

# Security and Privacy Concerns in Next-Generation Networks Using Artificial Intelligence-Based Solutions: A Potential Use Case



Murat Kuzlu, Ferhat Ozgur Catak, Yanxiao Zhao, Salih Sarp,  
and Evren Catak

**Abstract** Next-generation communication networks (NextG or 5G and beyond) have become more essential to be able to realize cutting-edge applications, such as autonomous cars, mobile healthcare and education, metaverse, digital twins, virtual reality, and many more. All those applications need high-speed, low latency, and secure data transmission. Artificial intelligence (AI) technologies are the main drivers and play a critical role because of their significant contribution to all layers in NextG, i.e., from the physical to the application layer. On the other hand, the security and privacy concerns for applications using AI-based methods in next-generation networks have not been fully investigated in terms of cyber vulnerabilities. This book chapter focuses on the AI-enabled applications on the physical layer of NextG networks, including multiple input multiple output (MIMO) beamforming, channel estimation, spectrum sensing, and intelligent reflecting surfaces (IRS), as well as provides a comprehensive analysis of the potential use case, i.e., channel estimation, along with its vulnerability under adversarial machine learning attacks with and without the defensive distillation mitigation method. According to simulations outcomes, AI-enabled Next-G applications are vulnerable to adversarial attacks, and the proposed mitigation methods are able to improve the robustness and performance of AI-enabled models under adversarial attacks.

---

M. Kuzlu

Batten College of Engineering & Technology, Old Dominion University, Norfolk, VA, USA  
e-mail: [mkuzlu@odu.edu](mailto:mkuzlu@odu.edu)

F. O. Catak (✉)

Department of Electrical Engineering and Computer Science, University of Stavanger, Rogaland,  
Norway  
e-mail: [f.ozgur.catak@uis.no](mailto:f.ozgur.catak@uis.no)

Y. Zhao · S. Sarp

Department of Electrical and Computer Engineering, Virginia Commonwealth University,  
Richmond, USA

E. Catak

Independent Researcher, Stavanger, Norway  
e-mail: [evren.catak@ieee.org](mailto:evren.catak@ieee.org)

**Keywords** Next-generation networks · 5G and beyond · Cybersecurity · Adversarial machine learning

## 1 Introduction

The next-generation networks, i.e., 5G and beyond, have been penetrated into all sectors, including infrastructure, computing, security, and privacy. The main goal of NextG networks is to realize cutting-edge applications, including metaverse, mobile healthcare, and education, autonomous cars, augmented reality (AR), virtual reality (VR), and others. It is expected that NextG networks will support very high data transmission (more than 100 Gbps), ultra-low latency (milliseconds), and a high cellular traffic capacity (10 million devices per square kilometer) [1–3]. Advanced communication technologies are key drivers to achieve these goals, which include millimeter wave (mmWave), massive multiple-input multiple-output (massive MIMO), and artificial intelligence (AI). In the literature, advanced communication technologies have been studied in [4–8]. In frequency bands above 24 GHz, mmWave provides many advantages in terms of throughput, capacity, and latency. The advanced version of MIMO, i.e., massive MIMO, can also significantly increase the quality throughput and capacity of the radio link by using a group of antennas at both the transmitter and receiver sides.

AI also plays an essential role in achieving these requirements to improve network applications' efficiency, latency, and reliability [9]. AI has been applied to especially several NextG applications at the physical layer, including beamforming, channel estimation, spectrum sensing, intelligent reflecting surfaces (IRS), and others. The authors in [4] investigate the role of AI-based solutions in deploying and optimizing 5G and beyond network operations. They stressed that NextG networks are different from current networks in terms of architecture, communication and computing technologies, and applications. The study [10] emphasized the contribution of AI-based solutions to NextG networks in terms of improving network performance and provided an extensive review of NextG networks using AI-based solutions, which focus on physical layer applications, including reconfigurable intelligent surface (RIS), massive MIMO, and multi-carrier (MC) waveform. These AI-based algorithms significantly improve the overall system performance for NextG networks.

On the other hand, AI-based algorithms brings security and privacy concerns. In the literature, there are several studies regarding this concern, e.g., model poisoning in the wireless research community is studied [11–16]. The authors in [17] proposed a robust framework to detect adversarial attacks for industrial artificial intelligence systems (IAISs). According to the results, the framework can detect several adversarial attacks, including DeepFool and fast gradient signed method (FGSM), with high accuracy and low delay. Since AI-enabled models could be vulnerable to adversarial attacks, AI-enabled models should be evaluated in terms of risk assessment, vulnerabilities, security and privacy concerns before deploying in the next-generation wireless communication networks.

This book chapter provides a comprehensive review of security and privacy concerns in the NextG network using AI-based solutions along with a potential use case. It also provides a brief description of widely used adversarial attacks and mitigation methods. The attacks include Fast Carlini & Wagner (C&W), Basic Iterative Method (BIM), Momentum Iterative Method (MIM), Projected Gradient Descent (PGD), and, Gradient Sign Method (FGSM), while mitigation methods include adversarial machine learning and defensive distillation. It also implements a potential use case, i.e., channel estimation, along with its vulnerability under adversarial attacks with and without the mitigation method.

## 2 Next Generation Networks Architecture

The next-generation networks (NextG or 5G and beyond) have been paying more attention from academia and industry to meet the demands of future applications, such as metaverse, mobile healthcare, autonomous cars, AR, VR, and many more. Significant improvements need to be performed in next-generation network architecture to meet requirements along with the driving force behind the evolution of wireless networks. Future applications have more rigid requirements in terms of data transmission and latency, which will force the limits of 5G networks. NextG networks are expected to enhance information transmission performance, i.e., up to 1 Tbps data rate and ultra-low latency (microseconds). One goal of NextG is to provide global coverage through satellite communication networks and underwater communications [18]. It is also expected NextG will offer energy-efficient and seamless wireless connections in a global scope as well as guarantee future application requirements, such as ultra-high throughput and ultra-low latency. The NextG architecture is also different from the traditional one, i.e., combined terrestrial and non-terrestrial networks, integration of fully AI-based models for all layers, and enhanced network protocol stack framework. Big data and AI will play a crucial role in NextG networks to meet the requirements in terms of efficient network management, distributed computing, resource sharing, and security and privacy concerns. The authors in [19] proposed an architecture to tackle these challenges. Figure 1 derived from [19] represents the NextG network architecture. The architecture consists of three layers, i.e., (1) Resource level, (2) Network function level, and (3) Service and application level. The first layer (resource level) provides the main resource for the upper layers, including communication, distributed cloud data, and computing resources. The second level (network function level) manages the resources and conducts the network functions for the service and application levels. The third level (service and application level) can generally be classified into two categories: (1) vertical services focusing on specific applications, e.g., vehicles or drones, and (2) horizontal services crossing different applications, e.g., reporting and tracking the location of users and their devices. This architecture also consists of four planes: (1) Sharing and cooperation plane, (2) Data collection plan, (3) AI plane, and (4) Security plane. The sharing and cooperation plane is the most important plane to address the decentral-

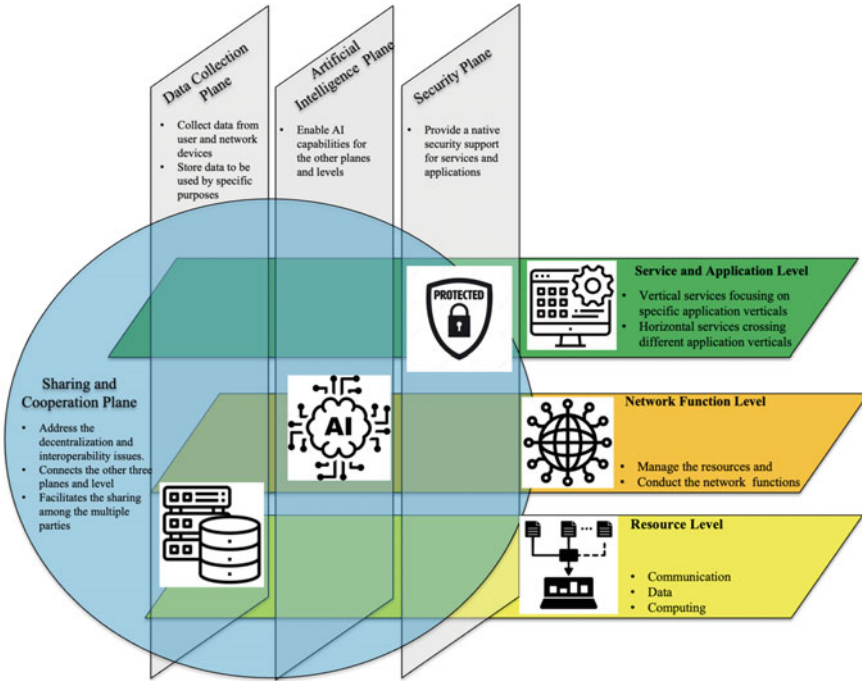


Fig. 1 Conceptual NextG network architecture [19]

ization and interoperability issues. It connects the other three planes and facilitates the sharing among multiple parties. The data collection plane is responsible for data collecting from the user and network devices as well as storing them to be used for specific purposes, such as network operation and optimization. The AI plane is the other important plane in this architecture. It provides AI-enabled capabilities for the security plane, resource level, network function level, and service and application level on demand. The last plane is the security plane, which provides native security support for networks, services, and applications.

### 3 CyberSecurity Framework for Next Generation Networks

Below is a proposed framework alongside some of widely used cybersecurity frameworks available. These frameworks help enterprises manage potential cyber risks efficiently and allow them to plan for future detection of cyber threats or investigation of security incidents during application and system development.

### **3.1 Available Cybersecurity Frameworks**

#### **3.1.1 ML Cyber Kill Chain**

Lockheed Martin created the Cyber Kill Chain methodology to support organizations understand and assess the risks they face from a potential cyber-attack. There are seven phases in a typical cyber-attack. These phases are *reconnaissance, weaponization, delivery, exploitation, installation, command and control/actuation, and actions on objectives*. Organizations can assess the potential effect of an effective cyber-attack on their operations by understanding the activities that occur during each cyber-attack phase.

#### **3.1.2 MITRE ATT&CK**

MITRE ATT&CK is designed to catalog the tradecraft and behaviors of adversaries to identify their activities better and generate an effective response strategy. By providing a common language and framework, organizations can more easily communicate their security processes and make attackers' techniques and tactics more identifiable.

#### **3.1.3 MITRE Atlas**

MITRE Atlas is a framework that includes information on how attackers might try to harm AI systems so that people can be better prepared to defend against those attacks. It is similar to MITRE Att&ck, which is a general framework for regular systems, not just AI systems. MITRE Atlas is a resource that includes information from security groups and academic research.

### **3.2 Proposed Framework**

Our study aims to address security threats and possible solutions by matching the Cyber Kill Chain and MITRE Atlas frameworks to catch and mitigate the vulnerabilities of AI models. These models will be a new part of potential AI-based 5G and beyond networks. Figure 2 illustrates 3 stages of the cyber kill chain for AI-based applications.

The first stage of creating an adversarial AI model is to gather information about the AI model we want to exploit. This can be done by finding datasets from publicly available sources like the weights and hyperparameters used in the training process. After this, the adversary can make their own replica of the AI model to make malicious inputs. The second stage is to build the replicated model, find its vulnerabilities, and generate malicious pilot signals that will be used as inputs to the target AI model. The third stage is to execute the target AI model with the malicious input signals.

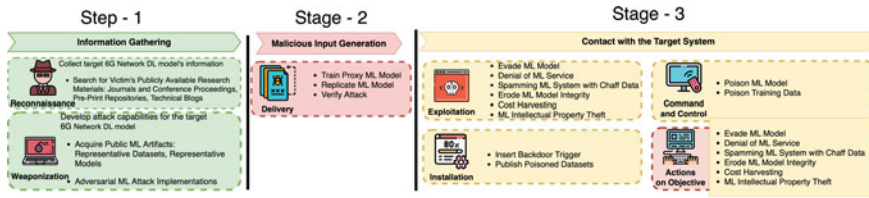


Fig. 2 Cyber kill chain for AI-based applications of 6G wireless communication networks

This will make the AI model fabricate incorrect outcomes, which the attacker can use to exploit the AI model and install a backdoor. With the backdoor, the adversary will take control of the AI model and the target system.

The tactics and methodologies described in the adversarial tactics and methodologies section of MITRE Atlas will take place in the Cyber kill chain stages.

- (i) The reconnaissance phase is when the adversary gathers information about the organization and its networks, systems, and employees. This information can create a profile of the organization, employees, network, and procedures. Social engineering attacks can be made with this information by the attackers.
- (ii) The weaponization phase occurs when the attacker utilizes the information collected during the reconnaissance phase to develop the tools they need to successfully make an attack against the organization. The adversary will use the information collected during the previous stage to choose the best delivery instrument to get the information it wants to deliver to the organization's IT infrastructure. The adversary can then concentrate on the delivery phase, using the same tools to provide information or files to the organization's IT infrastructure.
- (iii) The attacker must make use of a vulnerability in the organization's network once the information has been provided. The information gathered during the reconnaissance phase can be used to identify the software operated by the organization, operating systems, and applications running on the organization's systems.
- (iv) After the adversary has gathered information about the target organization during the reconnaissance phase, they will use this information to exploit the organization's network during the exploitation phase. The adversary will identify the best software, operating systems, and applications to exploit to install malicious software on the organization's systems. This malicious software will allow the adversary to manipulate or listen in the organization's network.
- (v) The command and control phase refers to when the attacker uses the malicious program installed during the exploitation phase to place further malicious software on the organization's systems. This allows them to control the organization's systems.

- (vi) The attacker may utilize the malicious program placed in the course of the exploitation phase to reach the organization's systems and loot information during the actions on objectives phase. They may also interfere with the organization's network.

The cyber kill chain is a process that details the steps an adversary takes to launch a successful cyberattack. Once the adversary has completed all the process steps, the organization's ability to employ its network can be affected.

### 3.3 Adversarial Machine Learning Attacks

There are two main types of adversarial machine learning models: the attacker's and the user's models. The attacker's goal is to manipulate the output of the user's model so that the attacker can benefit from the user's perspective [20]. Adversarial machine learning attacks are effective if the attacker accesses the training data. However, the proposed scheme is robust to the perturbations of the adversarial samples of the training data, which in turn makes the proposed scheme robust to adversarial machine learning attacks.

For example, to attack a deep learning model that predicts beamforming vectors, the attacker first needs to find a noise vector  $\sigma \in \mathbb{C}^k$  that will maximize the loss function  $\ell$  output. The attacker then uses the lowest possible budget to corrupt the inputs, which increases the distance (i.e., mean squared error (MSE)) between the model's prediction and the real beam vector. Therefore,  $\sigma$  is calculated as

$$\sigma^* = |\sigma|_p \leq \epsilon \arg \max \ell(\omega, \mathbf{x} + \sigma, \mathbf{y}) \quad (1)$$

where  $\mathbf{y} \in \mathbb{R}^m$  is the label (i.e., beamforming vectors), and  $p$  is the norm value, and it can be 0, 1, 2,  $\infty$ .

There are two primary methods of constructing adversarial examples: content-based and gradient-based [21]. Gradient-based attacks were chosen due to their simplicity and variety. Gradient-based attacks use the gradient of the loss function to generate adversarial examples, which are then incorrectly labeled.

- (i) Fast Gradient Sign Method (FGSM): FGSM tries to fool a neural network by changing the data given a little bit. The idea is to add noise to the data in the same direction as the loss function. The noise is controlled by a small number, epsilon. This makes the data look slightly different to the neural network, but enough to fool it.

$$\mathbf{x}^{adv} = \mathbf{x} + \epsilon \cdot \text{sign}(\nabla_{\mathbf{x}} \ell(\omega, \mathbf{x}, \mathbf{y})) \quad (2)$$

- (ii) Basic Iterative Method (BIM): The BIM attack is a variation of the FGSM single-step attack. It works by iteratively updating adversarial examples multiple times, with each value calculated in the neighborhood of the original input. The selected input with a smaller step size is manipulated by BIM iteratively.

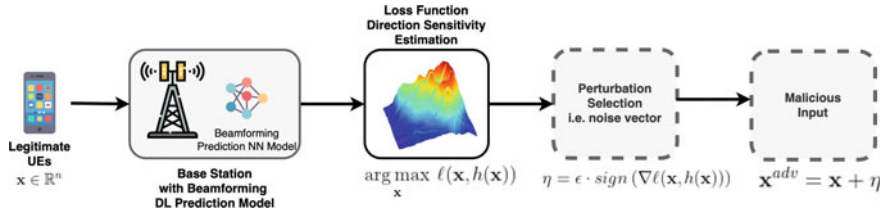


Fig. 3 Typical adversarial machine learning-based malicious input generation

FGSM is applied multiple times to a small step size  $\alpha$  instead of taking one significant step, i.e.,  $\epsilon/\alpha$ . By doing this, BIM creates less distortion while still fooling the neural network. However, this increases the computing cost and complexity. The BIM can be explained using the following equation.

$$\mathbf{x}_0^{adv} = \mathbf{x}, \mathbf{x}_{N+1}^{adv} = \text{Clip}_{\mathbf{x}, \epsilon} \{ \mathbf{x}_N^{adv} + \epsilon \cdot \text{sign}(\nabla_x \ell(\omega, \mathbf{x}_N^{adv}, y)) \} \quad (3)$$

- (iii) Projected Gradient Descent (PGD): PGD creates adversarial examples by starting the search at random points in a specified region and running several iterations to find an example that maximizes loss, which will be similar to a real input but different enough to trip up the ML model. PGD can generate more powerful attacks than BIM and FGSM. However, the size of the perturbation is kept smaller than a specified value, referred to as  $\epsilon$ , so that the adversarial example is still realistic and isn't just a random input.
- (iv) Momentum Iterative Method (MIM): MIM is another derivation of the BIM adversarial attack that improves the convergence of BIM by introducing a momentum term and integrating it into iterative attacks [22]. The step size of the  $\epsilon$  also determines the attack level of MIM as an attack parameter. MIM is better at finding the minimum amount of change needed to fool a model than BIM and can do so more quickly.

A characteristic adversarial ML-based malicious input generation process is indicated in Fig. 3.

### 3.4 Mitigation Methods for Wireless Networks

The 5G and future generations of networks relying on DL are vulnerable to adversarial machine learning attacks. Adversarial training and defensive distillation are two possible methods of mitigating these attacks and protecting wireless networks.



### 3.4.1 Adversarial Training

The goal of iterative adversarial training is to reduce the adversarial inputs’ impact on the training process. The DL model is first trained with the normal training data in iterative adversarial training. Then, the DL model is trained with the adversarial examples using the correct labels. The DL model is trained multiple times with normal and adversarial examples. However, iterative adversarial training is not practical. To increase the robustness of the victim model, it must train against all the different attack types and parameters which will take a quite long time.

The pseudo-code of adversarial training is shown in the algorithm 1.

---

#### Algorithm 1 Iterative adversarial training-based mitigation

---

**Input**  $h$ : vulnerable model,  $\Omega$ : attacks,  $\Pi$ : epsilon values,  $\mathbf{x}_{train}$ : training data,  $\mathbf{y}_{train}$  training data output,  $\mathbf{x}_{test}$ : test data,  $\mathbf{y}_{test}$ : test data output

**Output**  $\hat{h}$ : robust model

- 1: **for**  $\epsilon \in \Pi$  **do** {For each epsilon budget}
  - 2:   **for**  $attack \in \Omega$  **do** {For each epsilon budget}
  - 3:      $\mathbf{x}^{adv} \leftarrow attack(\mathbf{x}_{train}, \epsilon)$  {Generate malicious inputs with  $attack$  and  $\epsilon$  budget.}
  - 4:      $\mathbf{x}^{adv\_train} \leftarrow \mathbf{x} \cup \mathbf{x}^{adv}$  {Merge newly created malicious inputs  $\mathbf{x}^{adv}$  and  $\mathbf{x}_{train}$  }
  - 5:      $h.fit(\mathbf{x}^{adv\_train}, \mathbf{y}_{train})$  {Re-train the model  $h$  with new training data}
  - 6:   **end for**
  - 7: **end for**
- 

### 3.4.2 Defensive Distillation

Papernot et al. [23] proposed defensive distillation technique as an adversarial ML defense method against attacks. In knowledge distillation, a larger model (the teacher) is used to train a smaller model (the student). The teacher model is first trained with a high-temperature parameter to soften the softmax probability outputs of the DNN model. The student model is then trained using the outputs of the teacher model. The goal is for the student model to learn the knowledge of the teacher model but be smaller and faster. Equation 4 shows the modified softmax activation function as follows:

$$p_i = \frac{\exp(\frac{z_i}{T})}{\sum_j \exp(\frac{z_j}{T})} \tag{4}$$

where  $p_i$  is the probability of  $i$ -th class and  $z_i$  are the logits. The teacher model is used to predict each sample to acquire the training data’s soft labels which are used to train the student model. Figure 4 shows the overall steps for this technique.

The beamforming prediction model (i.e., student model) is drawn in Fig. 4 which is trained and used in base stations to protect against adversarial machine learning attacks. The student model inherits the training parameters after the teacher model’s

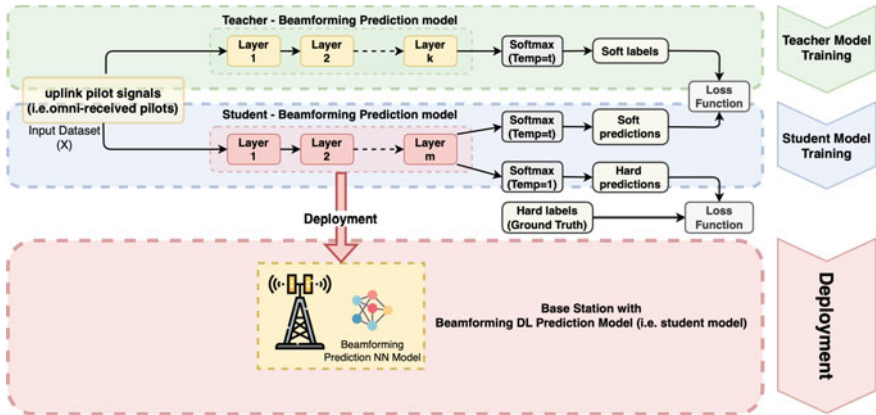


Fig. 4 Defensive distillation

training as a first step. In the second step, the student model is trained using the teacher models parameters, and the loss function is created with the actual labels and predictions of the student model. This technique allows preserving the teacher model’s knowledge to be compressed and transferred to the student model. And as a last step, student model is deployed to the base stations.

A technique called defense distillation can be used to reduce the effects of gradient-based untargeted attacks. This technique lowers the gradients to zero, making the standard objective function impractical.

## 4 Potential Use Cases

In this section, we will introduce several potential use cases including MIMO beamforming, spectrum sensing, channel estimation and IRS.

### 4.1 MIMO Beamforming

Signal to noise ratio is one of the key metrics for a channel that is affected by signal fading. Having diverse sources for a signal considerably reduces the error rate as each of the signal paths is not affected similarly. There are three ways to increase the diversity of the signals, i.e., Time diversity, Frequency, and space diversity. First, two uses use various times and frequencies, such as channel coding and OFDM. Space diversity benefits from the distribution of multiple antennas to capture different radio paths.

MIMO is one of the widely used RF technology that provides increased link capacity and spectral efficiency. MIMO systems utilize multiple antennas for both the receiver and transmitter ends to handle more data simultaneously.

Wireless signals can take various paths during the transmission between transmitter and receiver. Also, the change in the location of any antennas will create additional paths. This multipath propagation nature of the signal is caused by the objects along the transmission path. Previously, multipath propagation is seen as interference that causes signal degradation.

However, MIMO systems benefit from multipath propagation, where each additional signal path is considered as an additional channel to transmit additional data to the receiver. This is one of the main reasons that MIMO systems provide a robust link between the two ends. That is why the reliability of the MIMO systems depends on multipath propagation.

## 4.2 *Spectrum Sensing*

The electromagnetic spectrum that ranges 1 Hz to 3 THz is called the radio spectrum. It is one of the keys and limited resources that is not fully utilized due to region-based regulations and technical hardships. The majority of the existing radio spectrum is allocated to high-demand service providers, such as cellular communication, TV, and radio broadcasting. However, According to the report released by the Federal Communications Commission (FCC), there is still an underutilized spectrum, such as the licensed 0–6 GHz band having the 90% of underage [24, 25]. To increase the utilization of the limited spectrum, FCC recommends the use of free bands by a secondary user(s) until the primary user needs it. That's why "spectrum sensing" processes are developed to check specific bands to detect non-occupied frequency bands.

Spectrum sensing is also one of the notable research fields in cognitive radio (CR). CR is an intelligent software-based wireless communication concept that is introduced by Mitola in 1999 [26]. CR has a dynamic structure that senses and learns the wireless channels in its vicinity. It will then adopt the operating parameters to steer clear of user interference and congestion.

There has been a constant interest in spectrum sensing and related fields in the literature. For example, the study [27], provides a comprehensive survey of spectrum sensing for CR. Enabling algorithms, challenges, sensing standards, approaches, and cooperative and multi-dimensional spectrum sensing is presented. Also, the study [28], provides detailed spectrum sensing techniques such as the optimal likelihood ratio test, energy detection, matched filtering detection, cyclostationary detection, eigenvalue-based sensing, joint space-time sensing, and robust sensing methods.

Even though there are many studies and proposed methods for spectrum sensing, spectrum sensing is still subject to research because of the changeable nature of wireless communication channels, complexity, interferences, and noise in communication. AI methods would be a good alternative solution for spectrum sensing to deal with communication's complexity and changeable nature.

### 4.3 Channel Estimation

Transmitters and receivers utilize various mediums or channels to exchange information. In the case of wireless communication, a channel is simply the band of Radio Frequency that is used for the transmission of the signals. The characteristic and state of the channel is called channel state information (CSI). The transmitted signal ( $x(t)$ ) is exposed to three main distortions to some degree, i.e., attenuation by a factor of  $h_0$ , delay by a certain time  $\tau_0$  and noise, depending on the properties of the channel. The delay of  $\tau_0$  based on the electromagnetic wave's speed and attenuation  $h_0$  is determined by the transmitter/receiver gains, frequency, and propagation medium. To transmit a signal from one point to another point meaningfully, the received signal ( $y(t)$ ) needs to be decoded correctly. The first step to decode a signal is to understand the CSI such that the added noise and distortion can be rectified at the receiver. This process is called channel estimation. The signal at the receiver can be shown as:

$$y(t) = h_0 * x(t - \tau_0) \quad (5)$$

Scattered and reflected signals also reach the receiver with various delays and attenuation. These are also summed on the receiver side. Moreover, the mobility of the communication sides affects the attenuation  $h_l^t$  and delay  $\tau_l^t$  of the CSI by introducing a doppler frequency shift.

$$y(t) = \sum_{l=0}^l h_l^t * x(t - \tau_l^t) \quad (6)$$

where  $l$  is the specific path/tap at a time.

To fully utilize the capacity of the channel and increase the overall performance of the information transmission, channel estimation is one of the critical topics in wireless communication.

### 4.4 Intelligent Reflecting Surfaces (IRS)

IRS has been recognized as valuable ingenious technology [29]. This newly emerged technology could be perceived as the extension of massive MIMO [30]. It will enable increased data rate and channel capacity that NextG wireless communication requires without the vast amount of energy consumption and complexity of massive MIMO applications.

An IRS composes of a large number of predominantly passive elements, i.e., micro-strip type small antennas. Each of these elements' properties, such as load impedance, could be tunable by PIN diodes or varactors. PIN diodes are turned on or off to alter the phase-shift difference of IRS elements with different load impedances. Varactors' bias voltage is another parameter that can be utilized to tune phase shift by

altering the load impedance of each element. The reflected signals amplitude is also changed with the variable resistor's resistivity. By controlling the load impedance and resistivity of each element, different reflection coefficients are achieved individually.

If the phase shifts of individual elements are controlled in a way that the reflected signals are added constructively or destructively, the signals could be directed at certain guidance. An IRS controller is responsible for receiving the reconfiguration request communication. A field-programmable gate array (FPGA) could be employed to implement the IRS controller. Besides the passive elements, a few active IRS elements are also included in some of the IRS architectures. These active elements gather two orthogonal uplink communication links from both transmitter and receiver to predict the channel vectors and environment descriptors. AI-based techniques are adopted to utilize active elements as well.

## 5 A Potential Use Case: AI-Enabled Channel Estimation Model

In this section, we will take AI-based channel estimation modelling as a specific use case via presenting the dataset and experimental results. Experimental results cover the vulnerability analysis of the AI-enabled models to adversarial machine learning attacks with and without the selected mitigation method, i.e., defensive distillation. The model vulnerability will be evaluated through the MSE performance metric. MSE measures the average squared difference between the actual and predicted values. A high MSE score represents a high prediction error.

### 5.1 Dataset Preparation

In recent years, several network simulation tools have provided a wide range of examples for next-generation network communications systems, including NS3, OMNET++, NetSim, RemCom, MATLAB, and many more [31]. These tools are usually used for evaluating the performance of communication networks or dataset generation. In this study, a reference example in MATLAB 5G Toolbox [32], i.e., "Deep Learning Data Synthesis for 5G Channel Estimation," is selected to obtain datasets for DL-based models. It also allows to customize and generate communication components, such as waveforms, antennas, and channel models.

Channel estimation model is created with a single-input single-output (SISO) antenna by using demodulation reference signal (DM-RS) and the physical downlink shared channel (PDSCH) to generate 256 training datasets. Each dataset presents 8568 data points, i.e., 612X14X1 or 612 subcarriers, 14 OFDM symbols, and 1 antenna. Then, each data point is converted to a real-valued 612-14-2 matrix, i.e., from a complex (real and imaginary) 612-14 matrix. It is required to provide real

**Table 1** The channel estimation parameters with values

Channel parameter	Value	Channel parameter	Value
Delay profile	TDL-A/B/C/D/E	Modulation	16QAM
Delay spread	1–300 ns	Transmit antenna	1
Doppler shift	5–400 Hz	Receive antenna	1
NFFT	1024	Transmission direction	Downlink
Sample rate	30,720,000	Polarization	Co-Polar
Symbols per slot	14	Windowing	36
Slots per subframe	2	Slots per frame	20

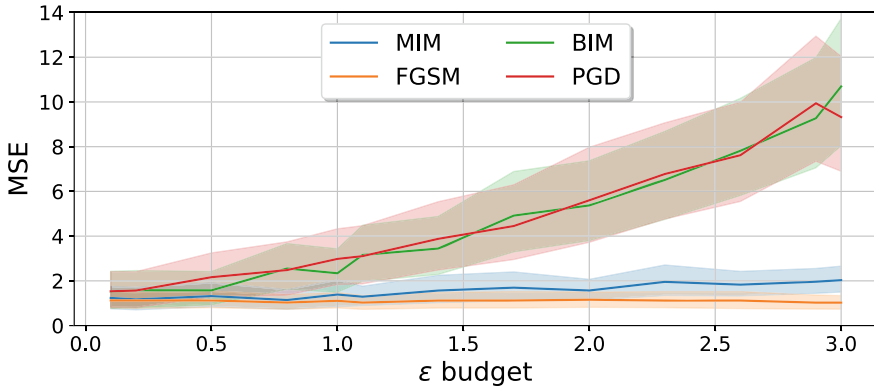
inputs instead of complex ones into the convolutional neural network (CNN) model used in the reference model during the training process. This is because the resource grids include complex data points, i.e., real and imaginary, in the channel estimation scenario. However, the CNN model handles the resource grids as 2-D images with real numbers. Finally, 4-D arrays (612-14-1-2N) are created from the training dataset with N as the number of training examples (256). In this study, 80% of the dataset is used for training, while 20% is used for testing.

For each dataset, a new channel characteristic is generated based on selected channel parameters and tuned through MATLAB 5G toolbox. Table 1 below provides the channel estimation scenario parameters with values.

## 5.2 Experimental Results

This section investigates the experimental results of an AI-powered channel estimation model against adversarial machine learning attacks. These results are represented in two ways: (1) line plots showing the impact of each adversarial machine learning attack (FGSM, MIM, BIM, and PGD) on the undefended and defended model performance, i.e., MSE, and (2) the table showing the performance (i.e., MSE) of the defended and undefended models for each adversarial attack. Figures 5 and 6 show the line plots, while Table 2 shows the prediction performance results of the defended and undefended AI-powered channel estimation models against adversarial attacks.

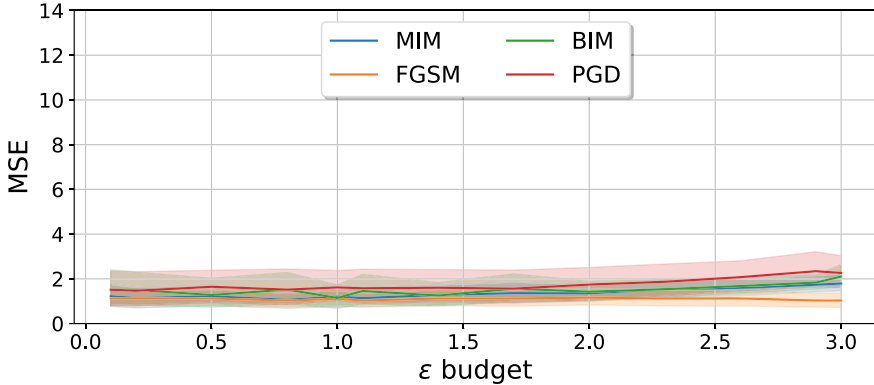
Figure 5 shows MSE values for the FGSM, MIM, BIM, and PGD attack methods for undefended models under attack powers from  $\epsilon = 0.01$  to  $\epsilon = 3.0$ . MSE values are close to each other for attack methods with a low power attack, i.e.,  $\epsilon < 0.5$ . However, these values dramatically increase along with higher power attacks ( $\epsilon > 0.5$ ). For example, MSE values can reach from 1.51, 153 to 10.69, 9.32 for BIM and PGD. The case is different for FGSM and MIM attacks, i.e., MSE values are low compared to BIM and PGD. The reason is that FGSM and MIM attacks are simple types of attacks, and then MSE values do not dramatically increase with high attack power. According to the results, the AI-powered models are exposed to



**Fig. 5** MSE comparison for undefended channel estimation model under adversarial attacks

adversarial attacks, especially PGD and BIM, and MSE can be very high under a heavy adversarial attack. Fortunately, the mitigation methods (such as adversarial training and defensive distillation) can significantly contribute to improving the AI-powered model’s robustness against adversarial attacks. In this study, the defensive distillation method is used as a mitigation method. The model performance is shown in Fig. 6 after applying the mitigation method for the selected adversarial attacks and attack powers in terms of MSE. According to Fig. 6, defended AI-powered models are still vulnerable to adversarial attacks. However, the models’ robustness is better under adversarial attacks. Models can resist high attack power. For example, MSE values can go from 1.51, 1.12, 1.22, and 1.51 to 2.1, 1.03, 1.79, and 2.26 with the lowest attack power ( $\epsilon = 0.1$ ) and the highest attack power ( $\epsilon = 3.0$ ) for BIM, FGSM, MIM, and PGD, respectively. The impact of the mitigation method on the model performance is different for some attack types. It has a high impact on the BIM and PDG attacks. This is because they are more complex attacks, and the MSE values can go very high under these attacks. As expected, the change in MSE is more compared to simple type attacks. For FGSM and MMI, the mitigation method has almost no impact on the models’ performance under adversarial attacks.

Table 2 shows the impact of attack power ( $\epsilon$ ) on undefended and defended models’ performance, i.e., MSE, for each adversarial attack in detail. The value of  $\epsilon$  ranges from 0.1 to 3.0. The higher value of  $\epsilon$  means a powerful attack. The lowest MSE value is 1.12 (under FGSM attack), and the highest MSE value is 10.69 (under BIM attack) for defended models. On the other hand, the lowest MSE value is 1.12 (under FGSM attack), and the highest MSE value is 2.26 (under PGD attack). MSE values dramatically go down from 10.69/9.32 to 2.10/2.26 for BIM/PGD after the mitigation method is applied. It is clear that the mitigation method significantly improves the model’s robustness, especially BIM and PGD. However, it cannot be said for FGSM and MIM attacks. According to Table 2, MSE values do not change as expected; they look closely to undefended and defended models, e.g., MSE values are 1.02 and 1.03 for undefended and defended models under an FGSM attack.



**Fig. 6** MSE comparison for defended channel estimation model under adversarial attacks

**Table 2** MSE results

$\epsilon$	Defended				Undefended			
	BIM	FGSM	MIM	PGD	BIM	FGSM	MIM	PGD
0.1	1.510613	1.121487	1.223872	1.513761	1.517611	1.123785	1.236382	1.534755
0.2	1.508010	1.121527	1.140600	1.468598	1.582042	1.123365	1.171341	1.566335
0.5	1.277997	1.121636	1.221769	1.646010	1.575610	1.122185	1.319650	2.164850
0.8	1.520606	1.031109	1.062509	1.520017	2.553312	1.029960	1.143569	2.482308
1.0	1.146857	1.109705	1.206056	1.617474	2.340146	1.108166	1.388278	2.982269
1.1	1.458215	1.031218	1.139865	1.580210	3.158410	1.029011	1.291160	3.105877
1.4	1.254450	1.121870	1.279377	1.603026	3.444848	1.119346	1.567613	3.878579
1.7	1.562587	1.124703	1.360767	1.563201	4.917432	1.121440	1.695858	4.451830
2.0	1.424730	1.160261	1.351300	1.744564	5.372514	1.156977	1.569844	5.602715
2.3	1.538028	1.122384	1.544590	1.869615	6.512692	1.117764	1.955341	6.778334
2.6	1.679046	1.125183	1.597183	2.076902	7.816463	1.120118	1.830526	7.618081
2.9	1.834858	1.032794	1.741633	2.342456	9.272282	1.026461	1.961652	9.940795
3.0	2.105044	1.032966	1.791616	2.264387	10.693936	1.026504	2.031071	9.321798

### 5.3 Observations

This study investigates undefended and defended AI-powered channel estimation models in NextG networks in terms of their vulnerabilities against adversarial attacks, i.e., FGSM, MIM, BIM, and PGD. Defensive distillation, as the migration method, is applied to the defended models. The overall results show that AI-powered models are vulnerable to adversarial attacks, and models’ vulnerabilities can be significantly reduced for some types of attacks, i.e., to be improved the models’ robustness. Observations can be given as follows:

- 1: AI-powered channel estimation models are vulnerable to adversarial attacks, espe-



cially, under a high attack power ( $\epsilon > 0.5$ ) for BIM and PDG.

2: The attack power ( $\epsilon$ ) has no impact on some adversarial attacks, i.e., FGSM and MIM.

3: The selected mitigation method can significantly increase the model robustness, especially for BIM and PGD.

4: The strongest attack is BIM, while the weakest is FGSM for undefended models.

5: The strongest attack is PDG, while the weakest is FGSM for defended models.

## 6 Security and Privacy Concerns

### 6.1 Homomorphic Encryption

Homomorphic encryption is a cryptosystem that enables computation on ciphertexts, producing an encrypted result that, when decrypted, matches the result of the operations as if they had been performed on the plaintext. The definition of homomorphic encryption (HE) scheme is given in [33] as follows:

**Definition 6.1** A homomorphic encryption scheme consists of a randomized polynomial-time algorithm,  $\mathcal{E}$ , which takes as input a security parameter  $\lambda$  and a message  $m \in M$  and outputs a ciphertext  $c = \mathcal{E}(1^\lambda, m)$ . The ciphertext space  $C$  is a polynomial-time deterministic function of  $\lambda$ . There is a randomized polynomial-time algorithm  $\mathcal{D}$ , which takes as input a security parameter  $\lambda$  and a ciphertext  $c \in C$ , and outputs a message  $m \in M$ , such that  $m = \mathcal{D}(1^\lambda, c)$ , with probability at least  $1 - \epsilon(\lambda)$ .

There are several homomorphic encryption schemes proposed in the literature, such as Paillier cryptosystem [34], ElGamal encryption scheme [35], Goldwasser-Micali (GM) scheme [36], Boneh-Goh-Nissim (BGN) scheme [37], and Paillier-HOM scheme [33]. Among these schemes, Paillier, ElGamal, and GM schemes are additive homomorphic and can support only simple operations on ciphertexts. On the other hand, BGN and Paillier-HOM are multiplicative homomorphic and can support more complex computations on ciphertexts.

A homomorphic encryption scheme is a pair of algorithms, **Enc** and **Dec**, with the following properties:

- (i) A polynomial-time randomized algorithm **Enc** which takes as input a security parameter  $\lambda \in \mathbb{N}$  and a message  $m \in M$  and outputs a ciphertext  $c = \text{Enc}(\lambda, m)$  such that  $c \in C$ .
- (ii) A polynomial-time randomized algorithm **Dec** which takes as input a security parameter  $\lambda \in \mathbb{N}$  and a ciphertext  $c \in C$  and outputs a message  $m \in M$  such that  $m = \text{Dec}(\lambda, c)$  with probability at least  $1 - \epsilon(\lambda)$ .

Additively homomorphic and multiplicatively homomorphic are the most common encryption types.

**Table 3** Computational cost and security of various HE schemes

Additively homomorphic schemes				
Scheme	Security	Runtime	Key	Ciphertext
Paillier [34]	IND-CPA	$O(n^6)$	$O(n^2)$	$O(n^2)$
ElGamal [35]	IND-CPA	$O(n^3)$	$O(n^2)$	$O(n^2)$
GM [36]	SEM-IND-CPA	$O(n^3)$	$O(n^2)$	$O(n^2)$
Multiplicatively homomorphic schemes				
Scheme	Security	Runtime	Key	Ciphertext
BGN [37]	IND-CPA	$O(n^4)$	$O(n^3)$	$O(n^3)$
Paillier-HOM [33]	IND-CPA	$O(n^6)$	$O(n^2)$	$O(n^2)$

**Definition 6.2** Homomorphic encryption  $\mathcal{E}$  is additively homomorphic if

- (i)  $\mathcal{E}(1^\lambda, m_1 + m_2) = \mathcal{E}(1^\lambda, m_1) + \mathcal{E}(1^\lambda, m_2)$
- (ii)  $\mathcal{E}(1^\lambda, m) = \mathcal{E}(1^\lambda, -m)$

**Definition 6.3** Homomorphic encryption  $\mathcal{E}$  is multiplicatively homomorphic if

- (i)  $\mathcal{E}(1^\lambda, m_1 m_2) = \mathcal{E}(1^\lambda, m_1) \times \mathcal{E}(1^\lambda, m_2)$
- (ii)  $\mathcal{E}(1^\lambda, m) = \mathcal{E}(1^\lambda, 1/m)$

Homomorphic encryption has several applications in distributed systems, cloud computing, data mining, and database security. In these applications, the data is stored in the cloud, and the data owner wants to keep its data private. The data owner can encrypt the data and store it in the cloud. The cloud user can perform computations on the encrypted data, and the result will also be encrypted. The data owner can decrypt the result and get the required information (Table 3).

## 6.2 Security of Homomorphic Encryption

Many HE schemes have been proposed in the literature in the past decade. The security of these schemes is analyzed under different security models. The security of HE schemes can be categorized under three different security models:

(i) **Partial homomorphic encryption**

With partial homomorphic encryption, the user can perform only limited operations on the ciphertext. In [38], Rivest et al. proposed a scheme that can support only a limited number of multiplications in the ciphertext. In this scheme, a ciphertext can be decrypted only if all the multiplications in the ciphertext are performed.

(ii) **Limited homomorphic encryption**

With limited homomorphic encryption, a ciphertext can be decrypted after any

number of operations are performed on the ciphertext. However, the number of operations that can be performed on the ciphertext is limited.

(iii) **Fully homomorphic encryption**

With fully homomorphic encryption, a ciphertext can be decrypted after any number of operations are performed on the ciphertext. In [33], Gentry, Sahai, and Waters proposed a scheme that supports both multiplications and addition in the ciphertext. In this scheme, a ciphertext can be decrypted after any number of multiplications and divisions are performed in the ciphertext. In [39], Brakerski and Vaikuntanathan proposed a scheme that supports only a limited number of multiplications in the ciphertext. In this scheme, a ciphertext can be decrypted after any number of multiplications are performed in the ciphertext.

The global model  $w_t$  can be trained on the aggregated dataset  $\cup_{i=1}^m D_i$  using any machine learning algorithm.

### 6.3 Federated Learning

In this section, we briefly describe the federated learning (FL) framework. We refer to [40, 41] for more details.

**Definition 6.4** (FL model) A federated learning (FL) model is a tuple  $M = (\{M_i\}_{i=1}^N, \{w_i\}_{i=1}^N, \{D_i\}_{i=1}^N, w, \{R_i\}_{i=1}^N)$ , where

1.  $M_i$  is the model trained on the local dataset  $D_i$  at client  $i$ ,
2.  $w_i$  is the weight of client  $i$ ,
3.  $D_i$  and  $D$  are the local and global datasets, respectively,
4.  $w$  is the global model trained on the global dataset  $D$ ,
5.  $R_i$  is the loss of client  $i$  on the global dataset  $D$ .

In the FL framework, the global model  $w$  is trained by optimizing the following objective:

$$\min_w \frac{1}{N} \sum_{i=1}^N w_i R_i(w, D_i). \quad (7)$$

The objective function (7) is minimized by training individual models  $M_i$  on local datasets and aggregating the models by averaging the weights. FL is an iterative approach to finding the best global model  $w$ . In each iteration, the client trains the individual model  $M_i$  on the local dataset  $D_i$  and sends the weights  $w_i$  to the central server. The server aggregates the weights and updates the global model  $w$ . The process is repeated until the global model converges. The FL framework has several advantages compared to traditional learning approaches, including improved privacy and security and lower communication and computational costs.

## 7 Summary

The NextG projects have been initiated to support a wide range of diverse applications, from AR/VR, metaverse, mobile healthcare, autonomous cars to digital twins and many more, by both the academia and the industry integrated with advanced cloud communication and data, computing, AI technologies in recent years. It has no doubt that AI is the most important tool in terms of significant contribution to all layers in NextG, i.e., from the physical to the application layer. On the other hand, the security and privacy concerns for NextG applications using AI-enabled solutions have not been fully addressed due to its complexity and multidisciplinary. This book chapter focuses on the AI-enabled applications on the physical layer of NextG networks, including beamforming, channel estimation, spectrum sensing, and IRS, and intends to investigate the vulnerability of AI-enabled channel estimation models under the selected adversarial attacks, such as FGSM, MIM, BIM, and PGD, with and without the selected mitigation (defensive distillation). According to the results, the AI-enabled channel estimation model is vulnerable to adversarial attacks. On the other hand, mitigation methods can significantly improve the performance and robustness of AI-enabled models under adversarial attacks.

## References

1. Agiwal M, Roy A, Saxena N (2016) Next generation 5G wireless networks: a comprehensive survey. *IEEE Commun Surv Tutor* 18(3):1617–1655. <https://doi.org/10.1109/COMST.2016.2532458>
2. Ziegler V, Yrjola S (2020) 6G indicators of value and performance. In: 2020 2nd 6G wireless summit (6G SUMMIT), pp 1–5. <https://doi.org/10.1109/6GSUMMIT49458.2020.9083885>
3. Johansson NA, Wang Y-PE, Eriksson E, Hessler M (2015) Radio access for ultra-reliable and low-latency 5G communications. In: *IEEE international conference on communication workshop (ICCW)*, pp 1184–1189. <https://doi.org/10.1109/ICCW.2015.7247338>
4. Letaief KB, Chen W, Shi Y, Zhang J, Zhang Y-JA (2019) The roadmap to 6G: AI empowered wireless networks. *IEEE Commun Mag* 57(8):84–90. <https://doi.org/10.1109/MCOM.2019.1900271>
5. Kaur J, Khan MA, Iftikhar M, Imran M, Haq QEU (2021) Machine learning techniques for 5G and beyond. *IEEE Access* 9:23472–23488
6. Wilhelmi F, Carrascosa M, Cano C, Jonsson A, Ram V, Bellalta B (2021) Usage of network simulators in machine-learning-assisted 5G/6G networks. *IEEE Wireless Commun* 28(1):160–166
7. Khan S, Hussain A, Nazir S, Khan F, Oad A, Alshehri MD (2022) Efficient and reliable hybrid deep learning-enabled model for congestion control in 5G/6G networks. *Comput Commun* 182:31–40
8. Piran MJ, Suh DY (2019) Learning-driven wireless communications, towards 6G. In: 2019 international conference on computing, electronics & communications engineering (ICCECE). IEEE, pp 219–224
9. Morocho Cayamcela ME, Lim W (2018) Artificial intelligence in 5G technology: a survey. In: 2018 international conference on information and communication technology convergence (ICTC), pp 860–865. <https://doi.org/10.1109/ICTC.2018.8539642>

10. Ozpoyraz B, Dogukan AT, Gevez Y, Altun U, Basar E (2022) Deep learning-aided 6G wireless networks: a comprehensive survey of revolutionary PHY architectures. [arXiv:2201.03866](https://arxiv.org/abs/2201.03866)
11. Dang S, Amin O, Shihada B, Alouini M-S (2020) What should 6G be? *Nat Electron* 3(1):20–29
12. Kuzlu M, Fair C, Guler O (2021) Role of artificial intelligence in the internet of things (IoT) cybersecurity. *Discov Internet Things* 1(1):1–14
13. Porambage P, Gür G, Osorio DPM, Liyanage M, Ylianttila M (2021) 6G security challenges and potential solutions. In: *Proceedings of IEEE joint European conference on networks and communications (EuCNC) 6G Summit*, pp 1–6
14. Siriwardhana Y, Porambage P, Liyanage M, Ylianttila M (2021) AI and 6G security: opportunities and challenges. In: *Proceedings of IEEE joint European conference on networks and communications (EuCNC) 6G Summit*, pp 1–6
15. Catak FO, Kuzlu M, Catak E, Cali U, Unal D (2022) Security concerns on machine learning solutions for 6G networks in mmwave beam prediction. *Phys Commun* 52:101626. <https://doi.org/10.1016/j.phycom.2022.101626>
16. Catak E, Catak FO, Moldsvor A (2021) Adversarial machine learning security problems for 6G: mmwave beam prediction use-case. In: *IEEE international Black Sea conference on communications and networking (BlackSeaCom)*, pp 1–6. <https://doi.org/10.1109/BlackSeaCom52164.2021.9527756>
17. Li G, Ota K, Dong M, Wu J, Li J (2020) Desvig: decentralized swift vigilance against adversarial attacks in industrial artificial intelligence systems. *IEEE Trans Ind Inf* 16(5):3267–3277. <https://doi.org/10.1109/TII.2019.2951766>
18. Yastrebova A, Kirichek R, Koucheryavy Y, Borodin A, Koucheryavy A (2018) Future networks 2030: architecture & requirements. In: *10th international congress on ultra modern telecommunications and control systems and workshops (ICUMT)*. IEEE, pp 1–8
19. Liu G, Huang Y, Li N, Dong J, Jin J, Wang Q, Li N (2020) Vision, requirements and network architecture of 6G mobile network beyond 2030. *China Commun* 17(9):92–104. <https://doi.org/10.23919/JCC.2020.09.008>
20. Faruk Tuna O, Ozgur Catak F, Taner Eskil M (2021) Exploiting epistemic uncertainty of the deep learning models to generate adversarial samples, *arXiv e-prints* [arXiv:2102.04150](https://arxiv.org/abs/2102.04150)
21. Vardhan R (2021) An ensemble approach for explanation-based adversarial detection, Ph.D. thesis
22. Fostiropoulos I, Shbita B, Marmarelis M. Robust defense against L p-norm-based attacks by learning robust representations
23. Papernot N, McDaniel P, Wu X, Jha S, Swami A (2016) Distillation as a defense to adversarial perturbations against deep neural networks. [arXiv:1511.04508](https://arxiv.org/abs/1511.04508)
24. Ma J, Li GY, Juang BH (2009) Signal processing in cognitive radio. *Proc IEEE* 97(5):805–823
25. Develi I et al (2020) Spectrum sensing in cognitive radio networks: threshold optimization and analysis. *EURASIP J Wireless Commun Netw* 2020(1):1–19
26. Mitola J, Maguire GQ (1999) Cognitive radio: making software radios more personal. *IEEE Personal Commun* 6(4):13–18
27. Yucek T, Arslan H (2009) A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Commun Surv Tutor* 11(1):116–130
28. Zeng Y, Liang Y-C, Hoang AT, Zhang R (2010) A review on spectrum sensing for cognitive radio: challenges and solutions. *EURASIP J Adv Signal Process* 1–15
29. Sarp S, Tang H, Zhao Y (2021) Use of intelligent reflecting surfaces for and against wireless communication security. In: *IEEE 4th 5G World Forum (5GWF)*. IEEE, pp 374–377
30. Rusek F, Persson D, Lau BK, Larsson EG, Marzetta TL, Edfors O, Tufvesson F (2012) Scaling up mimo: opportunities and challenges with very large arrays. *IEEE Signal Process Mag* 30(1):40–60
31. 5G Simulacion Software, Network Simulation Tools. <https://se.mathworks.com/products/5g.html>
32. Matlab 5G Toolbox. <https://www.mathworks.com/products/5g.html>. Accessed 30 Sept 2021

33. Gentry C (2009) Fully homomorphic encryption using ideal lattices. In: Proceedings of the forty-first annual ACM symposium on theory of computing, STOC '09, Association for Computing Machinery, New York, NY, USA, 2009, pp 169–178. <https://doi.org/10.1145/1536414.1536440>
34. Paillier P (1999) Public-key cryptosystems based on composite degree residuosity classes. In: Stern J (ed) *Advances in cryptology—EUROCRYPT '99*. Springer, Heidelberg, pp 223–238
35. Elgamal T (1985) A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans Inf Theor* 31(4):469–472. <https://doi.org/10.1109/TIT.1985.1057074>
36. Goldwasser S, Micali S (1982) Probabilistic encryption and how to play mental poker keeping secret all partial information. In: Proceedings of the fourteenth annual ACM symposium on theory of computing, STOC '82, Association for Computing Machinery, New York, NY, USA, 1982, pp 365–377. <https://doi.org/10.1145/800070.802212>
37. Boneh D, DeMillo RA, Lipton RJ (1997) On the importance of checking cryptographic protocols for faults. In: Fumy W (ed) *Advances in cryptology—EUROCRYPT '97*. Springer, Heidelberg, pp 37–51
38. Rivest RL, Dertouzos ML (1978) On data banks and privacy homomorphisms
39. Brakerski Z, Vaikuntanathan V (2011) Fully homomorphic encryption from ring-LWE and security for key dependent messages. In: Rogaway P (ed) *Advances in cryptology—CRYPTO 2011*. Springer, Heidelberg, pp 505–524
40. McMahan HB, Moore E, Ramage D, Hampson S, Arcas BA (2017) Communication-efficient learning of deep networks from decentralized data. In: International conference on artificial intelligence and statistics
41. Konečný J, McMahan HB, Yu FX, Richtarik P, Suresh AT, Bacon D (2016) Federated learning: strategies for improving communication efficiency. In: NIPS workshop on private multi-party machine learning. [arXiv:1610.05492](https://arxiv.org/abs/1610.05492)