whether it is already covered by intersections of the five recognised areas. This is beyond the scope of this work; however, it is recommended that such analysis be undertaken to understand the role cyber intelligence plays in the current landscape.

2.3 Defining Cyber Intelligence (CYBINT)

There is a need to not only define cyber intelligence, but also to situate it within the currently recognised IC. It is not the role of this report to define what cyber intelligence is; it is clear that this is a far larger task that must be completed in collaboration with the IC and the cybersecurity profession. Without defining an entirely new community (CYBINT), we will instead opt to use a home-grown definition in this report.

We define cyber intelligence as:

> intelligence activities focused on the collection, validation, exploitation and dissemination of information concerning the threat posed by an adversary in the cyber domain. This intelligence activity may or may not currently be encompassed by the overlapping of other formally recognised disciplines of the IC.

These key stages closely follow those highlighted in second-generation OSINT collection.[36] Cyber intelligence in the context of this report deals with both open and closed sources (legal and illegal applications) where the boundaries of intelligence collection are only restricted in so far as noting that this activity, like cyber conflict itself, cannot exist in a vacuum.[37] We simply define cyber intelligence as the bridging of intelligence disciplines to synthesis an actionable product in the cyber domain. This definition relies on the traditional intelligence cycle defined above and works to both create a product and serve to inform decision-makers at the policy and operational levels in the cyber domain. In this manner, we have synthesised the existing definitions without contradicting the current uses seen in the literature.

This definition of cyber intelligence is not contradictory to the definitions given in the CIA Intelligence Cycle.[38] It merges activities in the third and fourth stages of the Intelligence Cycle into a validation stage to allow more effort to be provided on the exploitation of the intelligence through analysis and contextualising. Neither does this definition play into the current discussion in the community regarding the position of cyber as a subdiscipline of SIGINT or a discipline in its own right. We do not aim to resolve these issues; rather, the working definition will allow us to work on the project focus of GAN in the context of cyber intelligence.

[34] Ibid.

[35] NATO Open Source Intelligence Handbook.

[36] Williams, H. J. & Blum, I., 2018.

[37] Fast, B., Johnson, M. & Schaeffer, D., 2011.

[38] Central Intelligence Agency, 2007. The Intelligence Cycle. [Online]
Available at: https://www.cia.gov/kids-page/6-12th-grade/who-we-are-what-we-do/the-intelligence-cycle.html
[Accessed 30 09 2020].

3. Fields of application and functionality

Machine learning technologies have been applied in various contexts to process information useful for predicting future events. In the vast majority of cases, the quality and quantity of information available largely determine the quality of the predictions, whether based on static or adaptive strategies. This implies that a substantial alteration of the informative base is a useful technique for interfering with the normal functioning of a system based on machine learning, consequently providing competitive advantages. In recent years, a research and application sector known as Adversarial Machine Learning has developed, capable of implementing different types of attack[39] to steal information and manipulate the evaluations of neural systems. In this research area, GANs[40] are widely used as they can generate outputs that reflect the distribution and nature of the examples used in the training phase of neural networks. However, although adversarial technologies represent a significant danger to intelligent systems, they also allow the development of strategies capable of countering them effectively based on available data. This elaboration process leads to the design of more robust and performing informative systems.

The ability to acquire information and manipulate it make GANs a valuable tool capable of meeting the particular requirements of the intelligence sector. In this chapter, we will analyse some of the fields of application that can be interesting for cyber intelligence purposes.

3.1 Counterterrorism

Terrorism is a social scourge that has been dogging mankind for many years and states must confront it to prevent the different types of attacks for which they may not be adequately prepared due to the vastness of the variables involved, such as the political and social context or the morphology of the scenario. The advent of computer vision has provided counter-back support which focuses on risk assessment based on regression models and probability theory. Artificial vision is a good candidate for supporting states in protecting their people from terrorism both by informing them about the consequences of a given attack and by defining the behaviours to be followed for their resolution. The first example is the prototype Adversarial Learning for counterterrorism (ALTer) framework[41] which tries to provide support by simulating complex terrorist scenarios to identify sensitive and vulnerable places and prepare a pre-planned response, thus offering the opportunity to predict, avoid or manage attacks.

ALTer uses GANs as a key design artifact to implement a proof-of-concept and simulate terrorist attacks through gamification because of the lack of data available about these scenarios and the sensitivity of the data. In particular, fragments are recorded in which criminal and terrorist activities are perpetrated in controlled play sessions and are saved with accurate data of the sitemap to create a synthetic dataset. From this dataset, StyleGAN is used to map the characteristics of terrorist attack scenes extracted in latent space and subsequently transfer the attacks to other locations. StyleGAN is a new architecture that combines GAN and AdaIN that can combine multiple features into a single feature using separate feature vectors for each training level. For example, it can generate a high-quality and realistic-looking third object using some obvious characteristics of one existing example and less obvious characteristics of another.

[39] Huang, L., Joseph, A., Nelson, B., Rubinstein, B., & Tygar, J. (2011). Adversarial Machine Learning.

[40] Goodfellow, J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). Generative Adversarial Networks.

[41] Cascavilla, G., Di Nucci, D., Slabber, J., Tamburri, D., Palomba, F., & van den Heuvel, W.-J. (2020). Counterterrorism for Cyber-Physical Spaces: A Computer Vision Approach.

ALTer analyses the factors to consider in the simulation of counterterrorism on public spaces and classifies them into three categories: environments, event and agents. For each, a set of factors is identified and extracted.

Environment

The environment reflects the place where the event takes place. In particular, the following factors are recognised:

- Location: 3D composition of the location of the main site, including all the buildings and access areas;
- # of Entries: number of places that people can enter the main site;
- # of Exits: number of places that people can exit the main site;
- # Access Control: number of secured entry points;
- # of Barriers: number of access points with barricades (including natural ones) to protect against the attack;
- Surveillance: if there are surveillance systems in place for early alerting against an attack.

Event

The event describes the type of terrorist scenario. In particular, the following factors are recognised:

- Activity: the type of event that is taking place;
- Attack: the type of attack;
- Crowd Density: how many people per square meter will be at the event.

Agents

The agents are all the individuals that we consider in the simulation. In particular, the following factors are recognised:

- Type: citizens, terrorists, police officers, firefighters, rescue personal, ambulances and hospitals;
- Groups: groups of agents (e.g., police, agents or families) should be together;
- Speed: the speed of the agent;
- Vision Radius: the distance in which an agent can detect another agent;
- Route: the route that the agent follows;
- Response Time: the response time of the agent.

Once the factors to be extracted have been identified, it is necessary to encode them so that they are understandable by a GAN and they can be used for its training according to the pipeline shown below which identifies two main process lines, one linked to the environment and one linked to events:

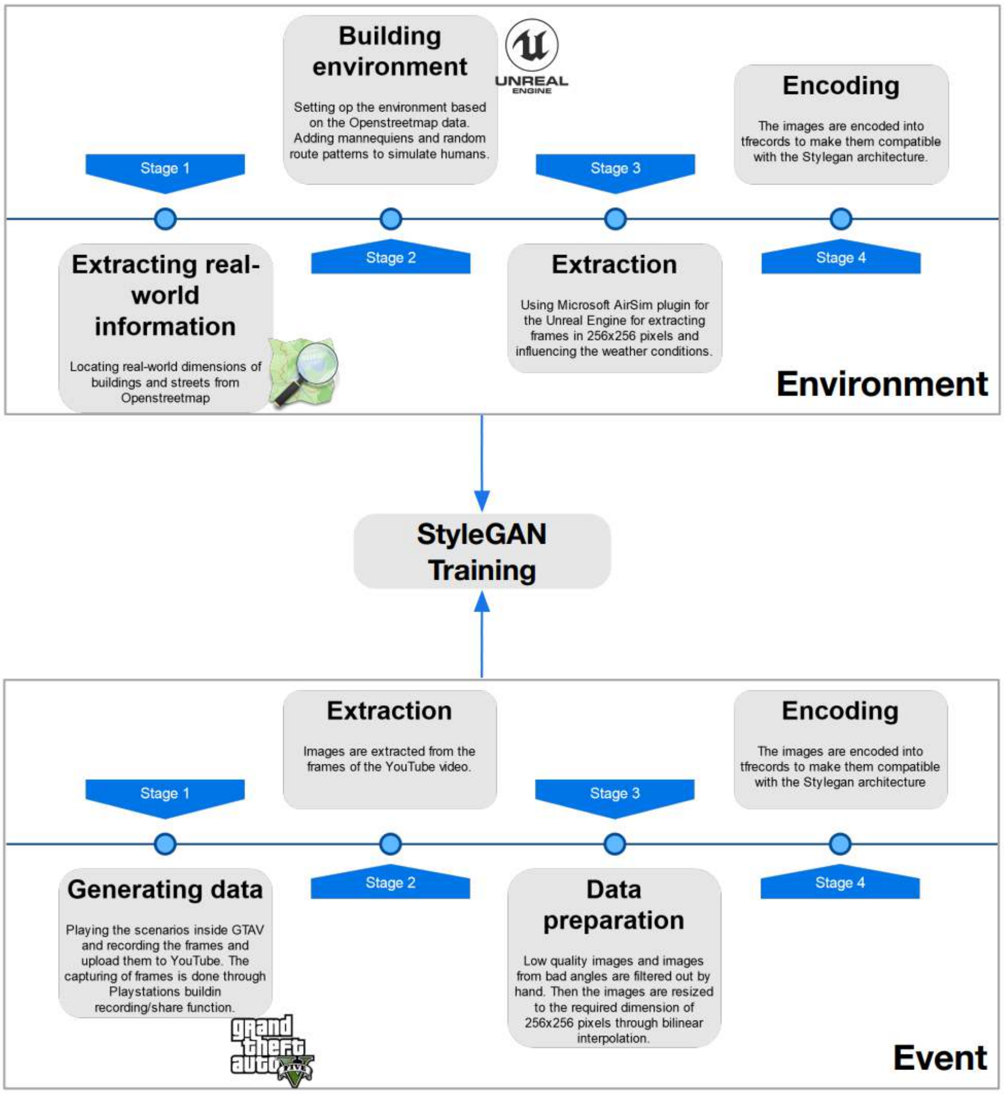

Building environment
Setting op the environment based on the Openstreetmap data. Adding mannequiens and random route patterns to simulate humans.
UNREAL ENGINE
Encoding
The images are encoded into tfrecords to make them compatible with the Stylegan architecture.
Stage 1
Stage 3
Stage 2
Stage 4
Extracting real-world information
Locating real-world dimensions of buildings and streets from Openstreetmap
Extraction
Using Microsoft AirSim plugin for the Unreal Engine for extracting frames in 256x256 pixels and influencing the weather conditions.
Environment
StyleGAN Training
Extraction
Images are extracted from the frames of the YouTube video.
Encoding
The images are encoded into tfrecords to make them compatible with the Stylegan architecture
Stage 1
Stage 3
Stage 2
Stage 4
Generating data
Playing the scenarios inside GTAV and recording the frames and upload them to YouTube. The capturing of frames is done through Playstations buildin recording/share function.
grand theft auto FIVE
Data preparation
Low quality images and images from bad angles are filtered out by hand. Then the images are resized to the required dimension of 256x256 pixels through bilinear interpolation.
Event

Figure 3: ALTer process flow

To evaluate the extent to which scenarios can be transferred from a simulated environment to their real-world cyber-physical counterparts, a real experiment was conducted in Malaga, Spain. In the experiment, it was simulated that the terrorist was lighting a fire in the main square of the city. From this experiment, it emerged that the training of GANs on cyber-physical spaces is challenging due to the stochastic variance of the real world, as the virtual world coded by programmers collides with the innumerable variables that require specific systems of approximation and hypothesis in the cyber-physical spaces of real life. However, the proof of concept showed the possibility of simulating real cyber-physical spaces as subject to terrorist threats, albeit not with a low margin of error. This approach, although still in its infancy, is worthy of further experimentation using architectural improvements, datasets that are as accurate as possible and data fusion approaches between real and simulated data. As part of this study, we show a new way to generate data for video game-based terrorism scenarios, thus addressing the lack of data and sensitivity issues on terrorism data.

3.2 Cross-Domain Relations

Cross-domain relations are natural to humans; for example, the sun and a tan. From an artificial intelligence perspective, this question can be reformulated as a conditional image generation problem, finding a mapping function from one domain to the other. However, pairing images can become tricky if corresponding images are missing in one domain or there are multiple best candidates.

Most of today's GAN's training approaches use explicitly paired data provided by humans or an algorithm, but DiscoGAN[42] pushes one step further by discovering relationships between two visual domains without any explicitly paired data. DiscoGAN is designed to discover relationships between two unpaired, unlabelled datasets. Figure 4 highlights the different implementations, distinguishing the standard GAN (a), the GAN with a reconstruction loss (b) and the DiscoGAN (c).

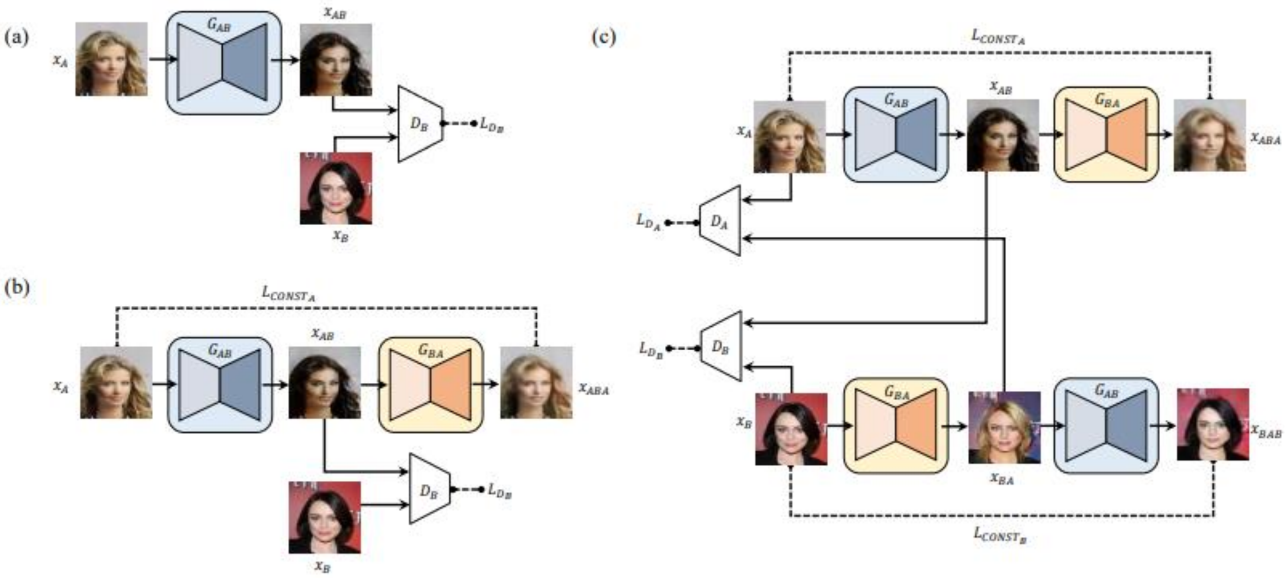


Figure 4: Difference between GAN and DiscoGAN

DiscoGAN can be trained with two sets of images without any explicit pair labels and does not require any pre-training. It takes one image in one domain as an input and generates its corresponding image in another. The model is based on two different GANs, each of them mapping each domain to its counterpart domain. It also learns the bidirectional mapping between two image domains, such as faces, cars, chairs, edges and photos, and successfully applies them in image translation. Translated images consistently change specified attributes such as hair colour, gender and orientation while maintaining all other components.

To empirically demonstrate the differences between the previous GAN models, an illustrative experiment based on synthetic data in 2-dimensional A and B domains can be used. Both source and target data samples are drawn from Gaussian mixture models.

[42] Taeksoo, K., Moonsu, C., Kim, H., Jung Kwon, L., & Jiwon, K. (2017). Learning to Discover Cross-Domain Relations with Generative Adversarial Networks.

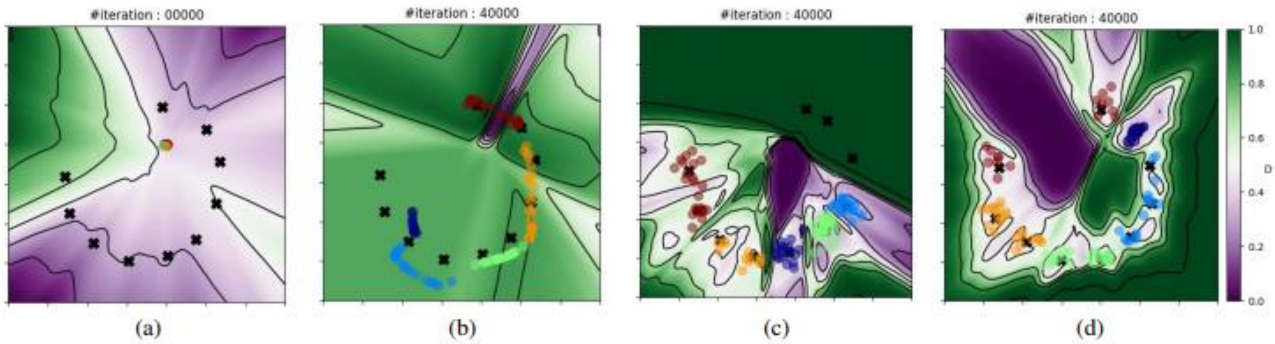
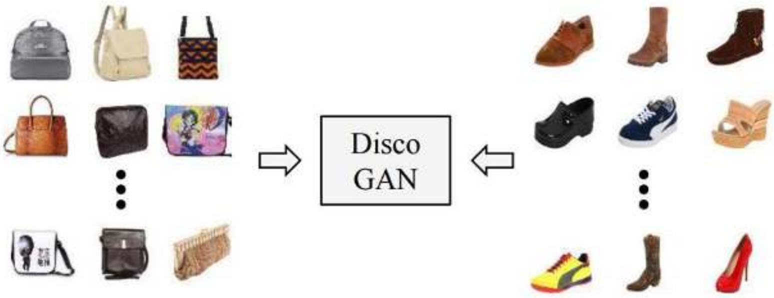


FIGURE 5: GANS DATA DISTRIBUTION

In Figure 5, there are four different elements for comparison: ten target domain modes and initial translations (a), standard GAN model (b), GAN with reconstruction loss (c) and DiscoGAN (d). Each of them is described by:

- The coloured background shows the output value of the discriminator;
- 'x' marks denote different modes in the B domain;
- Coloured circles indicate mapped samples of domain A to domain B, where each colour corresponds to a different mode.

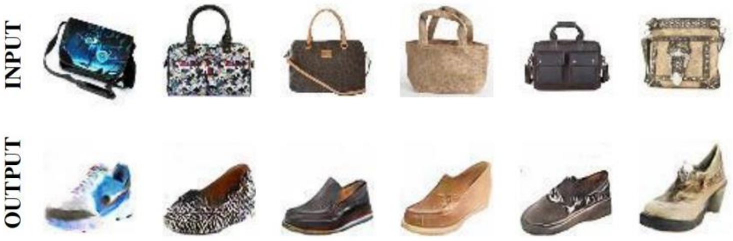
In the real world, to assess whether DiscoGAN successfully learns the underlying relationship between domains, training and testing uses different image-to-image translation tasks that require the use of discovered interdomain relationships between the source and destination domains. An example is expressed in Figure 6, which represents the result of the DiscoGAN training process. It shows a high-level overview of the unsupervised training procedure with two independent image sets (bags and shoes) without any additional annotation.



(a) Learning cross-domain relations **without any extra label**

FIGURE 6: DISCOGAN LEARNING SCHEMA

After the training process, DiscoGAN can combine and map the relationships between two domains such as the world of bags and the world of shoes and create the right shoes to match a certain bag as the following figures show.



(b) Handbag images (input) & **Generated** shoe images (output)

FIGURE 7: DISCOGAN RELATIONSHIPS

The process also happens correctly by inverting the source and destination domains, obtaining the same quality of results shown above:

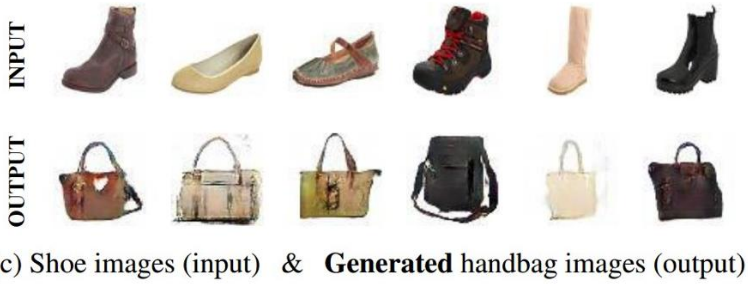

INPUT
OUTPUT
(c) Shoe images (input) & **Generated** handbag images (output)

FIGURE 8: DISCOGAN CROSS-RELATIONSHIPS

3.3 Radar discovery

Radar-based methods are commonly used to non-destructively detect concealed objects such as buried landmines, roots, breast tumours and concealed weapons on people. It is a recent trend to use a multimodal screening procedure for deceptive behaviours using the computer vision and machine learning approach using high-quality radar signal data often difficult to obtain from reliable sources. In the absence of such data, the computer AI-based approach fails to exceed the performance of human inspection of radar data. Most current radar-based algorithms for concealed object classification use simulated data that are free of clutter and generally only contain simple noise sources.

GANs have seen widespread application in the field of image processing and unsupervised image generation and can produce one-dimensional data, as to radar signal data, in audio applications. Audio data, like the ultra-wideband radar signals used for object detection, are complicated, nonstationary signals which are prone to external sources of noise and are difficult to process, often requiring qualitative human analysis to analyse the results. GANs can be used in the generation of radar signals indistinguishable from real ones by human observers to simplify the process of detecting hidden objects. Most applications of GANs and neural networks using radar data focus on images generated from radar signals using synthetic aperture radar (SAR) and time-of-flight algorithms. Applications would include data augmentation on rare events such as buried explosive detection in the ground and concealed object detection on people.

An existing implementation[43] based on WaveGAN and DCGAN uses three GANs trained with samples focused on concealed object detection on humans and generated using a FiniteDifference Time-Domain (FDTD) method. The FDTD method is extremely popular in the field of computational electromagnetics and the details on implementing the method are largely covered[44]. Each GAN can discover a different class of object: no concealed object, a large, concealed object, or a small concealed object. The model is described as follows:

[43] Truong, T., & Yanushkevich, S. (2020). Generative Adversarial Network for Radar Signal Generation.

[44] Taflove, A., & Hagness, S. (2000). Computational electrodynamics: the finite-difference time-domain method.

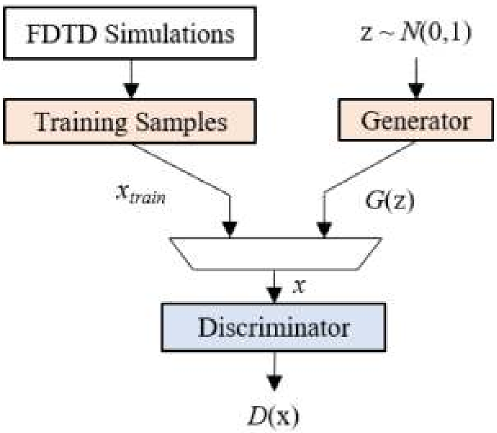


FIGURE 9: RADAR DISCOVERY ARCHITECTURE

The measuring, learning and testing systems are designed to emulate a simplified real-life scenario where a suspect is attempting to conceal a highly reflective object underneath layers of clothing. The simulations are simulated over 20 cm on the vertical axis and 50 cm on the horizontal axis with absorbing boundary conditions. With the scenario described, it was possible to measure the parameters to be used as a test set with a radar transceiver.

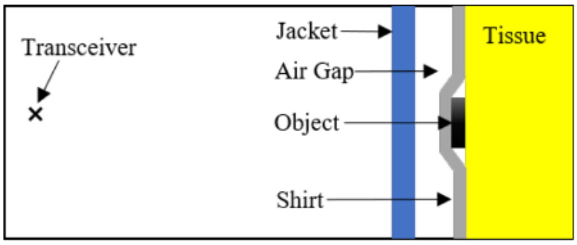


FIGURE 10: RADAR DISCOVERY SCENARIO

Figures 11 and 12 are annotated with the Early Time Response (ETR) and Late Time Response (LTR). The ETR exists in approximately the first 1.5 ns of the reflected signal and often captures the first reflections of the source signal from the system under test. The LTR consists of the measured response after 1.5 ns and contains smaller amounts of energy which have had multiple transmissions and reflections between layers in the system under test before returning to the transceiver.

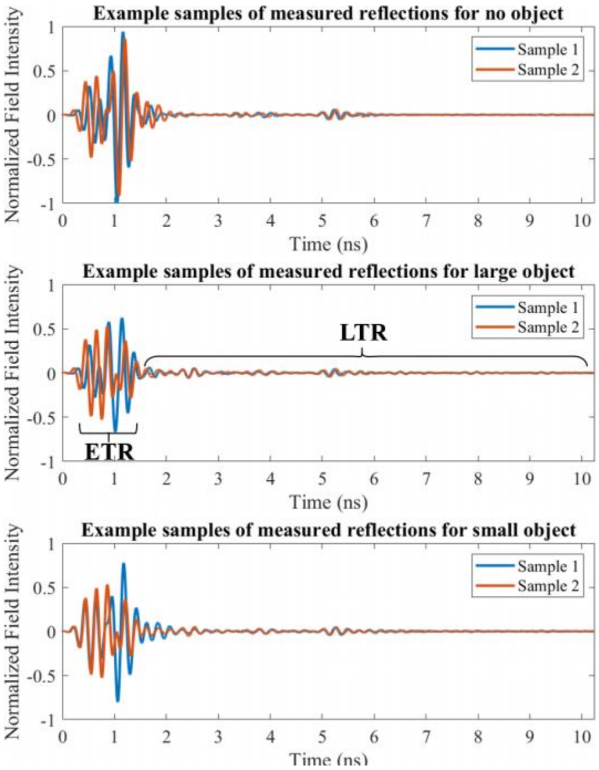


FIGURE 11: MEASURED REFLECTIONS FOR DIFFERENT OBJECT SIZES

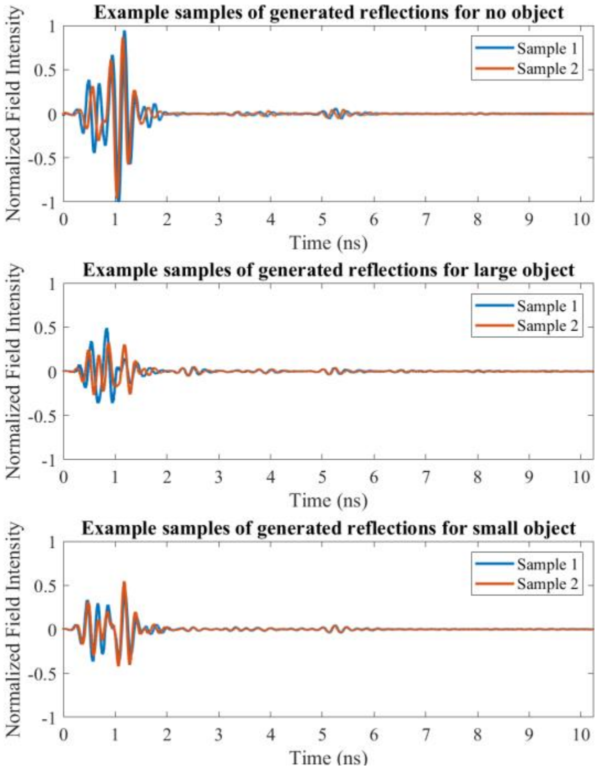


FIGURE 12: MEASURED GENERATED REFLECTIONS FOR DIFFERENT OBJECT SIZES

Figures 13 and 14 show the relative spectrograms for a 700 time-sample length window and 680 overlapped time samples. Spectrograms are a useful tool commonly used in audio analysis, and their application here reveals visual differences between each of the classes of data simulated. The 'no object' class contains little signal energy in the 3.1 - 5.3GHz range past 6ns. The large object class contains significant energy in those frequencies past 6ns. The small object class contains energy around 4.0 to 6.0GHz past 6ns.

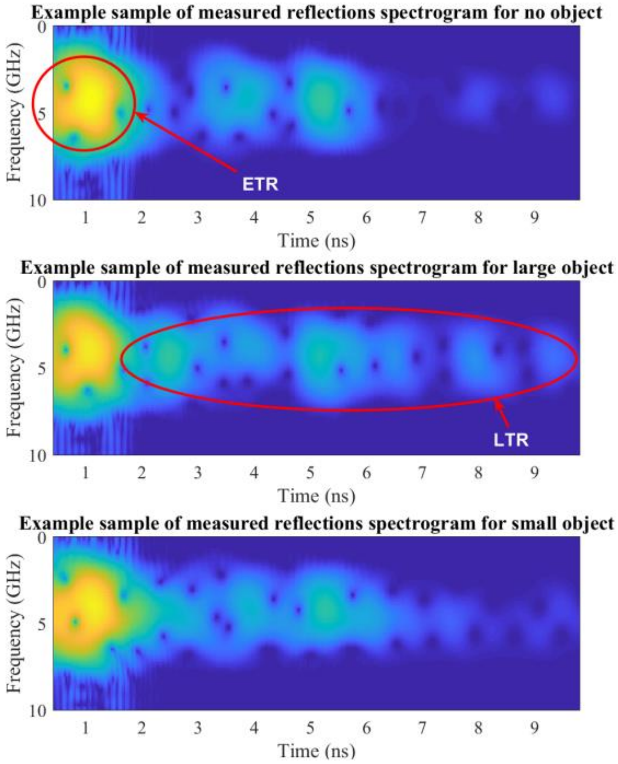


FIGURE 13: MEASURED REFLECTIONS SPECTROGRAM FOR DIFFERENT OBJECT SIZES

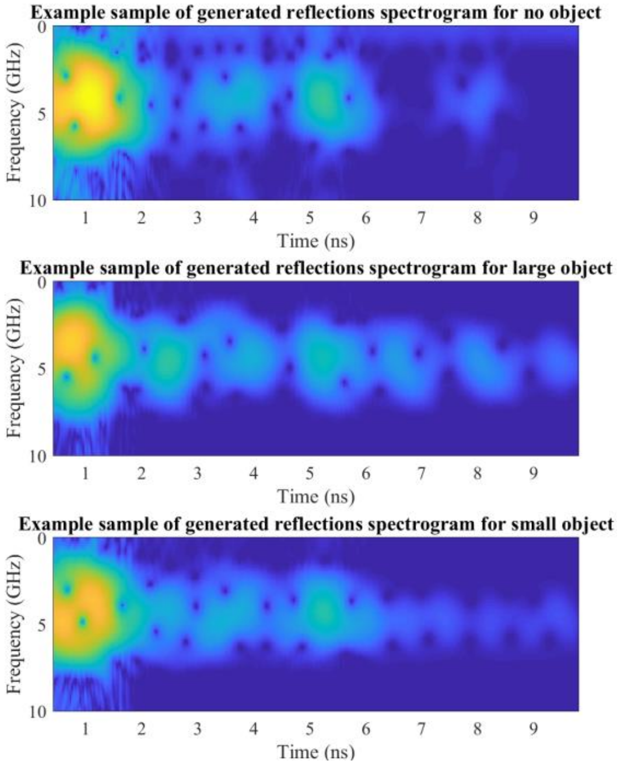


FIGURE 14: MEASURED GENERATED REFLECTIONS SPECTROGRAM FOR DIFFERENT OBJECT SIZES

To compare the ensemble variances of the training samples and generated samples, the mean squared error (MSE) is used in each object class category. The calculated MSE is 3.3e−5 for the no object generator, 1.2e−5 for the large object generator, and 9.0e−7 for the small object generator. Figure 15 shows the variance between the training samples and the samples generated for each object class.

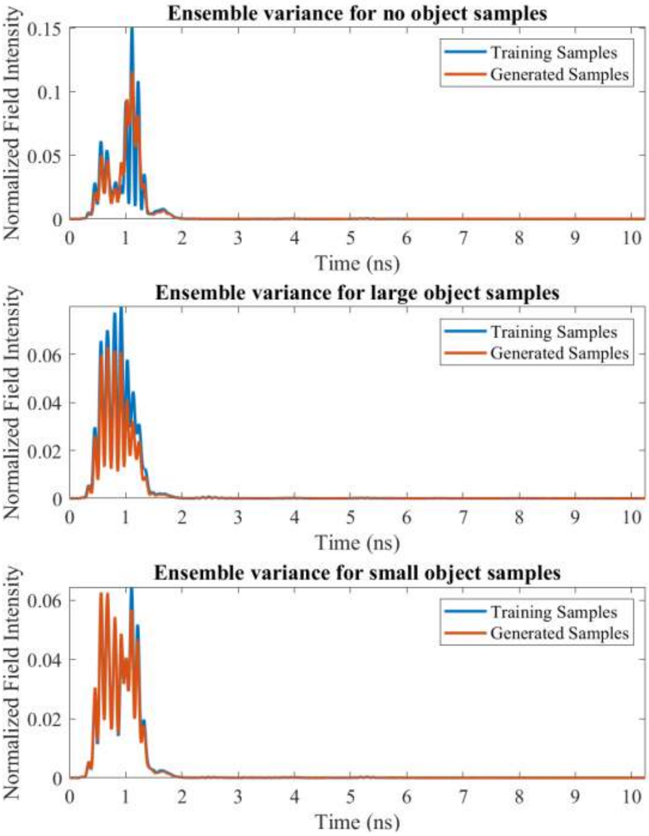


FIGURE 15: VARIANCES OF THE TRAINING SAMPLES AND GENERATED SAMPLES

The results of the GAN show promising results for the generation radar signal data, generating samples that are indistinguishable (by humans) from the training samples. This proof of concept lays the foundation for future research into the field of radar signal generation using GANs. With additional research, they may be capable of performing data augmentation on tedious, time-consuming and expensive-to-collect radar signals.

3.4 Wireless communications

Nowadays, remote communications are more and more frequent. Very often, physical or architectural barriers prevent cable communications, necessitating the use of wireless technologies. Due to the open and shared nature of wireless communication, various attacks can be performed against wireless systems. For this reason, it is important to have a mechanism to authenticate wireless signals at the physical layer before they proceed through the receiver chain. An example of an attack is a spoofing attack in which an adversary aims to mimic a legitimate user. One common approach for wireless signal spoofing is the replay attack where an attacker records a legitimate user’s transmission and repeats it. Fortunately, deep learning helps protect against these unpleasant events, finding many applications in wireless communications including spectrum sensing and modulation recognition. The adversary can also apply deep learning to launch wireless attacks to learn the underlying transmission behaviour by training a deep neural network and effectively jamming data transmissions.

A powerful technology such as a GAN can spoof wireless signals as if they originate from intended legitimate or primary users generating wireless signals that cannot be reliably discriminated from intended signals[45]. To train the GAN, four wireless devices are needed:

- an intended transmitter (T) to generate wireless communication;
- an intended receiver (R) used to validate or not the transmission;
- an attacker transmitter (AT) is used to train the generator;
- an attacker receiver (AR) is used to train the discriminator.

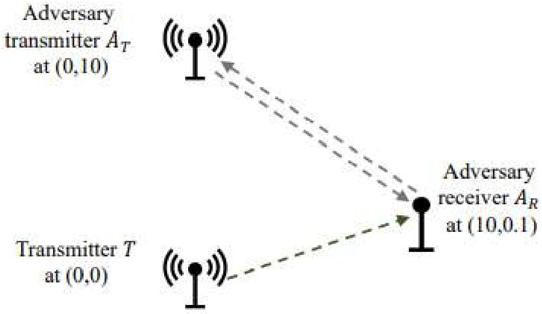


FIGURE 16: WIRELESS SPOOFING TRAINING ARCHITECTURE

AR must be placed close to R so that the receivers can receive the same signals from the transmitters. When AT makes its transmissions, it sends a flag signal such that AR knows these signals are from AT. The discriminator receives the signals from AR and compares them to classify signals and transmits the classification results to AT as feedback. Then the generator improves its transmitted signals such that these signals are more similar to T's signals. This is an iterative process that ends when the GAN converges. The topology of the spoofing attack phase is:

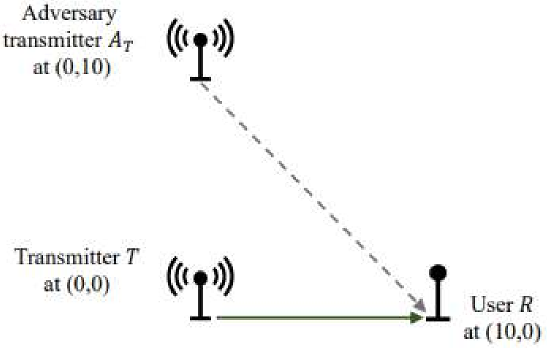


FIGURE 17: WIRELESS SPOOFING TEST ARCHITECTURE

The advantage of this attack is that the adversary does not need any prior knowledge of T, which will be learned by AR. Neither do they need to learn the channel effect explicitly. Instead, channel effects such as phase shift and propagation gain are learned implicitly through the collaboration of AT with AR.

From the spoofing tests, it is clear that the GAN-based spoofing attack provides a major improvement in attack success probability over the random signal and replay attacks. Table 1 reports the success probability of spoofing attacks by different methods.

TABLE 1: SPOOFING ATTACK SUCCESS PROBABILITY

Method of spoofing attack	Success probability
Random signal	7.89%

[45] Shi, Y., Davaslioglu, K., & Sagduyu, Y. (2019). Generative Adversarial Network for Wireless Signal Spoofing.

Replay	36.2%
GAN-based	76.2%

If the position of the attacker transmitter changes after its training, the effectiveness of the spoofing attack tends to decrease while remaining the best between the attack based on random signals or the replay attack.

Table 2: Adversary transmitter location success probability

AT location	Success probability
(0, 10)	76.2%
(0, 11)	65.2%
(0, 15)	61.0%
(0, 20)	56.2%

As the GAN opens us new opportunities to effectively spoof wireless signals, new defence mechanisms are called for as future work.

3.5 Credential guessing

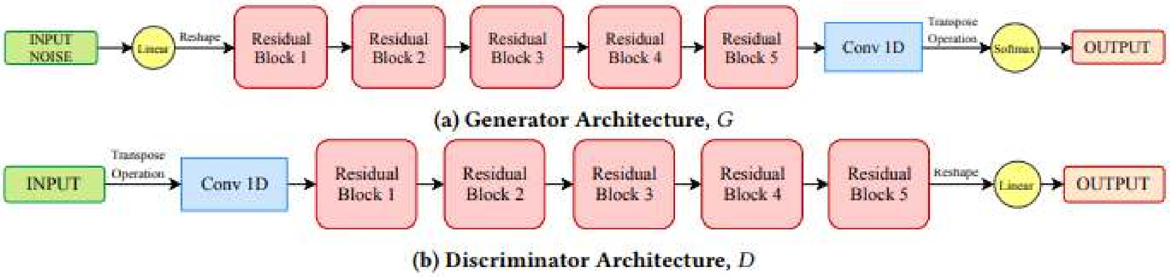
Password authentication is one of the most commonly used methods by users who tend to choose easy-to-guess passwords as common strings. These types of strings are subject to attacks called password guessing in which an attacker tries to log in using a database of common strings, dictionaries words and previous password leaks. The effectiveness of the attack relies on the ability to quickly test a large number of highly likely passwords against each password hash. An advanced technique is based on intuition on how users choose passwords by defining a heuristic for password transformations, which include combinations of multiple words and upper-case and lower-case letters in conjunction with Markov models.

Developing and testing new rules and heuristics is a time-consuming task that requires specialised expertise, and therefore has limited scalability. Thus, it is important to find a way to replace rule-based password guessing and password guessing based on simple data-driven techniques such as Markov models. As the password is a text-encoded string, a GAN-based approach can be used and particular IWGAN is the most stable approach for text generation. A password-based variant of IWGAN is represented by PassGAN[46], a novel line based on deep learning where a neural network is trained to determine autonomously password characteristics and structures and to leverage this knowledge to generate new samples that follow the same distribution. Deep neural networks are expressive enough to capture a range of properties and structures that describe the majority of user-chosen passwords and can be trained without any prior knowledge or assumptions. This involves a wide range of password-guessing knowledge that includes and surpasses what is captured in human-generated rules and Markovian password generation processes.

GANs are designed to perform density estimation in high-dimensional spaces since they perform implicit generative modelling by training a deep neural network architecture powered by simple random distribution and generating samples that follow the distribution of available data. To learn the generative model, a deep generative network (G) tries to mimic the underlying distribution of the samples. Then, a discriminating deep neural network (D) tries to distinguish between the original training samples and the

[46] Hitaj, B., Ateniese, G., Gasti, P., & Perez-Cruz, F. (2019). PassGAN: A Deep Learning Approach for Password Guessing.

samples generated by G and leak the relevant information on training data to it. This information helps G to adequately reproduce the original data distribution. The architecture of the PassGAN model is shown in Figure 18 which shows a high-level diagram of the layers that make up G and D. Each is a deep neural network that contributes to the growth in the performance of the other.



(a) Generator Architecture, G

(b) Discriminator Architecture, D

FIGURE 18: PASSGAN ARCHITECTURE

Using this architecture, PassGAN can generate an unlimited number of passwords because, unlike the password generation rules, the number of unique passwords that can be generated is not defined by the number of rules and the size of the password dataset used. Since PassGAN is not limited to a small subset of the password space, it can generate more passwords than any other tool, even though all tools have been trained on the same password dataset.

The results that will be shown are based on the use of two different datasets:

- RockYou: a common password list containing 3.094.199 entries;
- LinkedIn: a list of leaked passwords containing 43.354.871 entries.

The following image shows the number of unique passwords generated by PassGAN on various checkpoints, matching the RockYou testing set related considering a sampling size of 10^8 entries for each checkpoint.

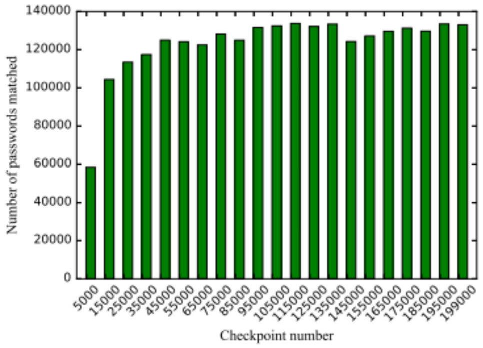


FIGURE 19: UNIQUE PASSWORDS GENERATED AT CHECKPOINTS

To evaluate the quality of PassGAN, the output is compared with the outputs of length 10 characters or less from HashCat Best64, HashCat gen2, JTR SpiderLab, FLA, PCFG and a Markov model. Table 3, based on the RockYou testset, show that, for each of the tools, PassGAN was able to generate at least the same number of matches. To achieve this, PassGAN needed to generate many passwords that were one order of magnitude higher than each of the other tools.

TABLE 3: PASSGAN COMPARISON CHART BASED ON THE ROCKYOU TESTSET

Approach	Unique Passwords	Matches	Number of passwords for PassGAN to outperform	PassGAN Matches
JTR Spyderlab	10^9	461,395 (23.32%)	$1.4 \cdot 10^9$	461,398 (23.32%)
Markov Model 3-gram	$4.0 \cdot 10^8$	532,961 (26.93%)	$2.47 \cdot 10^9$	532,962 (26.93%)
HashCat gen2	10^9	597,899 (30.22%)	$4.8 \cdot 10^9$	625,245 (31.60%)
HashCat Best64	$3.6 \cdot 10^8$	630,068 (31.84%)	$5.06 \cdot 10^9$	630,335 (31.86%)
PCFG	10^9	486,416 (24.59%)	$2.1 \cdot 10^9$	511,453 (25.85%)
FLA	$7.4 \cdot 10^9$	8 652,585 (32.99%)	$6 \cdot 10^9$	653,978 (33.06%)

The results from PassGAN for the previous test set are repurposed with a test set based-on LinkedIn leaked passwords as shown in Table 4.

TABLE 4: PASSGAN COMPARISON CHART BASED ON THE LINKEDIN TEST SET

Approach	Unique Passwords	Matches	Number of passwords for PassGAN to outperform	PassGAN Matches
JTR Spyderlab	10^9	461,395 (23.32%)	$1.4 \cdot 10^9$	461,398 (23.32%)
Markov Model 3-gram	$4.0 \cdot 10^8$	532,961 (26.93%)	$2.47 \cdot 10^9$	532,962 (26.93%)
HashCat gen2	10^9	597,899 (30.22%)	$4.8 \cdot 10^9$	625,245 (31.60%)
HashCat Best64	$3.6 \cdot 10^8$	630,068 (31.84%)	$5.06 \cdot 10^9$	630,335 (31.86%)
PCFG	10^9	486,416 (24.59%)	$2.1 \cdot 10^9$	511,453 (25.85%)
FLA	$7.4 \cdot 10^9$	8 652,585 (32.99%)	$6 \cdot 10^9$	653,978 (33.06%)

PassGAN proved able to match 35% of unique passwords from the RockYou dataset, and 34% from the LinkedIn dataset. However, the results show that the best password guessing strategy is to use multiple tools. By combining the output of PassGAN with the output of the HashCat Best64 rules, it can guess between 51% and 73% more unique passwords than HashCat alone.

The ability of PassGAN has been implemented[47,48] and reported in various articles from Science Magazine, Threatpost, Dark Reading, Sensors Online and the model was selected by Dark Reading as one of the coolest hacks of 2017 and has now been improved[49] by several researchers.

[47] https://github.com/brannondorsey/PassGAN%20

[48] https://github.com/d4ichi/PassGAN%20

[49] https://github.com/Riathoir/PASSGAN-IWGAN-Tensorflow-2

3.6 Digital defence evasion

In recent years, the use of digital protection systems has evolved considerably to protect the quantity of increasingly sensitive data that is deposited on the network every day. In the field of security from digital threats such as malware, many algorithms based on machine learning have been proposed to detect threats based on the extraction and analysis of their functionality. Considerable efforts have been made to improve detection performance at the expense of the toughness of the solutions adopted, but while deep learning has favoured manufacturers of security solutions, it has also favoured digital threat producers bent on studying and demolishing malware detection algorithms. Many machine learning algorithms are vulnerable to intentional attacks and hardly usable to use in real-world applications.

The first attack models developed assumed full access to the parameters of the malware detection model. From a practical point of view, this approach is impractical as the malware detection algorithms integrated into antimalware software are well protected. The birth and development of GANs have allowed the experimentation of new attacks to black-box detection algorithms. This is the case of MalGAN[50] which is based on the administration and evaluation of punctual test cases against a specific black-box to understand and extract the functionalities considered by the detection algorithm.

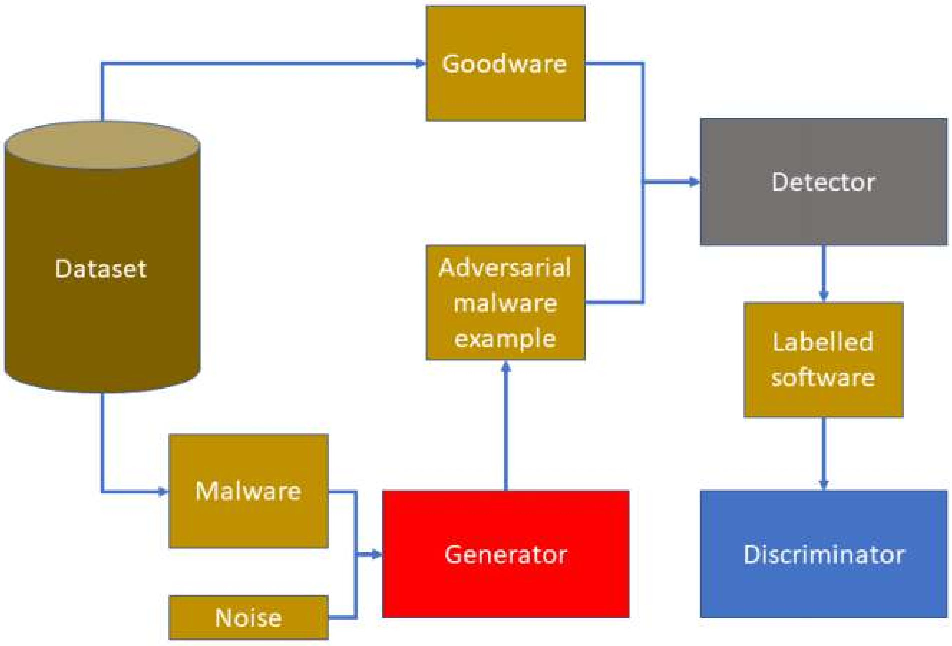


FIGURE 20: MALGAN ARCHITECTURE

The MalGAN model just shown, contains four key elements:

- a detector, which is a black-box machine-learning-based malware detection algorithm;
- a generator and a discriminator, which are both feed-forward neural networks;
- a dataset of malware and goodware features used to train the detector and the generator.

The generator and discriminator train each other to successfully attack a machine learning-based black-box malware detector. The discriminator is trained to fit the detector and the generator to fool the detector, producing adversarial examples. An excellent way to train the neural networks of this GAN is to use binary functionalities such as the presence or absence of some Application Program Interface (API), since they are widely used by malware detection systems and can guarantee a high accuracy

[50] Weiwei, H., & Ying, T. (2017). Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN.

detection. To make a machine learning algorithm effective, the samples in the training set and the test set must follow the same or similar probability distributions. Because the generator can change the probability distribution of adversarial examples from that of the black-box detector's training set, it can lead the detector to misclassify malware as benign, producing more complex and flexible examples to fool the target model.

Table 5 shows the results obtained by training the machine learning-based components using a large set of features extracted from malware and goodware. The results show the percentage of detection by the detectors of the original samples and the adversarial samples in both the training and in the test sets.

Table 5: MalGAN evasion performances

Algorithm	Training Set		Test Set	
	Original	Adversarial	Original	Adversarial
Random forest	95.10%	0.71%	94.95%	0.80%
Logistic regression	91.58%	0.00%	91.81%	0.01%
Decision trees	91.92%	2.18%	91.97%	2.11%
Support vector machines	92.50%	0.00%	92.78%	0.00%
Multi-layer perceptron	94.32%	0.00%	94.40%	0.00%

The results obtained are astounding because the superiority of MalGAN over traditional gradient-based adversarial example generation algorithms is that MalGAN can decrease the detection rate to nearly zero and make the retraining-based defensive method against adversarial examples hard to work. Malware authors can frequently retrain MalGAN, preventing the black-box detector from keeping up with it and making it unable to learn stable patterns from it. Once the black-box detector is updated, malware authors can immediately crack it. This process makes machine learning-based malware detection algorithms unable to work.

MalGAN is can also fool further defensive methods of detection algorithms and some of its uses have been published[,51,52,53] Architectural[54] and process flow-enhanced[55,56] versions of the GANs have also emerged. Neither is there any shortage of implementations that have been made public[57]. It can well be imagined that this tool, used inappropriately by malicious people, can represent a significant danger to the entire international community as it appears to be strongly interconnected within it. Fortunately, solutions have emerged to counter this type of phenomenon by developing more efficient detection[58] models based on deep learning approaches (Lu, et al. 2019). At the same time, from a practical point of view, the producers of anti-malware solutions[59] are committed to the continuous improvement of their products to ensure the safety of their users from this type of threat.

[51] https://github.com/yanminglai/Malware-GAN

[52] https://github.com/ZaydH/MalwareGAN

[53] https://www.codeproject.com/Articles/1260883/How-to-fool-Machine-Learning-malware-detectors-usi

[54] Kawai, M., Ota, K., & Dong, M. (2019). Improved MalGAN: Avoiding Malware Detector by Leaning Cleanware Features.

[55] Buonocore, G. (2019). MalGAN: Evasione e rilevamento di malware neurali.

[56] Labaca Castro, R., Schmitt, C., & Dreo Rodosek, G. (2019). Poster: Training GANs to Generate Adversarial Examples Against Malware Classification.

[57] https://github.com/tubutubucorn/Improved_MalGAN

[58] Huang, A., & Huang, Y. (2018). Towards Robust Malware Detection.

[59] Bühler, T. (2019, August 22). Defending against GAN-made malware. Available at https://www.avira.com/en/blog/gan-made-malware

4. Understanding GAN Applications

Before we can develop a framework, we must first develop an understanding of how GAN is being used.

4.1 Traditional Uses of GAN

Currently, GANs focus mainly on image-based applications or applications that can be made to look like the same problem space such as handwriting analysis, image classification, object detection and image editing. GANs use supervised and unsupervised learning methods that are best matched to input data and which allow automatic feature extraction. To achieve this, the underlying data structure for any input data must have both value and direction: i.e., the underlying data structure must be easily represented as a two-dimensional array of $\mathrm{m} \times \mathrm{n}$ data. Images are a good match as they have both a signal (pixel value) and location in the image to enable features to be extracted. This does not mean that GANs are only for image-based applications. Other novel applications include audio processing (signal and time), drug discovery[60] (drug molecule success in prior trials and time) and even the development of molecules that may be used in the fight for cancer (molecules and cancer growth data). A full analysis of GAN applications can be found in the literature.[61]

Focusing on image-based GANs, *This Person Does Not Exist* is a website implementation of StyleGAN built by Phil Wang[62] and drawing on the work of Karras et al.[63] StyleGAN allows for the production of images of faces of people who, as the name implies, do not exist. When additional items are included in the images such as fingers and hats, the images start to show signs of artefacts that indicate to the layman that they are not real. Indeed, the images generated by StyleGAN are easily identified due to the pinning of key features in the image itself such as the right eye and the location of the teeth. These specific identification issues were overcome in the further work of Karras et al. coined StyleGAN2.[64] While these images are believable when these defects are not known, 'at present, classifier-based methods can quite reliably detect generated images, regardless of their exact origin'.[65]

[60] Medicine I (2019) Artificial intelligence for drug discovery, biomarker development and aging research. URL https://insilico.com/.

[61] Alqahtani, H., Kavakli-Thorne, M. & Kumar, G. Applications of Generative Adversarial Networks (GANs): An Updated Review. Arch Computat Methods Eng 28, 525–552 (2021). https://doi.org/10.1007/s11831-019-09388-y.

[62] Wang, P., 2019. This person does not exist. [Online] Available at: https://thispersondoesnotexist.com.

[63] Karras, Tero, Samuli Laine, and Timo Aila. 'A style-based generator architecture for generative adversarial networks'. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[64] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T., 2020. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp.8110-8119).

[65] Ibid. 8116

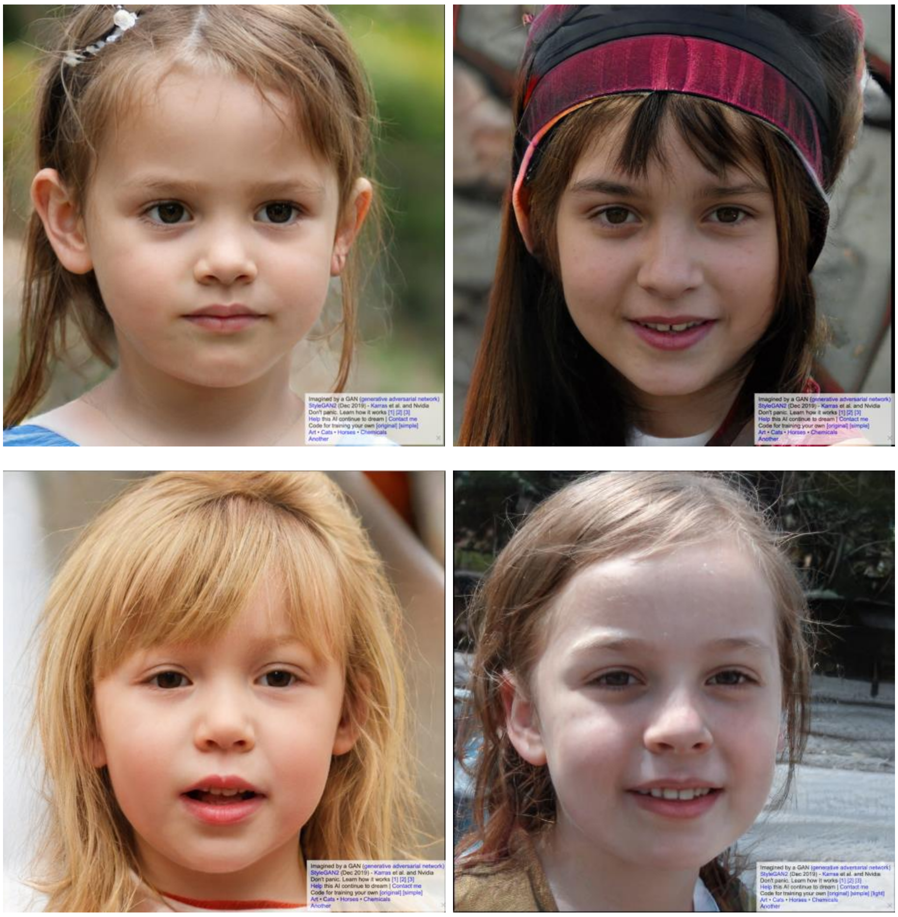

Imagined by a GAN (generative adversarial network)
StyleGAN2 (Dec 2019) - Karras et al. and Nvidia
Don't panic. Learn how it works [1] [2] [3]
Help this AI continue to dream | Contact me
Code for training your own [original] [simple]
Art • Cats • Horses • Chemicals
Another

FIGURE 21: EXAMPLE IMAGES OF PICTURES OF GIRLS GENERATED FROM THE WEBSITE THISPERSONDOESNOTEXIST.COM.[66]

Generating fake faces is not the only current application of GAN. Two other image-based applications of GAN that are currently showing promise are OpenAI's CLIP and DALL-E. The idea behind DALL-E is to use a GAN to generate visual images from natural strings of plain text,[67] while CLIP extrapolates the final product of an image.[68] Both have applications in the cyber intelligence space. For example, CLIP has theoretical uses where partial data has been obtained and needs to be completed whereas DALL-E can be used to assist in the reconstruction of faces from eyewitness accounts. Neither application currently exists as the technology is at an early stage of development. There is clear evidence to suggest that such GAN uses will dominate in the next few years to assist in reconstructing digital evidence obtained in the field in video, image and other electronic data.

4.2 Intelligence Cycle

From our discussion of intelligence, we understand that for GAN to be useful as a tool in the intelligence cycle it must be effective in one or more from the activities of collection, validation, exploitation or

[66] Wang, P., 2019. This person does not exist. [Online] Available at: https://thispersondoesnotexist.com.

[67] Ramesh, A., Pavlov, M., Gray, S., Goh, G., Wang, J., Chen, M., Chen, R., Misra, V., Mishkin, P., Krueger, G., Agarwal, S. and Sutskever, I., 2021. DALL·E: Creating Images from Text. [online] OpenAI. Available at: <https://openai.com/blog/dall-e/> [Accessed 3 February 2021].

[68] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. and Krueger, G., Learning Transferable Visual Models From Natural Language Supervision. Image, 2, p.T2.

reporting. This poses an interesting philosophical point since the purpose of a GAN is to develop synthetic data that can pass for real data. Actionable intelligence is typically collected, not invented. That being said, machine learning is being used to assist in the collection, validation, exploitation or reporting of intelligence information, but not usually through the use of GAN.

Tundis et al.[69] have produced a regression model to automate the assessment of counter intelligence (CTI) acquired from Twitter, enabling the results to be disseminated some 32 hours earlier than by traditional workflows. A list of features was obtained by interviewing 30 experts from a pool of cybersecurity professionals and academic researchers in the field of cyber threat intelligence. Among these features included the breakdown by four main characteristics to assess the CTI based on the level of detail, credibility, timeliness and actionable content. Using four regression models, a CTI relevance score was determined based on the output of the assessment from these experts. A metadata analysis was then conducted from tweets and Twitter profiles on a range of cybersecurity-related hashtags during the 31 days of May 2019). The analysis showed that the regression model was capable of getting CTI into the hands of analysts some 32 hours earlier. While this result is an impressive use of machine learning, it does not use a GAN.

Yang and Lam[70] have also looked for a solution to the information explosion issue facing Security Operation Centres (SOCs). In their work, six separate machine learning classifiers were used to pre-classify and sort incoming CTI reports before being analysed by traditional means. Machine learning methods were used to improve the efficiency and ability to rapidly assess and classify large volumes of CTI reports in SOCs. Their results indicate that, by processing large volumes of data, early warnings can be given to help mitigate or even avoid cyberattacks. We refer to their approach as effective processing of reports as opposed to the need to triage currently being used in the IC.

Neither Tundis et al. nor Yang and Lam specifically use a GAN in their work. Instead, they use a regression model and so we see no need to put this work through the assessment of our analysis. However, their work offers a valuable insight into how a GAN might be used to conduct similar work and provide further justification for the extension of this work into other machine learning algorithms, not just GAN. Both groups use machine learning tools to automate processes in the intelligence cycle. Their combined work focuses on the dissemination stages of the intelligence cycle, relying on the fact that intelligence has already been collected, validated and exploited into actionable items. The value of an individual actionable item of intelligence will increase with more samples of the same report being observed from different sources, hence both tools also assist in the validation, exploitation and dissemination of intelligence, but only when intelligence already exists.

While regression models are useful, the focus of this work is on the use of GAN whose benefits are in specific problem domains such as classification. In this context, it has been found that GANs are not being used but rather other machine learning algorithms are employed.

4.3 Red Teaming

While we stated above that an aim of this project should be to explore how GAN might be used to assist in gaining initial footholds into systems, to propagate or pivot and finally to achieve actions on an objective, this typically falls within the scope of counter-espionage and not intelligence-gathering operations. As a result, this has now been deemed beyond the scope of this report and these

[69] Tundis, A., Ruppert, S. and Mühlhäuser, M., 2020, June. On the Automated Assessment of Open-Source Cyber Threat Intelligence Sources. In *International Conference on Computational Science* (pp.453-467). Springer, Cham.

[70] Yang, W. and Lam, K.Y., 2019, December. Automated cyber threat intelligence reports classification for early warning of cyber attacks in next generation soc. In *International Conference on Information and Communications Security* (pp.145-164). Springer, Cham.

applications should be assessed against a separate framework than the one we will propose in the next section.

The framework we propose looks at cyber intelligence from the perspective of the intelligence cycle and not active red team capabilities (counterintelligence and espionage). While this makes up a significant portion of the domain, due to the difficulties in defining the domain we have excluded it from this report as beyond scope. Future work should specifically target these applications using the same methodology we have described, adjusted for a unified kill chain approach instead of the intelligence cycle. Given the capabilities of GAN to generate data that can be used to increase disinformation and misinformation, we recommend that this work be conducted.

5. Framework Design

To assess the uses of GAN, we must understand how advanced research into GAN as a product for intelligence exploitation has become. To understand this, we use a ranked weighted decision matrix (WDM) method to assess the applications, thus answering the questions laid down in Section 1.

The WDM is broken down into three segments, following the three questions. Each question has a key set of associated criteria. To answer question 1 (see Table 6), an assessment against an intelligence framework (IFA) will be conducted. To gain insight into how GAN is used with footholds, propagation or pivots in systems an assessment may be conducted against the unified kill chain (UKC). However, this has been left as beyond the scope of this report. Finally, to showcase how GAN is being used traditionally and how ready it is to be used in a cyber intelligence operation, the technology readiness of the GAN feature is documented. The weights for the weighted decision matrix are 50%, 20% and 10% respectively. This is summarised in Table 6.

TABLE 6: WEIGHTED DECISION MATRIX CRITERIA TO ASSESS THE CYBER INTELLIGENCE APPLICATIONS OF GAN.

	Intelligence Framework Assessment	**Technology Readiness Level**	**Final Score**
Weight	1.0	0-100%	100%
Criteria	How does the technology assist in the intelligence cycle? Record the final score from the IFA table.	What is the current technology readiness level of the research as reported in the literature? Record the final score from the TRL Table	Add the score based on Equation (1) below.

The score from the weighted decision matrix is obtained from the following logical formula:

$$\text{Final Score} = \text{TRL} \times \left(\text{C}'\frac{\text{c}}{\text{C}} + \text{V}'\frac{\text{v}}{\text{V}} + \text{E}'\frac{\text{e}}{\text{E}} + \text{D}'\frac{\text{d}}{\text{D}}\right) \; (1)$$

where c, v, e and d are the individual scores from each stage of intelligence framework assessment; C, V, E and D are the total potential scores obtained in each stage of the IFA; $\text{C}', \text{V}', \text{E}', \text{and D}'$ are the weightings given in Table 77; and TRL is the scaling factor applied from the technology readiness level (TRL). The TRL is used as a weight for the entire equation (Equation 1) to emphasise that technologies that are more mature pose more of a threat or opportunity. The IFA is interchangeable with an assessment against a UKC for red team applications in the cyber intelligence domain; however, this is beyond the scope of this report.

5.1 Intelligence Framework Assessment – Intelligence Cycle Capabilities

The first part of the WDM is an assessment against the intelligence collection orientation identified through the definition of cyber intelligence used in this report. This collection cycle is influenced by the CIA intelligence cycle[71] and advances in OSINT.[72] By merging the current literature, a four-step

[71] Central Intelligence Agency, 2007.
[72] Williams, H. J. & Blum, I., 2018.

collection cycle has been identified focusing on the collection, validation, exploitation and communication of intelligence information.

TABLE 7: INTELLIGENCE FRAMEWORK ASSESSMENT (IFA)

	Collection	Validation	Exploitation	Communication	IFA Score
Weighting	0.25	0.25	0.25	0.25	1.0

The definition of cyber intelligence given above is useful to understand the potential or current uses of GAN in cyber intelligence activities. For GAN to be used in cyber intelligence activities, it must be involved in at least one of the collection, validation, exploitation or communication stages.

To assess the framework, we ask a series of questions to determine the score for the GAN application. Examples of these questions are in the tables below and are expected to be adopted and expanded on by the users of this framework as the needs change. These questions are broken down into sub-frameworks based on the stages of the intelligence cycle.

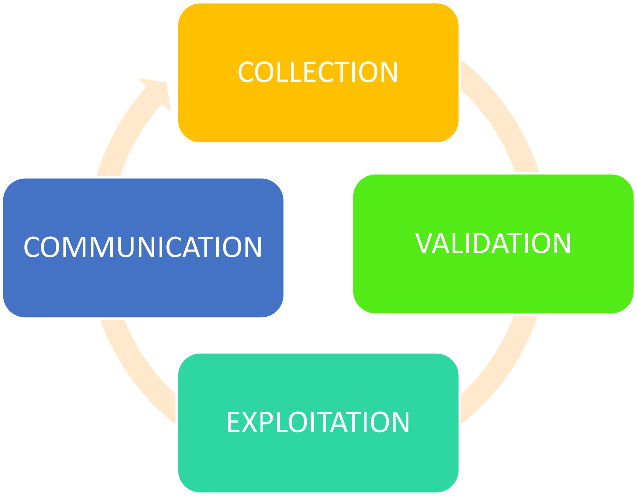


FIGURE 22: THE INTELLIGENCE CYCLE.

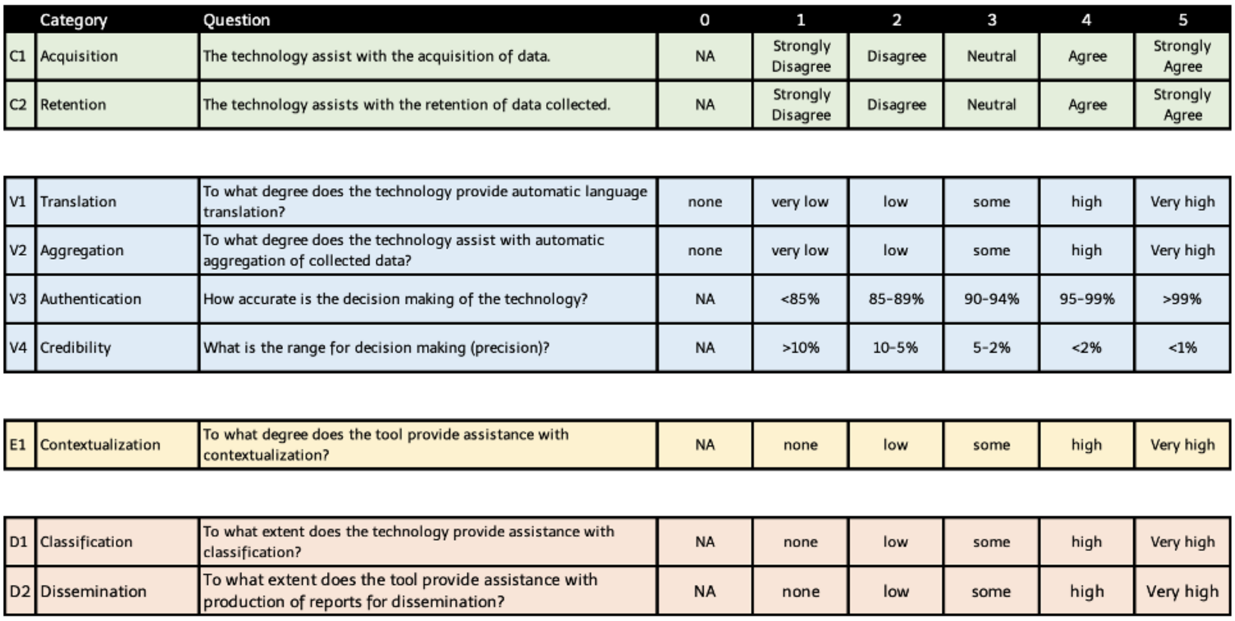
The first stage of the intelligence framework assessment is the collection sub-framework. This measures how well a piece of technology can deliver on the acquisition and retention of intelligence. This applies not just to cyber intelligence, but to all INT domains.

Following the collection framework is the validation framework. This assesses the ability of the technology under evaluation to meet the information needs of translation, aggregation, authentication and credibility.

It is not expected that any GAN tools will offer an ability for exploitation and that this will still require human interaction to create actionable intelligence. However, the third sub-framework is still the exploitation framework to create actionable intelligence.

Finally, the last sub-framework is based on the ability of any tool to classify and automatically disseminate actionable intelligence.

TABLE 8: INTELLIGENCE FRAMEWORK ASSESSMENT BROKEN DOWN BY CATEGORY IN THE INTELLIGENCE CYCLE.



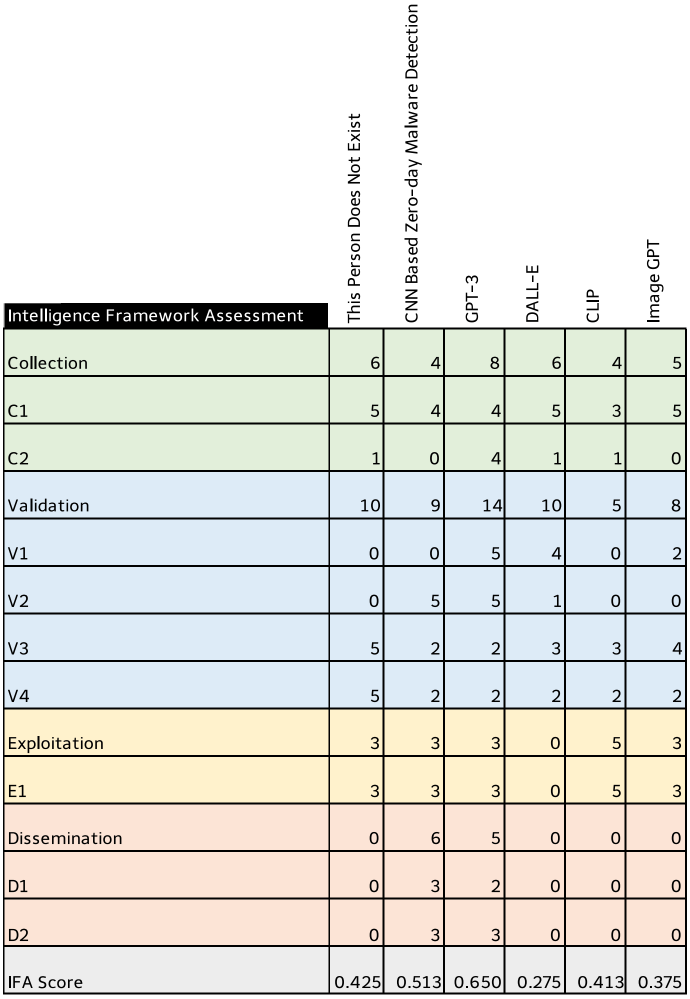
	Category	Question	0	1	2	3	4	5
C1	Acquisition	The technology assist with the acquisition of data.	NA	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
C2	Retention	The technology assists with the retention of data collected.	NA	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
V1	Translation	To what degree does the technology provide automatic language translation?	none	very low	low	some	high	Very high
V2	Aggregation	To what degree does the technology assist with automatic aggregation of collected data?	none	very low	low	some	high	Very high
V3	Authentication	How accurate is the decision making of the technology?	NA	<85%	85-89%	90-94%	95-99%	>99%
V4	Credibility	What is the range for decision making (precision)?	NA	>10%	10-5%	5-2%	<2%	<1%
E1	Contextualization	To what degree does the tool provide assistance with contextualization?	NA	none	low	some	high	Very high
D1	Classification	To what extent does the technology provide assistance with classification?	NA	none	low	some	high	Very high
D2	Dissemination	To what extent does the tool provide assistance with production of reports for dissemination?	NA	none	low	some	high	Very high

Individual tools are unlikely to score highly in all areas. Tools are likely to be designed to act on one or more of these areas in the intelligence collection framework. Each stage must be assessed individually on the assumption that any preconditions have been satisfied. For example, for the dissemination sub-framework to be assessed it must be assumed that actionable intelligence has already been created and is ready for dissemination regardless of whether the tool can generate such actionable intelligence. This will allow any tool to be assessed fairly based on the object of the tool.

It is unlikely that any individual tool will score highly in combined intelligence framework assessment, but will score highly in one of these sub-frameworks. Even then, it is not determined that this tool is ready for adoption or has been adopted by the IC. For this reason, we then weight any result from the IFA by the tool's technology readiness.

For the examples above, we now list the relevant scores under this framework in Table 9

TABLE 9: IFA ASSESSMENT FOR THE EXAMPLES GIVEN.



Intelligence Framework Assessment	This Person Does Not Exist	CNN Based Zero-day Malware Detection	GPT-3	DALL-E	CLIP	Image GPT
Collection	6	4	8	6	4	5
C1	5	4	4	5	3	5
C2	1	0	4	1	1	0
Validation	10	9	14	10	5	8
V1	0	0	5	4	0	2
V2	0	5	5	1	0	0
V3	5	2	2	3	3	4
V4	5	2	2	2	2	2
Exploitation	3	3	3	0	5	3
E1	3	3	3	0	5	3
Dissemination	0	6	5	0	0	0
D1	0	3	2	0	0	0
D2	0	3	3	0	0	0
IFA Score	0.425	0.513	0.650	0.275	0.413	0.375

5.2 Technology Readiness Level Assessment

The Technology Readiness Level (TRL) is an assessment framework developed by the National Aeronautics and Space Administration (NASA) over the last 50 years.[73] TRLs were initially used as a method of determining the 'acceptable readiness (of technology) for flight applications'.[74] It has since found applications in other areas of technology assessment and it is commonplace to see this framework adopted in similar industries to assess the technology needs of a specific organisation.[75] For example, the US Department of Energy (DoE),[76] Department of Homeland Security (DHS)[77] and Department of

[73] Straub, J., 2015. 'In search of technology readiness level (TRL) 10', Aerospace Science and Technology, vol. 46, pp.312-320.

[74] Sadin, S.R., Povinelli, F.P. and Rosen, R., 1989. The NASA technology push towards future space mission systems. In Space and Humanity (pp.73-77). Pergamon.

[75] Straub 2015.

[76] DoE, G., 2009. 413.3-4. US Department of Energy Technology Readiness Assessment Guide.

[77] McGarvey, D., Olson, J., Savitz, S., Diaz, G. and Thompson, G., 2009. *Department of Homeland Security Science and Technology Readiness Level Calculator* (ver. 1.1). Arlington, USA: Homeland Security Studies and Analysis Institute.

Defense[78] all use their own version of the TRL system, as do the Australian Department of Science and Technology Group and the Australian Defence Force (ADF), among many others.[79] We see that TRLs have been readily adapted to suit specific organisational needs beyond the initial applications in the aerospace industry.

First introduced in 1989, the TRL system was based on seven levels, the pursuit of which was found to be demonstratable to the success or failure of a particular technology in the NASA Advance Research and Technology programme. [80] The seven were originally defined as follows:

LEVEL 1 – BASIC PRINCIPLES OBSERVED AND REPORTED

LEVEL 2 – POTENTIAL APPLICATION VALIDATED

LEVEL 3 – PROOF-OF-CONCEPT DEMONSTRATED, ANALYTICALLY AND/OR EXPERIMENTALLY

LEVEL 4 – COMPONENT AND/OR BREADBOARD LABORATORY VALIDATED

LEVEL 5 – COMPONENT AND/OR BREADBOARD VALIDATED IN SIMULATED OR REAL-SPACE ENVIRONMENT

LEVEL 6 – SYSTEM ADEQUACY VALIDATED IN SIMULATED ENVIRONMENT

LEVEL 7 – SYSTEM ADEQUACY VALIDATED IN SPACE[81]

The purpose of these levels was to 'provide [NASA], and the communities with which it interacts, with a more precise means of describing the depth to which a research and technology program is to be pursued'.[82] While these initial levels lacked the precise descriptions needed to avoid ambiguity, common applications with a minimum TRL to be pursued were provided to illustrate how the levels were to be used. This included black box electronic circuitry systems to level 5 and propulsion systems to level 6 with level 7 as the conservative goal. 'This more demanding need (pursuit of level 7) is, of course, what makes these relatively complex technologies so time-consuming and expensive to develop and to prove flight ready'.[83] TRL is intended to determine the level of research that should be adopted on a specific technology before it could be safely adopted. This level is scalable and should change based on the technology under assessment.

In 1995, John Mankins[84] documented the TRL as used at NASA and as incorporated in the NASA Management Instruction (NMI 7100). This extension of the original seven-level TRL framework now included levels one through nine, broken down into five categories of technology development. These technology stages are described as:

> '(a) 'basic' research in new technologies and concepts (targeting identified goals, but not necessarily specific systems), (b) focused technology development addressing specific technologies for one or more potential identified applications, (c) technology development and demonstration for each specific application before the beginning of full system development of that application, (d) system development (through first unit fabrication), and (e) system 'launch' and operations'.[85]

[78] Sauser, B., Ramirez-Marquez, J.E., Magnaye, R. and Tan, W., 2009. A systems approach to expanding the technology readiness level within defense acquisition. STEVENS INST OF TECH HOBOKEN NJ SCHOOL OF SYSTEMS AND ENTERPRISES.

[79] DST, (n.d.). Technology Readiness Levels Definitions and Descriptions. Australian Government. https://www.dst.defence.gov.au/sites/default/files/basic_pages/documents/TRL%20Explanations_1.pdf.

[80] Sadin et al. (1989).

[81] Ibid. 74

[82] Ibid. 74

[83] Ibid. 75

[84] Mankins, J.C., 1995. Technology readiness levels. White Paper, April 6 (1995), p.1995.

[85] Ibid.

Along with the definition of stages, the TRLs were further broken down to be more specific. TRLs 1, 2 , 3, and 4 remained unchanged. The remainder of the TRL system was redefined as follows:

> LEVEL 5 – 'COMPONENT AND/OR BREADBOARD VALIDATED IN RELEVANT ENVIRONMENT'[86]
>
> LEVEL 6 – 'SYSTEM/SUBSYSTEM MODEL OR PROTOTYPE DEMONSTRATION IN A RELEVANT ENVIRONMENT (GROUND OR SPACE)'[87]
>
> LEVEL 7 – 'SYSTEM PROTOTYPE DEMONSTRATION IN A SPACE ENVIRONMENT'[88]
>
> LEVEL 8 – 'ACTUAL SYSTEM COMPLETED AND 'FLIGHT QUALIFIED' THROUGH TEST AND DEMONSTRATION (GROUND OR SPACE)'[89]
>
> LEVEL 9 – 'ACTUAL SYSTEM 'FLIGHT PROVEN' THROUGH SUCCESSFUL MISSION OPERATIONS'[90]

No classification of levels into stages was given at this point. It was not until Mankins published his work in 2009 that the official classifications used in the NASA literature reached peer review.[91] In this work, it was noted that the scale was developed in 1995 based on work originating from the mid-1970s.[92]

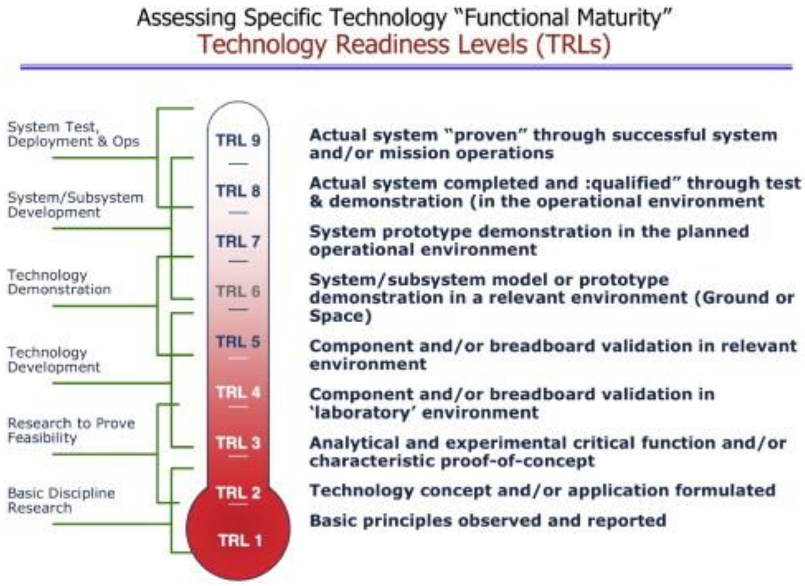

Assessing Specific Technology "Functional Maturity"
Technology Readiness Levels (TRLs)
System Test, Deployment & Ops
System/Subsystem Development
Technology Demonstration
Technology Development
Research to Prove Feasibility
Basic Discipline Research
TRL 9
TRL 8
TRL 7
TRL 6
TRL 5
TRL 4
TRL 3
TRL 2
TRL 1
Actual system "proven" through successful system and/or mission operations
Actual system completed and :qualified" through test & demonstration (in the operational environment
System prototype demonstration in the planned operational environment
System/subsystem model or prototype demonstration in a relevant environment (Ground or Space)
Component and/or breadboard validation in relevant environment
Component and/or breadboard validation in 'laboratory' environment
Analytical and experimental critical function and/or characteristic proof-of-concept
Technology concept and/or application formulated
Basic principles observed and reported

FIGURE 23: OVERVIEW OF THE TECHNOLOGY READINESS LEVEL SCALE AS SEEN IN MANKINS 2009.[93]

In 2014, Straub noted that a final TRL was needed which documented the final stage of technology readiness, that being the use of technology in extended operations.[94] He justified this final level based on the need to use space-based technologies commercially beyond the initial requirements that NASA defines. Such a level can also be seen as having been attained when product support is ongoing but not routinely needed to patch and rectify issues.[95] Such a definition is suitable not only for hardware systems, but also for software.

[86] Ibid.
[87] Ibid.
[88] Ibid.
[89] Ibid.
[90] Ibid.
[91] Mankins, J.C., 2009. Technology readiness assessments: A retrospective. Acta Astronautica, 65(9-10), pp.1216-1223.
[92] Ibid.
[93] Ibid.
[94] Straub, J., 2015. In search of technology readiness level (TRL) 10. Aerospace Science and Technology, 46, pp.312-320.
[95] Ibid.

We propose the adoption of the TRL-based explanations given by both Sadin et al. and Mankins. This version shown in Figure 24 is based on the work showing that the TRL framework is readily adopted based on organisational need. The main change from the Sadin framework is the inclusion of the Mankins and Straub extensions listed above. We have retained the linguistic stylings from the original Sadin framework and used them throughout the Makins and Straub extensions through levels 8, 9 and 10. Finally, we have adopted the decision questions to assess the technology readiness level as determined by Straub, modified for the generality of environment.

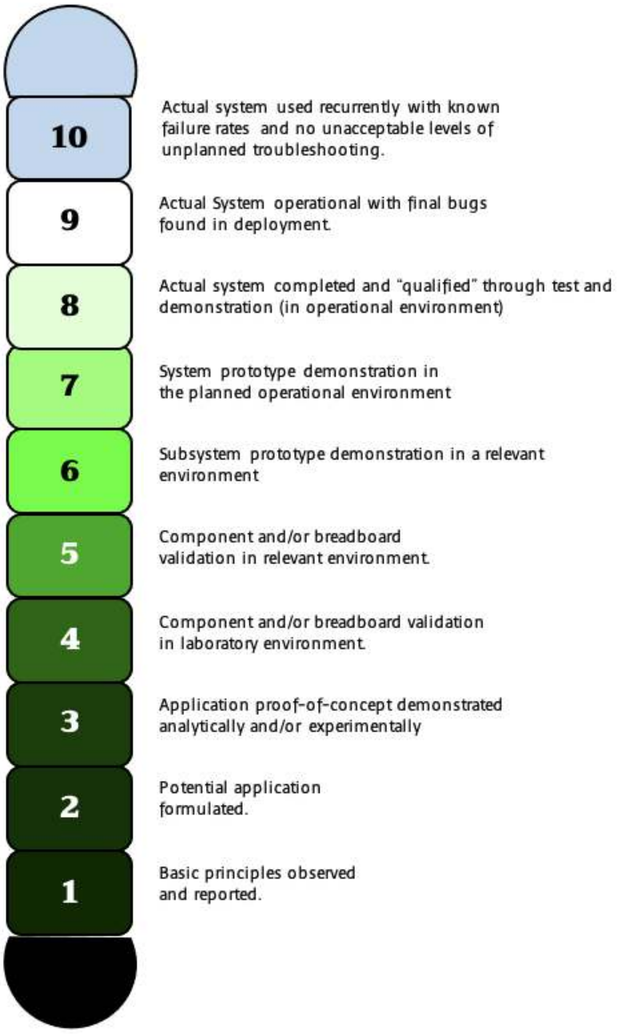

10
Actual system used recurrently with known failure rates and no unacceptable levels of unplanned troubleshooting.
9
Actual System operational with final bugs found in deployment.
8
Actual system completed and "qualified" through test and demonstration (in operational environment)
7
System prototype demonstration in the planned operational environment
6
Subsystem prototype demonstration in a relevant environment
5
Component and/or breadboard validation in relevant environment.
4
Component and/or breadboard validation in laboratory environment.
3
Application proof-of-concept demonstrated analytically and/or experimentally
2
Potential application formulated.
1
Basic principles observed and reported.

FIGURE 24: THE TEN LEVEL TECHNOLOGY READINESS LEVELS (TRL) USED FOR TECHNOLOGY ASSESSMENT BASED ON THE COMBINED SADIN ET AL, MANKINS AND STRAUB MODELS.

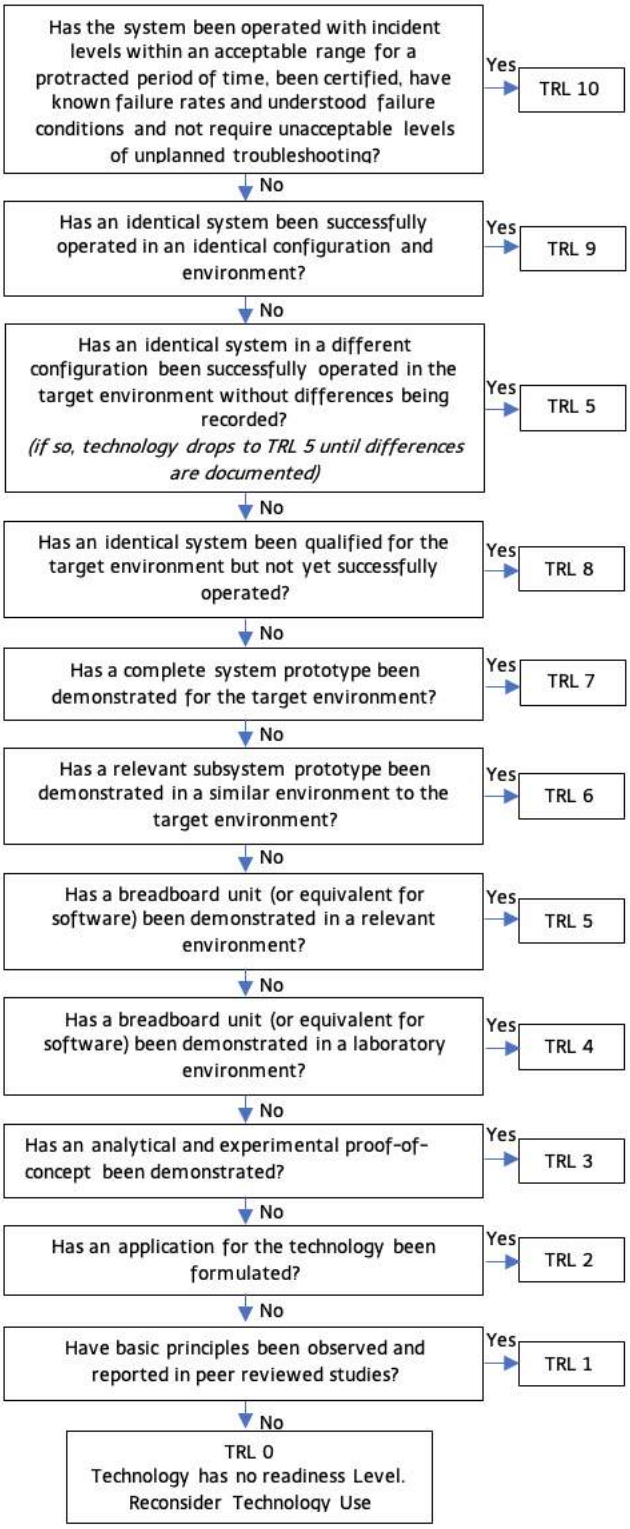

Has the system been operated with incident levels within an acceptable range for a protracted period of time, been certified, have known failure rates and understood failure conditions and not require unacceptable levels of unplanned troubleshooting?
Yes
TRL 10
No
Has an identical system been successfully operated in an identical configuration and environment?
Yes
TRL 9
No
Has an identical system in a different configuration been successfully operated in the target environment without differences being recorded?
(if so, technology drops to TRL 5 until differences are documented)
Yes
TRL 5
No
Has an identical system been qualified for the target environment but not yet successfully operated?
Yes
TRL 8
No
Has a complete system prototype been demonstrated for the target environment?
Yes
TRL 7
No
Has a relevant subsystem prototype been demonstrated in a similar environment to the target environment?
Yes
TRL 6
No
Has a breadboard unit (or equivalent for software) been demonstrated in a relevant environment?
Yes
TRL 5
No
Has a breadboard unit (or equivalent for software) been demonstrated in a laboratory environment?
Yes
TRL 4
No
Has an analytical and experimental proof-of-concept been demonstrated?
Yes
TRL 3
No
Has an application for the technology been formulated?
Yes
TRL 2
No
Have basic principles been observed and reported in peer reviewed studies?
Yes
TRL 1
No
TRL 0
Technology has no readiness Level.
Reconsider Technology Use

FIGURE 25: QUESTIONS TO ASSESS THE TRL OF A TECHNOLOGY AS ADAPTED FROM STRAUB.

By adopting a ten-level TRL framework, we can then use the weightings in Table 10 for our WDM assessment to obtain a final score for the efficacy of the tool under assessment.

TABLE 10: TECHNOLOGY READINESS ASSESSMENT

Technology Readiness Level											
Weighting	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Criteria	TRL 0	TRL 1	TRL 2	TRL 3	TRL 4	TRL 5	TRL 6	TRL 7	TRL 8	TRL 9	TRL 10

Using Figure 24, Figure 25 and Table 10 in conjunction yields our Technology Readiness Assessment.

For example, to obtain the final score in this assessment, the adjusted TRL of the application is multiplied by the weighting of the maximum level obtainable. An application that has been demonstrated in an attack environment would be assessed at TRL 7. The final score from the framework will be multiplied by the scaling factor of 0.7. This would provide a 70% weighting to the overall assessment in Table 6.

Focusing on the technology readiness levels of the examples discussed above we now display these in Figure 26.

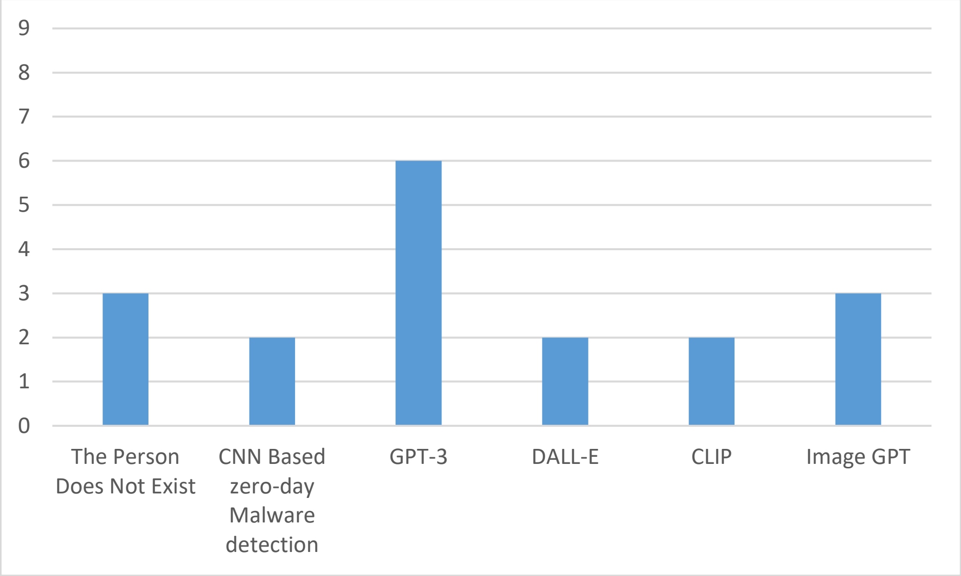


FIGURE 26: TRLS OF EXAMPLE GANS

6. Consideration and conclusions

GANs-powered systems are of great interest in the AI scenario because they open the door to more prediction capabilities, sometimes also more subtle. We know that AI-enhanced malware could take over industrial equipment or learn to mimic people's behaviours, but also simulate and stimulate them. We must pay attention to simulation capabilities because an intelligent system could deceive the human by establishing a relationship of trust, thus opening the way to far greater danger such as influencing or creating addiction. A possible and very devastating end is the real adversarial simulation using GAN technology, something not possible today but which might be attractive to the ill-disposed in the future. Some states have begun to understand the possibilities of simulation and deception by digital systems and try to counter its use through IT and political solutions as in the case of the State of California which, with bills AB-602[96] and AB -730[97] restricts the use of deceptive material.

To determine whether the international community has a particular interest in the various fields of application described, a study based exclusively on open sources was conducted. No particular evidence emerged on the massive use of these technologies, however, it immediately became clear how much the world of cybersecurity, probably due to its strongly IT nature, is closer and more inclined to the use of GANs. The amount of material on credential guessing and digital defence evasion stands out above all. What is most worrying is the ease of finding ready-to-use model implementations which, if they end up in the wrong hands, could represent a good starting point for designing adversarial attacks. Detection systems need to keep some of their components secret by hiding some parameters of the machine learning process in order not to be overwhelmed and to be considered adversarial-aware. But the 'security through obscurity' approach is ineffective alone and there is thus a need to introduce protection mechanisms such as randomisation to make exact replication of the machine learning procedure difficult.

From the reading and analysis of documents relating to these sectors, it is clear that technologically advanced cyber-states such as China and Japan are very interested in the development of cyber-GAN and in providing resources for research into these new technologies. It is also evident that a real use of GAN in production environments is currently limited to products that, through the discovery of anomalies, try to detect information that is harmful to the integrity of systems which are often subject to targeted and explorative attacks.

GANs have gone from being a purely mathematical reality to a reality applied in image generation. Although they are a great invention, they also suffer, like all artificial intelligence models, from problems related to the dataset such as obsolescence, data scarcity and context polarisation. In the field of intelligence, this represents a considerable problem as the assessments must be based on as reliable information as possible to produce intelligence of quality and which does not endanger the originator or whoever uses its properties. Therefore, the use of GANs for intelligence purposes cannot be limited to the development of increasingly sophisticated and accurate models, but requires the adoption of collateral solutions for the collection and enhancement of the large amount of data used to feed the neural networks, these might include:

- the use of adaptive and incremental learning models tailored to the target;
- representative and current data samples that do not bias predictions;
- systems capable of triaging and prioritising the available data; or support methodologies for forecasting trends.

[96] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602

[97] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730

We are at the beginning of an artificial intelligence revolution; just think of self-driving vehicles, robotics and new medical treatments. The prospects for improvement and the ease of access to these new technologies suggest that in the not-too-distant future, they could also be used for malicious purposes, so the intelligence community cannot fail to seize the opportunity to have mastery of the subject and be prepared.

Across all applications studied, we see that GANs generally have a low technology readiness level. While this is to be expected given the recent developments in the field, it provides ample evidence that further research should be conducted in this space; even if to simply lift the TRL of the stated application. Much of the literature does not even elucidate a specific benefit of GAN for the intelligence community

The applications assessed show promise that GAN and other machine learning algorithms are applicable to the intelligence cycle and will make the collection, validation, exploitation and dissemination of actionable intelligence easier, especially when confronted with the information explosion or big data problem. Currently, there is insufficient evidence to suggest the widespread adoption of the current tools without further development to elevate existing methods to a higher TRL. Further research should focus on the elevation of strategic machine learning applications to higher TRLs.

GANs are particularly strong when faced with image data due to the underlying data structures required to train them. This provides ample opportunity to focus on emerging applications such as data recovery, image validation and interpretation. Further work in this area is recommended and warranted. While offensive capabilities have not been widely assessed in this report, it is an emerging application of GAN and an exciting area for the cyber intelligence discipline. GANs should not be limited to just cyber intelligence. Research should focus on how they may be exploited to assist other areas of the IC that have large image-based analysis needs. This is particularly of relevance to the GEOINT and IMINT communities.

While we have gone beyond the scope initially described in this report by broadening our study to the application of other machine learning algorithms to the questions posed, a more robust idea of how GANs may be used in this space has been developed by understanding how machine learning is being applied. This report has defined cyber intelligence and generative adversarial networks. A brief literature review on both GAN and cyber intelligence has been provided to synthesise a working definition for the project. From this definition, a set of questions has been developed which will be answered in future documents to evaluate the current uses of GAN that apply to the cyber intelligence field.

Several recommendations have been made in this report:

(1) Further work should be conducted, expanding this study into cyber intelligence applications that use any machine learning method, not just GAN.
(2) Analysis should be undertaken to understand the role cyber intelligence plays in the current intelligence community and determine if a definition needs to be officially adopted.
(3) This initial assessment should be repeated with a focus on penetration and exfiltration (offensive cyber capabilities) using GAN.
(4) To assess the interactions of GAN with current use technologies, further study should be undertaken on technologies that yield TRL 7 or higher. Much work is ongoing in the artificial intelligence and machine learning space. Applications are being developed by OpenAI that may be suitable for such an assessment, noting the OpenAI Charter may prevent application in the intelligence space.

Noting the work in Tundis et al., our work could be automated using a machine learning approach. This would enable a larger sample to be rapidly assessed against our framework, including the input of domain-specific specialist knowledge. Such a larger study would yield greater precision in our method and would itself create a novel intelligence tool for the assessment of other cyber intelligence tools, current and emerging.

The framework that has been designed for the assessment of GAN applications relevant to the cyber intelligence field includes both lawful and illegal uses of the technology. A weighted matrix analysis method has been developed to enable the assessment of each known application. It includes an assessment against the intelligence activity cycle and the technology readiness level of the GAN application. The framework is ready for integration with other components of this project that have identified the uses of GAN relevant to cyber intelligence.

The current pattern suggested by the intelligence tools uncovered in the unclassified academic literature point to the use of GAN as a method of classifying large volumes of intelligence data, regardless of the domain specificity. Further applications suggest GANs will be useful in image-based applications and thus, should be used in IMINT applications where the crossover between IMINT and cyber intelligence exists. Many of the applications examined here are still under development. This is to be expected, as any tools that would be developed for use in the IC that are sufficiently advanced would be protected or closely guarded. With advances in technologies such as GPT-3, while the use would be a breach of the conditions, current examples suggest that OpenAI's tools can easily be weaponised for use in the intelligence community, specifically for use in cyber intelligence.

It is recommended that the frameworks created in this work be used in evaluating how pervasive GAN is in the cyber intelligence collection discipline. Our analysis has shown that GANs are not yet ubiquitous in the IC; however, some applications warrant exploration in both the traditional intelligence cycle and the counterintelligence domains. GAN can be used to assist in the cyber intelligence domain and further work should be conducted to develop specific tools to at least TRL 7 so that the benefits of GAN can be found. Once the technology has reached maturity at TRL 7 and beyond, further research will be able to assess the suitability for widespread adoption. This is an ongoing process and is vital to maintain a competitive edge.

6.1 Summary

In this report, we have explored the changing definition of cyber intelligence and placed it in the context of the five intelligence disciplines currently recognised in the IC. While cyber intelligence operations most likely fit in the current model of signals intelligence, there is sufficient overlap in all disciplines for cyber intelligence to be its own discipline when properly defined. We have avoided this problem by offering a simple definition of cyber intelligence which has allowed us to explore the use of generative adversarial networks in the collection, processing, integration, evaluation, analysis and interpretation of intelligence information. We have used a combined weighted decision matrix to assess novel GAN applications based on the intelligence cycle and technology readiness frameworks. While we note that much work on GANs is still at a relatively early stage, the applications currently theorised make this emerging topic one that requires further evaluation with targeted research to meet specific intelligence needs.

7. List of Figures

8. List of Tables

9. References

CHAPTER 3:

Buonocore, G. (2019). MalGAN: Evasione e rilevamento di malware neurali.

Bühler, T. (2019, August 22). Defending against GAN-made malware. Available at https://www.avira.com/en/blog/gan-made-malware

Cascavilla, G., Di Nucci, D., Slabber, J., Tamburri, D., Palomba, F., & van den Heuvel, W.-J. (2020). Counterterrorism for Cyber-Physical Spaces: A Computer Vision Approach.

Goodfellow, J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). Generative Adversarial Networks.

Hitaj, B., Ateniese, G., Gasti, P., & Perez-Cruz, F. (2019). PassGAN: A Deep Learning Approach for Password Guessing.

Huang, A., & Huang, Y. (2018). Towards Robust Malware Detection.

Huang, L., Joseph, A., Nelson, B., Rubinstein, B., & Tygar, J. (2011). Adversarial Machine Learning.

Kawai, M., Ota, K., & Dong, M. (2019). Improved MalGAN: Avoiding Malware Detector by Leaning Cleanware Features.

Labaca Castro, R., Schmitt, C., & Dreo Rodosek, G. (2019). Poster: Training GANs to Generate Adversarial Examples Against Malware Classification.

Shi, Y., Davaslioglu, K., & Sagduyu, Y. (2019). Generative Adversarial Network for Wireless Signal Spoofing.

Taflove, A., & Hagness, S. (2000). Computational electrodynamics: the finite-difference time-domain method.

Taeksoo, K., Moonsu, C., Kim, H., Jung Kwon, L., & Jiwon, K. (2017). Learning to Discover Cross-Domain Relations with Generative Adversarial Networks.

Truong, T., & Yanushkevich, S. (2020). Generative Adversarial Network for Radar Signal Generation.

Weiwei, H., & Ying, T. (2017). Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN.

CHAPTER 4 - 5:

Appleman, Roy Edgar. United States Army in the Korean War. South to the Naktong, North to the Yalu. Washington D.C.: Center of Military History, 2012.

Barreno, Marco, Blaine Nelson, Russell Sears, Anthony D. Joseph, e J. D. Tygar. «Can machine learning be secure?» 2006.

Borys, Christian. The day a mysterious cyber-attack crippled Ukraine. 4 July 2017. http://www.bbc.com/future/story/20170704-the-day-a-mysterious-cyber-attack-crippled-ukraine.

Bühler, Thomas. Defending against GAN-made malware. 22 August 2019. https://www.avira.com/en/blog/gan-made-malware.

Buonocore, Giuseppe. «MalGAN: Evasione e rilevamento di malware neurali.» 2019.

Cascavilla, Giuseppe, Dario Di Nucci, Johann Slabber, Damian A. Tamburri, Fabio Palomba, e Willem-Jan van den Heuvel. «Counterterrorism for Cyber-Physical Spaces: A Computer Vision Approach.» 2020.

Central Intelligence Agency. The Intelligence Cycle. 2007. https://www.cia.gov/kids-page/6-12th-grade/who-we-are-what-we-do/the-intelligence-cycle.html (consultato il giorno 09 30, 2020).

Chirgwin, Richard. IT 'heroes' saved Maersk from NotPetya with ten-day reinstallation blis. 25 January 2018. https://www.theregister.co.uk/2018/01/25/after_notpetya_maersk_replaced_everything/.

Cimpanu, Catalin. BleepingComputer. 6 July 2017. https://www.bleepingcomputer.com/news/security/m-e-doc-software-was-backdoored-3-times-servers-left-without-updates-since-2013/.

CNBC. There are 20 billion cyber attacks every day: Cisco. 11 May 2017. https://www.cnbc.com/video/2017/05/11/there-are-20-billion-cyber-attacks-every-day-cisco-.html.

Denning, Dorothy. «Cybersecurity's Next Phase: Cyber Deterrence.» The Conversation, 13 December 2016.

ESET North America. ‘Petya’ Ransomware: What we know now. 27 June 2017. https://www.eset.com/us/about/newsroom/corporate-blog/petya-ransomware-what-we-know-now-3/.

Fast, Barbara, Michael Johnson, e Dick Schaeffer. Cyber Intelligence: setting the landscape for an emerging discipline. Intelligence and National Security Alliance, 2011.

Glenn, Marcus. Failure of Nuclear Deterrence in the Cuban Missile Crisis. Montgomery: Air War College, Air University, 2017.

Goodfellow, Ian, Yoshua Bengio, e Aaron Courville. Deep Learning. Online. MIT Press, 2016.

Goodfellow, Jan J., et al. «Generative Adversarial Networks.» 2014.

Greenberg, Andy. How an Entire Nation Became Russia's Test Lab for Cyberwar. 2017 June 2017. https://www.wired.com/story/russian-hackers-attack-ukraine/.

Gronholt-Pedersen, Jacob. Maersk says global IT breakdown caused by cyber attack. 27 June 2017. https://www.reuters.com/article/us-cyber-attack-maersk/maersk-says-global-it-breakdown-caused-by-cyber-attack-idUSKBN19I1NO.

Grossman, Nadav. EternalBlue – Everything There Is To Know. 29 September 2017. https://research.checkpoint.com/eternalblue-everything-know/.

Harper, Michael. Energy Company RasGas Is Infected With Shamoon Virus. 31 August 2012. http://www.redorbit.com/news/technology/1112685657/shamoon-virus-rasgas-aramco-083112/.

Hitaj, Briland, Giuseppe Ateniese, Paolo Gasti, e Fernando Perez-Cruz. «PassGAN: A Deep Learning Approach for Password Guessing.» 2019.

Huang, Alex, e Yangyang Huang. «Towards Robust Malware Detection.» 2018.

Huang, Ling, Anthony. D. Joseph, Blaine Nelson, Benjamin I. P. Rubinstein, e J. D. Tygar. «Adversarial Machine Learning.» 2011.

Iasiello, Emilio. «Is Cyber Deterrence an Illusory Course of Action?» Journal of Strategic Security, Spring 2014: 54-67.

ICS-CERT. Joint Security Awareness Report (JSAR-12-241-01B) Shamoon/DistTrack Malware (Update B). 16 October 2012. https://ics-cert.us-cert.gov/jsar/JSAR-12-241-01B.

InferKit. Talk to Transformer. 2020. https://app.inferkit.com/demo (consultato il giorno 09 30, 2020).

Kaplow, Louis. «Pareto Principle and Competing Principles.» The Harvard John M. Olin Discussion Paper Series, 2005.

Katz, Brian. «The Collection Edge: Harnessing Emerging Technologies for Intelligence Collection.» CSIS Briefs, July 2020: 1-9.

Kawai, Masataka , Kaoru Ota, e Mianxing Dong. «Improved MalGAN: Avoiding Malware Detector by Leaning Cleanware Features.» 2019.

Kerras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Lehtinen Jaakko, e Timo Aila. «Analyzing and improving the image quality of stylegan.» Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

Khrushchev, Nikita. Khrushchev Remembers: The Last Testament. Boston: Bantam, 1976.

Kubecka, Chris. «How to Implement IT Security After a Cyber Meltdown.» [Slideshow]. 3 August 2015. https://www.blackhat.com/docs/us-15/materials/us-15-Kubecka-How-To-Implement-IT-Security-After-A-Cyber-Meltdown.pdf.

Labaca Castro, Raphael, Corinna Schmitt, e Gabi Dreo Rodosek. «Poster: Training GANs to Generate Adversarial Examples Against Malware Classification.» 2019.

Lebow, Richard Ned. Between Peace and War: The Nature of International Crisis. Baltimore: The Johns Hopkins University Press, 1981.

Libicki, Martin C. «Cyberdeterrence and Cyberwar.» In Cyberdeterrence and Cyberwar, di Martin C. Libicki, 27-37. Santa Monica, CA: RAND, 2009.

LogRhythm. NotPetya Technical Analysis. Boulder: July, 2017.

Lonsdale, David J. «Warfighting for Cyber Deterrence: a Strategic and Moral Imperative.» Springer, 02 February 2017.

Lowenthal, Mark, e Robert Clark. The Five Disciplines of Intelligence Collection. London: SAGE, 2016.

Lu, Shuqiang, et al. «New Era of Deeplearning-Based Malware Intrusion Detection: The Malware Detection and Prediction Based On Deep Learning.» 2019.

Malware Tech. Petya Ransomware Attack - What's Known. 27 June 2017. https://www.malwaretech.com/2017/06/petya-ransomware-attack-whats-known.html.

Mandt, EJ. «Integrating Cyber-Intelligence Analysis and Active Cyber-Defence Operations.» Journal of Information Warfare 16, n. 1 (2017): 31-48.

Martin Gilje Jaatun, Maria B Line, Tor Olav Grotan. «Secure Remote Access to Autonomous Safety Systems: A Good Practice Approach.» International Journal of Autonomous and Adaptive Communications Systems Vol. 2 No. 3, 2009: 297-312.

Microsoft. New ransomware, old techniques: Petya adds worm capabilities. 27 June 2017. https://cloudblogs.microsoft.com/microsoftsecure/2017/06/27/new-ransomware-old-techniques-petya-adds-worm-capabilities/?source=mmpc.

Mike Oppenheim, Steve Stone. A 'Wiper' in Ransomware Clothing: Global Attacks Intended for Destruction Versus Financial Gain. 29 June 2017. https://securityintelligence.com/a-wiper-in-ransomware-clothing-global-attacks-intended-for-destruction-versus-financial-gain/.

New Petya / NotPetya / ExPetr ransomware outbreak. 27 June 2017. https://www.kaspersky.com/blog/new-ransomware-epidemics/17314/.

Nicole Perlroth, Mark Scott, Sheera Frenkel. Cyberattack Hits Ukraine Then Spreads Internationally. 27 June 2017. https://www.nytimes.com/2017/06/27/technology/ransomware-hackers.html.

Nye, Joseph S. «Nuclear Lessons for Cyber Security.» Strategic Studies Quarterly, Winter 2011: 18-38.

Nye., Joseph S. «Deterrence and Dissuasion in Cyberspace.» International Security, President and Fellows of Harvard College and the Massachusetts Institute of Technology, 2017: 44-71.

Orme, John. «Deterrence Failures: A Second Look.» International Security, Spring 1987: 96-124.

Pagliery, Jose. The inside story of the biggest hack in history. 5 August 2015. http://money.cnn.com/2015/08/05/technology/aramco-hack/index.html.

Panikkar, K M. In Two Chinas: Memoirs of a Diplomat. London: Allen and Unwin, 1955.

You Won't Believe What Obama Says in this Video! Diretto da Jordan Peele. Prodotto da Buzz Feed . 2018.

Perlroth, Nicole. In Cyberattack on Saudi Firm, U.S. Sees Iran Firing Back. 23 October 2012. http://www.nytimes.com/2012/10/24/business/global/cyberattack-on-saudi-oil-firm-disquiets-us.html.

Philbin, Michael J. Cyber Deterrence: An Old Concept in a New Domain. Carlisle, PA, USA: U.S. Army War College, 2013.

Popper, Karl. Conjectures and Refutations: The Growth of Scientific Knowledge (2002 ed.). London: Routledge, 1963.

—. The Logic of Scientific Discovery. United Kingdom: Hutchinson & Co, 1959.

SANS ICS. Analysis of the Cyber Attack on the Ukrainian Power Grid. Washington DC: SANS, 2016.

Shi, Yi, Kemal Davaslioglu, e Yalin E. Sagduyu. «Generative Adversarial Network for Wireless Signal Spoofing.» 2019.

Slayton, Rebecca. «Why Cyber Operations Do Not Always Favor the Offense.» International Security, Harvard Kennedy School, February 2017: 1-3.

Symantec. Shamoon: Back from the dead and destructive as ever. 30 November 2016. https://www.symantec.com/connect/blogs/shamoon-back-dead-and-destructive-ever.

—. The Shamoon Attacks. 16 August 2012. https://www.symantec.com/connect/blogs/shamoon-attacks.

Symantec. Petya ransomware outbreak: Here's what you need to know. 24 October 2017. https://www.symantec.com/blogs/threat-intelligence/petya-ransomware-wiper.

Taeksoo, Kim, Cha Moonsu, Hyunsoo Kim, Lee Jung Kwon, e Kim Jiwon. «Learning to Discover Cross-Domain Relations with Generative Adversarial Networks.» 2017.

Timothy Naftali, Philip Zelikow. The Presidential Recordings, John F. Kennedy, The Great Crisis Volume II. New York: W.W. Norton and Company, 2001.

Truong, Thomas, e Svetlana Yanushkevich. «Generative Adversarial Network for Radar Signal Generation.» 2020.

Wang, Phil. This person does not exist. 2019. https://thispersondoesnotexist.com (consultato il giorno 09 30, 2020).

Weiwei, Hu, e Tan Ying. «Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN.» 2017.

Whitting, Allen S. China Crosses the Yalu: The Decision to Enter the Korean War. Stanford: Stanford University Press, 1960.

Williams, Heather J, e Ilana Blum. Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise. Santa Monica: RAND National Defense Research Institute, 2018.