ISSN: 0974-6471, Vol. 16, No. (1) 2023, Pg. 12-45

Oriental Journal of Computer Science and Technology

www.computerscijournal.org

A Reinforcement Learning Paradigm for Cybersecurity Education and Training

GABRIEL KABANDA[1]*, COLLETOR TENDEUKAI CHIPFUMBU[2], and TINASHE CHINGORIWO[3]

[1]Machine Learning Woxsen School of Business, Woxsen University, Hyderabad, India.
[2]Department of Information and Masrketing Sciences Midlands State University
Faculty of Business Sciences.
[3]Faculty of Technology, Zimbabwe Open University.

Abstract

Reinforcement learning (RL) is a type of ML, which involves learning from interactions with the environment to accomplish certain long-term objectives connected to the environmental condition. RL takes place when action sequences, observations, and rewards are used as inputs, and is hypothesis-based and goal-oriented. The key asynchronous RL algorithms are Asynchronous one-step Q learning, Asynchronous one-step SARSA, Asynchronous n-step Q-learning and Asynchronous Advantage Actor-Critic (A3C). The paper ascertains the Reinforcement Learning (RL) paradigm for cybersecurity education and training. The research was conducted using a largely positivism research philosophy, which focuses on quantitative approaches of determining the RL paradigm for cybersecurity education and training. The research design was an experiment that focused on implementing the RL Q-Learning and A3C algorithms using Python. The Asynchronous Advantage Actor-Critic (A3C) Algorithm is much faster, simpler, and scores higher on Deep Reinforcement Learning task. The research was descriptive, exploratory and explanatory in nature. A survey was conducted on the cybersecurity education and training as exemplified by Zimbabwean commercial banks. The study population encompassed employees and customers from five commercial banks in Zimbabwe, where the sample size was 370. Deep reinforcement learning (DRL) has been used to address a variety of issues in the Internet of Things. DRL heavily utilizes A3C algorithm with some Q-Learning, and this can be used to fight against intrusions into host computers or networks and fake data in IoT devices.

Article History

Received: 16 December 2022

Accepted: 23 January 2023

Keywords

Artificial Intelligence;
Cybersecurity;
Deep Learning;
DRL applications;
E-learning;
Machine Learning;
Reinforcement Learning.

CONTACT Gabriel Kabanda ✉ gabrielkabanda@gmail.com Machine Learning Woxsen School of Business, Woxsen University, Hyderabad, India.

Doi: http://dx.doi.org/10.13005/ojcst16.01.02

Introduction

With the growth of the Internet, cyberattacks are evolving quickly, and the state of cybersecurity is not optimistic. Computers, networks, programs, and data are all protected by a variety of technologies and procedures called "cybersecurity" in order to prevent assaults and unauthorized access, modification, or destruction.[1,21] Firewalls, antivirus software, and intrusion detection systems are all components of a network security system (IDS). Unauthorized system behavior, such as use, copying, modification, and destruction, can be found, ascertained, and identified with the aid of IDSs.[1] Both internal and external invasions constitute security breaches.

Machine Learning (ML) is the use and development of computer systems that can learn and adapt without explicit direction, including using algorithms and statistical models to examine data patterns and draw inferences from them.[2] The two fundamental characteristics of machine learning (ML) are automatic analysis of large datasets and creation of models for broad relationships between data. A type of Artificial Intelligence (AI) technology called Machine Learning (ML) enables a system to learn without any explicit programming.[3] As a result, machine learning can also be thought of as a subset of artificial intelligence, allowing a machine to make predictions, automatically learn from data, and improve performance based on prior experiences.[4] The primary goal of the ML technique is to allow computers to learn on their own without the aid of humans.[5,3] As shown in Figure 1, ML is primarily split into four groups based on the learning procedures, including supervised, unsupervised, semi-supervised, and reinforcement learning approaches. Image recognition, automatic language translation, medical diagnosis, stock market trading, speech recognition, traffic prediction, online fraud detection, virtual personal assistants, email spam and malware filtering, self-driving cars, and product recommendations are examples of real-world applications for machine learning.[6]

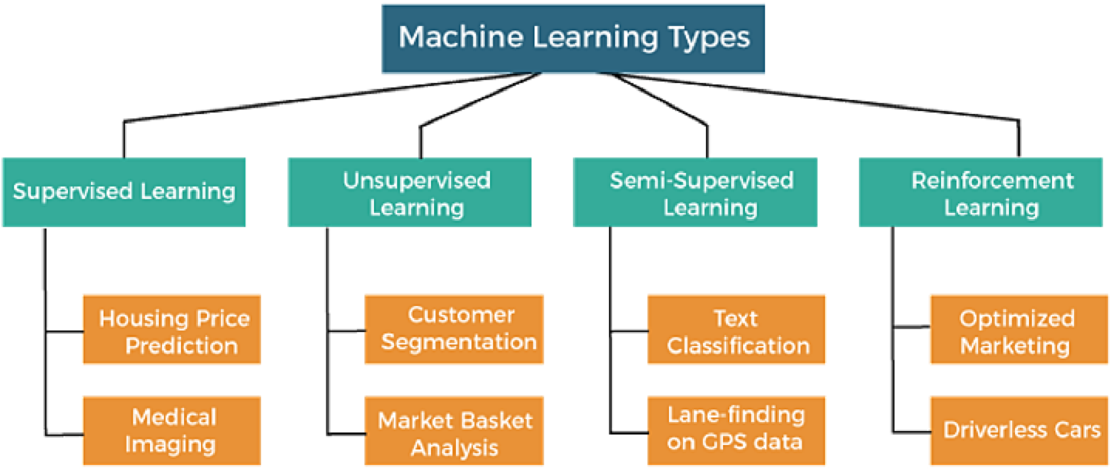


Fig. 1: Machine learning algorithms and their uses.

Background

Reinforcement Learning (RL) is a special type of ML where the input is a set of actions, observations and incentives.[2] RL is goal-oriented learning from interactions. RL learns about, from, and while interacting with the external environment. It uses systems (agents) that improve performance based on their interaction with the environment. Agents use reinforcement learning through interaction with the environment to learn a sequence of actions (inputs) that maximizes the reward signal as measured by the reward function. In reinforcement learning, the raw data (or input features) are also unlabeled and algorithms find patterns through trial and error and reward human experts accordingly. A goal is defined by a reward signal that needs to be maximized. A state is usually described by a feature vector. Robots, games and process control are some of the commonly used RL examples. The concept of RL is illustrated on Figure 2.

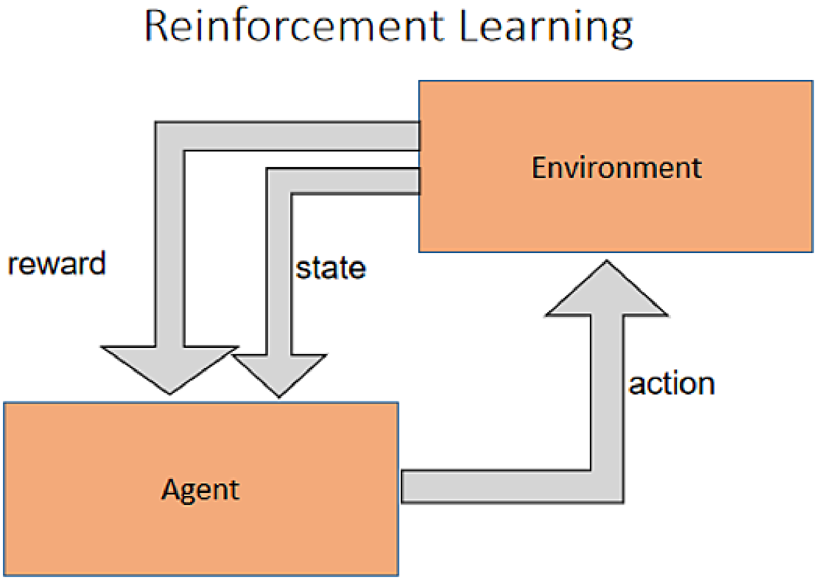


Fig. 2: Reinforcement Learning

The characteristics of RL include that the learner is not given instructions on what to do, therefore through a process of trial and error, there is a need to explore and take advantage of opportunities while there is a chance of a delayed reward. The entire issue of a goal-directed agent interacting with an uncertain environment is taken into account by RL.

The four major components of Reinforcement Learning (RL) systems are.

a) *Policy,* which instructs what to do. A policy is a mapping of the perceived environmental conditions to the actions to be taken when such states are present. Given the current environment state, a reinforcement learning agent chooses behaviors based on a policy.

b) *Reward signal,* which specifies what, is good. The reward signal identifies the objective. The environment transmits a single number known as the reward to the reinforcement learning agent on each time step. The agent's goal is to maximize its overall reward over the long run. Changes to the policy are made using the reward signal.

c) *Value function* that forecasts what is favourable. The value function shows what is excellent over the long term, whereas the reward signal shows what is beneficial over the short term. The total reward an agent can anticipate to accrue over the course of their lifetime, starting in that condition, is the value of that state. One should calculate the value using the incentives that would be offered in the states that would likely follow the current condition. Future prizes may be subject to a temporal discount using a factor between [0, 1]. It is important to utilize the values when making and assessing decisions. Decisions about what to do are based on value assessments. Instead of seeking the largest reward, one would like to take behaviors that result in the highest value states. The environment directly rewards you. Values must be updated on a regular basis based on the series of observations an agent makes throughout its existence.

d) Optional environment management *model* that controls what happens next. The RL policy objective function is $\pi\theta$ (a|s)

Asynchronous gradient descent is a technique for optimizing controllers in asynchronous reinforcement learning. This is helpful for deep reinforcement learning because deep neural networks, which are labor-intensive to train, are used as controllers. A reinforcement learning algorithm called Q-learning searches for the optimum course of action given the current situation.[2] The goal of Q-learning is to discover a policy that maximizes overall reward. In Q-learning, the "q" stands for quality. In this context, quality refers to how helpful a particular activity is in obtaining a potential reward. By preserving an estimated utility value Q(s,a) for each action in each stage, Q-learning strengthens value iteration. The highest possible Q value among all the potential actions at a state is simply the utility of that state, or U(s), or Q(s). The key asynchronous RL algorithms are.

i. Asynchronous one-step Q-learning
ii. Asynchronous one-step SARSA
iii. Asynchronous n-step Q-learning
iv. Asynchronous Advantage Actor-Critic (A3C)

The challenges of RL occur when

- there are many states and actions
- the episode can end without reward
- there is a ‘narrow’ path to reward.

With the rise of sophisticated and effective targeted cyberattacks worldwide, the threat of cyberattacks is only getting worse. Cybersecurity experts who are sufficiently motivated and skilled to prevent, detect, respond to, or even lessen the impact of such threats are desperately needed to address this problem. To this purpose, numerous graduate and undergraduate cybersecurity educational programs and concentrations have been formed in recent years. Additionally, a number of government standard-setting projects in the area of cybersecurity have emerged to support the development of cybersecurity education. Educational institutions confront numerous challenges when developing a cybersecurity curriculum due to the transdisciplinary (and occasionally multidisciplinary) nature of cybersecurity. In this context, we think it is desirable to present a broad overview of the entire efforts made so far in the direction of developing cybersecurity curricula. Like supervised or unsupervised learning, reinforcement learning (RL) is a machine learning paradigm that teaches an agent the optimum course of action to take in order to maximize its rewards in a certain environment.[2]

Statement of the Problem

It has been established that RL research has significantly aided in the defense of distributed cyberphysical systems. Given that artificial intelligence has the capacity to raise the security level of the defended distributed systems to the cutting-edge level often reached by attackers, its contribution to cybersecurity is of utmost importance. The methods used by computer programs to learn to produce results from experiments are divided into three paradigms in the subject of machine learning: supervised, unsupervised, and reinforcement learning (RL). In supervised learning, labels from the input data are used to train the model; in unsupervised learning, patterns found in the input data are used to train the model; and in real-time learning (RL), a software agent learns to respond autonomously to an environment that it is not yet familiar with.

Machine Learning (ML) is primarily split into four groups based on the learning procedures, including supervised, unsupervised, semi-supervised, and reinforcement learning approaches. Semi-supervised machine learning is a sort of algorithm that falls in between supervised and unsupervised machine learning. In order to create the desired output, it therefore combines supervised and unsupervised learning methods and uses a combination of labeled and unlabeled datasets during the training phase. Although semi-supervised learning combines supervised and unsupervised learning and uses data with a few labels, the majority of the data it uses is unlabeled. Instead of using solely labeled data as in supervised learning, the primary goal of semi-supervised learning is to make effective use of all available data. Reinforcement Learning (RL) is a special type of ML where the input is a set of actions, observations and incentives. The four major components of Reinforcement Learning (RL) systems are policy, reward signal, value function and an optional environment management model that controls what happens next. The key asynchronous RL algorithms are Asynchronous one-step Q-learning, Asynchronous one-step SARSA, Asynchronous n-step Q-learning and Asynchronous Advantage Actor-Critic (A3C).

Deep Learning is a recent area of study in machine learning. The creation of a neural network that mimics the human brain for analytical learning is the driving force behind it. In terms of performance and training speed, the on-policy A3C algorithm appears to be the most effective asynchronous reinforcement learning algorithm. It is of paramount importance to establish the Reinforcement Learning (RL) paradigm for cybersecurity education and training. What actual benefits can reinforcement learning provide for the cybersecurity industry? We will be able to determine the cybersecurity sectors that benefit from RL-based contributions the most, as well as the kind of contribution and the amount of research devoted to it, by providing an answer to this question. These areas include software/system protection, attacker/defender games, security policy development, and malware/intrusion detection.

Main Purpose

The main purpose for the research paper is to establish the Reinforcement Learning (RL) paradigm for cybersecurity education and training.

Research Objectives

The key research objectives of the research are to

a) Differentiate between Reinforcement Learning and other machine learning taxonomies (supervised, semi-supervised and unsupervised).

b) Determine the effect of ML on cybersecurity training and education.

c) Examine how ML is being used by commercial banks in Zimbabwe to tackle cybersecurity threats in order to highlight the significance of cybersecurity education and training.

d) Evaluate the performance of the Q-learning algorithm and the A3C method for reinforcement learning.

e) Identify the Reinforcement Learning paradigm that may affect cybersecurity training and education.

Research Questions

The research questions include the following

a) How is Reinforcement Learning distinguished from other categories of Machine Learning (ML)?

b) What is the impact of ML on cybersecurity education and training?

c) How is machine learning (ML) used to address cybersecurity concerns as observed in cybersecurity education and training in Zimbabwean commercial banks?

d) How do the structure and functionality of the Reinforcement Learning Q-Learning algorithm and the Reinforcement A3C Algorithm compare?

e) Which unique Reinforcement Learning traits might affect cybersecurity training and education?

Literature Review

As was shown in Figure 1, Machine Learning (ML) is primarily split into four groups based on the learning procedures, including supervised, unsupervised, semi-supervised, and reinforcement learning approaches. The four different learning types of ML and their associated algorithms are covered in detail in the following sections. These applications, however, depend on the type of ML types used.

Machine Learning paradigms

Supervised learning Approach

Supervised learning approaches need humans to give input and required output, in addition to providing feedback about the prediction accuracy in the training process.[3] This type of approach uses classes that are predetermined and the classes are created in a method of finite set normally defined by the human.[4] This practically means that a certain segment of data will be labeled with these classifications. Figure 3 illustrates the supervised learning process.

Training and testing are the two crucial phases of the supervised learning process. In the training stage, training data is taken into consideration as the input, and the learning algorithm studies the structures to build the learning model.[8] On the other side, during the testing phase, the learning model makes a prediction for the test data using the execution engine.[9] Although it provides the final prediction, supervised learning does not define the probability for inputs where the predicted outcome is clear because the objective is frequently to train the computer to learn a classification system that has already been developed by a human. Results for a data set with features and labels are produced via supervised learning.[10] As a result, the algorithm will next be given a set of characteristics together with the intended outputs as inputs. It then learns by comparing its intended output with the intended outputs to identify any faults and then modifying the model as necessary.[11] As long as the inputs are available, it is not necessary to use the constructed model. However, nothing can be inferred about the outputs if part of the input values are unavailable or absent. Regression and classification are the two categories into which supervised machine learning problems are divided.[12] The common uses of supervised learning include speech recognition, fraud detection, medical diagnosis, image segmentation, and spam and fraud detection.[13] The discussion of unsupervised learning will be covered in the next section.

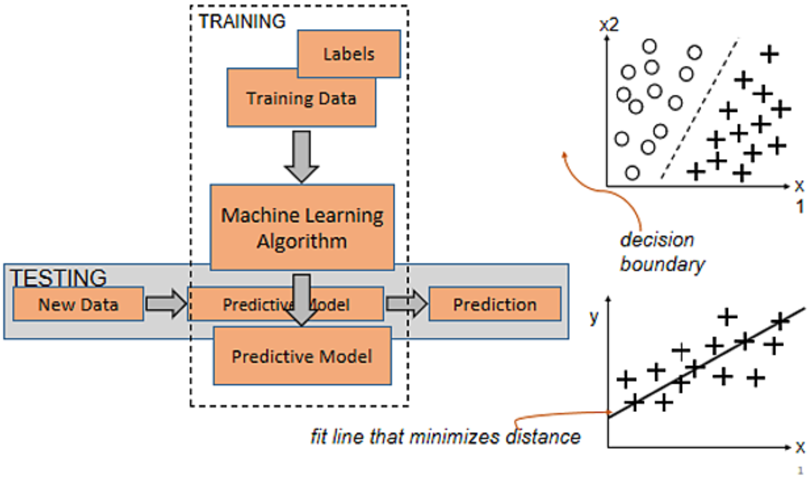


Fig. 3: Supervised learning process

Unsupervised Learning Approach

Models are not supervised using training datasets when utilizing the machine learning technique known as unsupervised learning.[3] There is no requirement for supervision, as the name suggests. In order to extract hidden patterns and insights from the provided data, it models itself. As a result, unsupervised learning is also described as a subset of machine learning in which models are taught with unlabeled datasets and then let to act on the data unchecked.[14] Finding the underlying structure of a dataset, classifying the data into groups based on similarities, and representing the dataset in a compressed format are the basic goals of unsupervised learning. To put it another way, it classifies or groups the unsorted dataset in accordance with the patterns, similarities, and differences.[15] As a result, the unsupervised learning algorithm is never trained on the input dataset that contains items of diverse categories, which means it has no knowledge of the dataset's characteristics. The unsupervised learning algorithm's objective is to automatically identify an item's features.[16] The unsupervised learning algorithm is anticipated to complete this work in the end by grouping the items dataset into groups based on their similarity.

In contrast to supervised learning approaches, unsupervised learning approaches do not require any sort of training procedure.[3] Unsupervised learning is when an algorithm investigates input data without being provided a specific output variable, for instance by looking for patterns in customer demographic data.[17] Unsupervised learning is typically employed when it is unclear how to categorize the data and when an algorithm is required to identify patterns and categorize the data. Again, unlike supervised learning, where we have the input data but no corresponding output data, unsupervised learning cannot be applied immediately to a regression or classification task. Compared to supervised learning, unsupervised learning algorithms enable more complicated tasks. Unsupervised learning strategies are more complex than supervised learning strategies, nevertheless. Unsupervised learning is depicted in Figure 4.

There are two different kinds of unsupervised learning methods: clustering and association. Web usage mining, continuous manufacturing market basket analysis, network analysis, recommendation systems, anomaly detection, and singular value decomposition are only a few examples of unsupervised learning applications.[18] Detail on the semi-supervised learning strategy is covered in the section that follows.

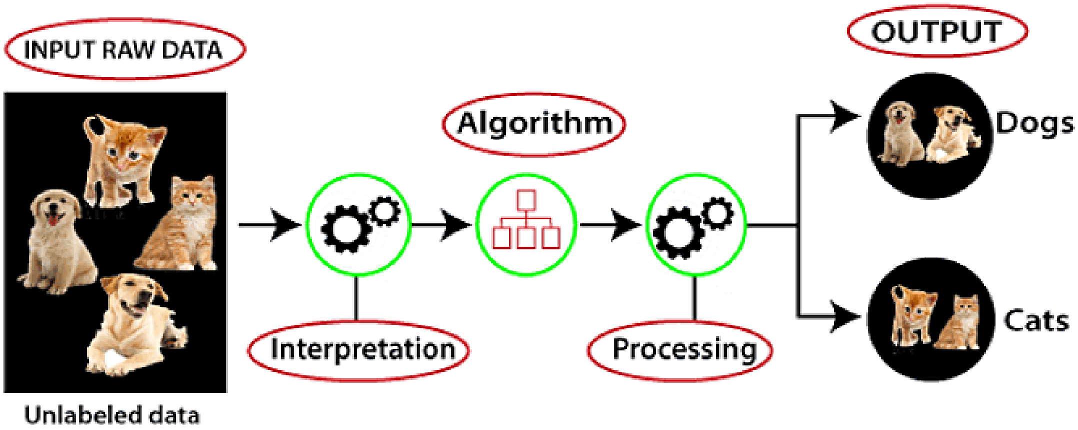


Fig. 4: Unsupervised learning process

Semi-Supervised Learning Approach

Figure 5 illustrates how semi-supervised learning techniques mix labeled and unlabeled instances to produce an appropriate function or classifier.[19] In other words, semi-supervised machine learning is a sort of algorithm that falls in between supervised and unsupervised machine learning. In order to create the desired output, it therefore combines supervised and unsupervised learning methods and uses a combination of labeled and unlabeled datasets during the training phase. Although semi-supervised learning combines supervised and unsupervised learning and uses data with a few labels, the majority of the data it uses is unlabeled. As supervised and unsupervised learning are dependent on the presence or lack of labels, it cannot be compared to those two types of learning. To reduce the drawbacks of supervised learning and unsupervised learning methods, use semi-supervised learning. Instead of using solely labeled data as in supervised learning, the primary goal of semi-supervised learning is to make effective use of all available data.[20] An unsupervised learning technique is used to cluster comparable data in the first place, and it also aids in labeling the unlabeled data into labeled data. The reason behind this is that labeled data is more expensive to acquire than unlabeled data.

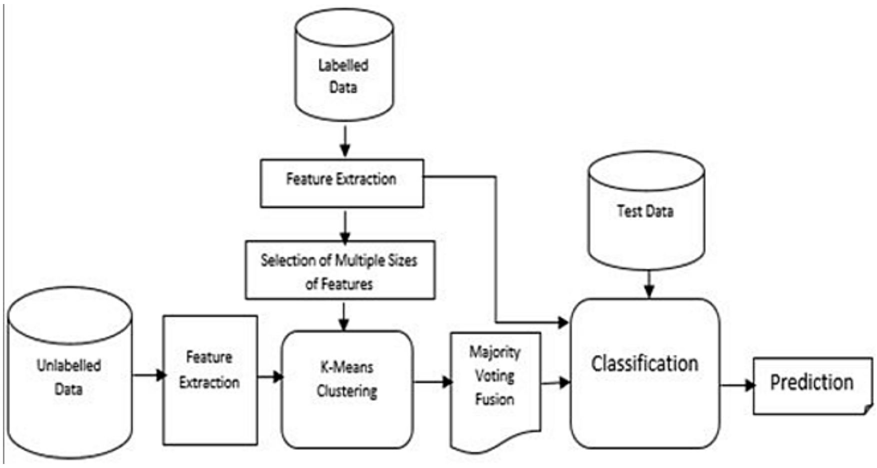
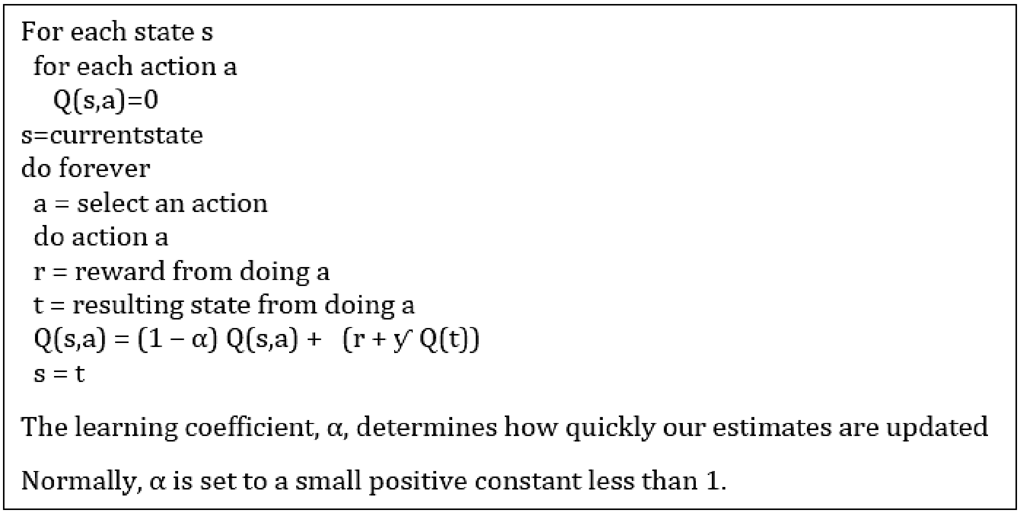


Fig. 5: Semi-supervised learning process

Reinforcement Learning, Deep Learning and Cybersecurity

Actions per state, agents, and states (S) are all components of reinforcement learning (A). Since RL algorithms frequently employ dynamic programming techniques, Markov decision processes frequently depict this reward-based world. These procedures show a basic explanation of the issue so that people can learn how to solve it. In reality, agents continually decide which course of action to take as the environment in which they behave reacts and confronts them with new challenges. RL algorithms don't know the precise Markov decision processes, in contrast to traditional dynamic programming techniques. The goal of the RL algorithm Q-Learning is to learn the policy that instructs agents on what action to take in specific circumstances.[2] This policy is optimal and provides all the subsequent steps required to accomplish a task while maximizing the incentive gain. Agents that study their environment are constantly forced to decide between using the knowledge they have learnt and investigating new potential actions to take. Thus, the e-greedy parameter—which denotes the ratio of exploration to exploitation actions—should be taken into account when developing RL algorithms.

The basic Q-learning algorithm is shown on Figure 6 below.

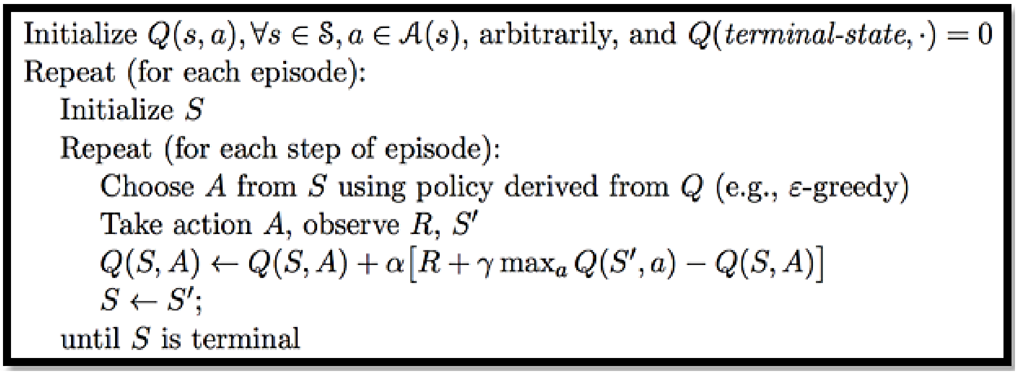


```
For each state s
 for each action a
   Q(s,a)=0
s=currentstate
do forever
 a = select an action
 do action a
 r = reward from doing a
 t = resulting state from doing a
 $Q(s,a) = (1 - \alpha)\ Q(s,a) + \ \ (r + \gamma\ Q(t))$
 s = t
```

The learning coefficient, α, determines how quickly our estimates are updated

Normally, α is set to a small positive constant less than 1.

Fig. 6: Basic Q-Learning Algorithm



```
Initialize $Q(s, a), \forall s \in \mathcal{S}, a \in \mathcal{A}(s)$, arbitrarily, and $Q(\textit{terminal-state}, \cdot) = 0$
Repeat (for each episode):
   Initialize $S$
   Repeat (for each step of episode):
      Choose $A$ from $S$ using policy derived from $Q$ (e.g., $\varepsilon$-greedy)
      Take action $A$, observe $R$, $S'$
      $Q(S, A) \leftarrow Q(S, A) + \alpha\big[R + \gamma \max_a Q(S', a) - Q(S, A)\big]$
      $S \leftarrow S'$;
   until $S$ is terminal
```

Fig. 7: Asynchronous one-step Q-Learning Algorithm

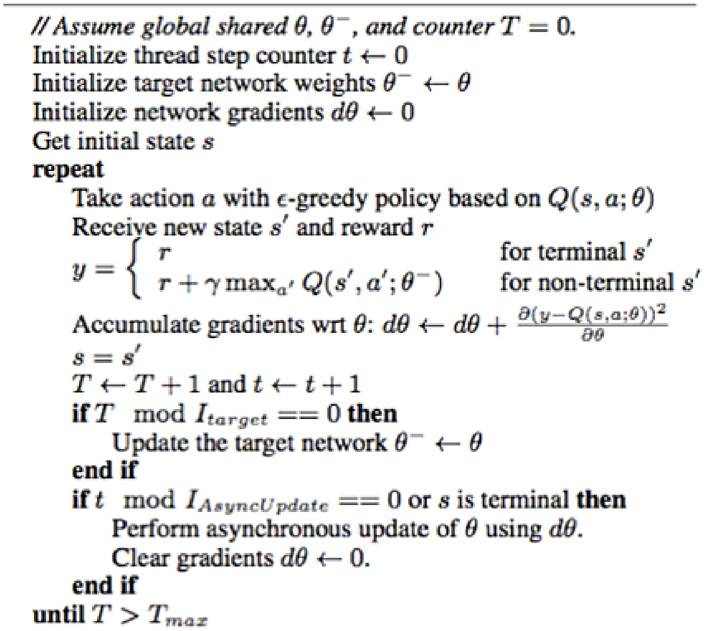
As an improvement to the basic Q-Learning Algorithm, the resultant Asynchronous one-step Q-Learning Algorithm is shown on Figure 7 below, where,

Q(st,at) ← Q(st, at) + α(rt+yQ(st+1, π(St+1))- Qst,at))
π(s) ← arg max Q(s,a)

Q(st,at) ← Q(st,at) + α (rt+ y max Q(st+1,a`) -Q(st,at))

For ease of implementation of the Asynchronous one-step Q-Learning Algorithm, the following steps are followed in order to produce the pseudo-code shown on Figure 8 below which was implemented using the Python programming language.

- Each thread interacts with its own copy of the environment and at each step, computes the gradient of the Q-learning loss.
- It uses a globally shared and slowly changing target network with parameters θ when computing the Q-learning loss.
- The gradients are accumulated over multiple time-steps before being applied (similar to using mini-batches), which reduces chance of multiple actor-learners overwriting each other's updates.



```
// Assume global shared $\theta$, $\theta^-$, and counter $T = 0$.
Initialize thread step counter $t \leftarrow 0$
Initialize target network weights $\theta^- \leftarrow \theta$
Initialize network gradients $d\theta \leftarrow 0$
Get initial state $s$
repeat
    Take action $a$ with $\epsilon$-greedy policy based on $Q(s, a; \theta)$
    Receive new state $s'$ and reward $r$
    $y = \begin{cases} r & \text{for terminal } s' \\ r + \gamma \max_{a'} Q(s', a'; \theta^-) & \text{for non-terminal } s' \end{cases}$
    Accumulate gradients wrt $\theta$: $d\theta \leftarrow d\theta + \frac{\partial (y - Q(s,a;\theta))^2}{\partial \theta}$
    $s = s'$
    $T \leftarrow T + 1$ and $t \leftarrow t + 1$
    if $T \mod I_{target} == 0$ then
        Update the target network $\theta^- \leftarrow \theta$
    end if
    if $t \mod I_{AsyncUpdate} == 0$ or $s$ is terminal then
        Perform asynchronous update of $\theta$ using $d\theta$.
        Clear gradients $d\theta \leftarrow 0$.
    end if
until $T > T_{max}$
```

Fig. 8: Asynchronous One-Step Q-Learning Algorithm pseudo-code

The Asynchronous Advantage Actor-Critic (A3C) Algorithm is much faster, simpler, and scores higher on Deep Reinforcement Learning tasks.[2] A3C completely blows most algorithms like Deep Q Networks (DQN). The A3C Algorithm is explained below.

Asynchronous

In contrast to DQN, which uses a single neural network to represent a single agent that interacts with a single environment, A3C makes use of various iterations of the aforementioned to learn more effectively. There are numerous worker agents in A3C, each with their own set of network parameters, and there is a global network. While the other agents are interacting with their environments, each of these agents is interacting with its individual copy of the environment. Beyond the expedited completion of more work, this is preferable to having a single agent because each agent's experience is distinct from that of the others. The range of experiences that are generally available for training increases in this way.

Actor-Critic

This series has so far concentrated on value-iteration techniques like Q-learning or policy-iteration techniques like Policy Gradient. Actor-Critic combines the advantages of the two methods. For A3C, our network will estimate a value function V(s) (how good it is to be in a specific state) as well as a policy (s) (a set of action probability outputs). These will all be distinct, fully-connected layers at the network's top. In comparison to conventional policy gradient approaches, the agent updates the policy (the actor) more intelligently by using the value estimate (the critic).

Advantage

The update rule informed the agent which of its acts were "good" and which were "bad" based on the discounted returns from a set of experiences. The network was then adjusted to suitably promote and discourage actions.

Advantage: A = Q(s,a) - V(s)

The structure of Reinforcement Learning (RL) with the A3C Algorithm is illustrated by the schema on Figure 9 below. A3C is faster, simpler, and scores higher on Deep Reinforcement Learning tasks. In terms of characteristics, the A3C Algorithm.

- Maintains a policy π(at| st; θ) and an estimate of the value function V(st; θv)
- Is an n-step Actor-Critic method
- As with value-based methods, it uses parallel actor-learners that accumulate updates to improve training stability
- In practice, the policy approximation and the value function share some parameters
- It uses a neural network that has one softmax output for the policy π(at| st; θ) and one linear output for the value function V(st; θv) with all non-output parameters shared
- For a one-step actor-critic method with a learned policy and a learned value function, the advantage is defined as follows,

A(at, st) = Q(at, st) – V(st)

- The advantage estimate for a n-step actor-critic method is shown below,

$$\sum_{i=0}^{k-1} \gamma^i r_{t+i} + \gamma^k V(s_{t+k};\theta_v) - V(s_t;\theta_v)$$

The A3C pseudo-code is shown on Figure 10 and was implemented in Python in this research paper.

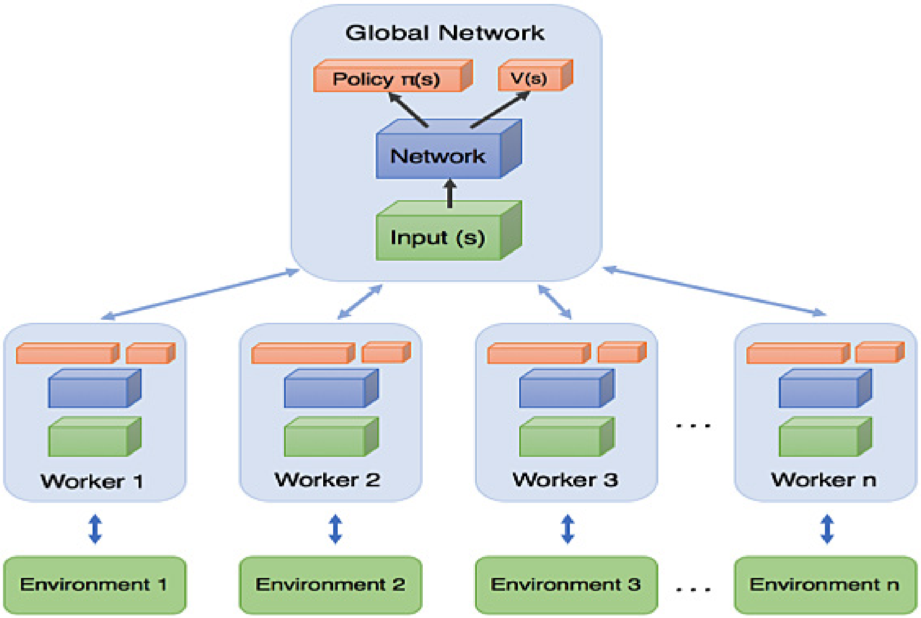
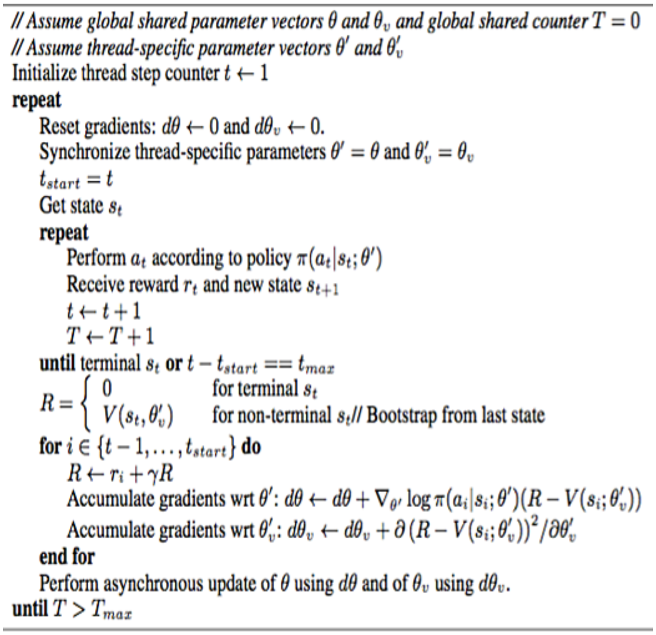


Fig. 9: Reinforcement Learning with the A3C Algorithm (Source:https://www.coursera.org/lecture/practical-rl/case-study-a3c-XNMfH)



```
// Assume global shared parameter vectors $\theta$ and $\theta_v$ and global shared counter $T = 0$
// Assume thread-specific parameter vectors $\theta'$ and $\theta'_v$
Initialize thread step counter $t \leftarrow 1$
repeat
    Reset gradients: $d\theta \leftarrow 0$ and $d\theta_v \leftarrow 0$.
    Synchronize thread-specific parameters $\theta' = \theta$ and $\theta'_v = \theta_v$
    $t_{start} = t$
    Get state $s_t$
    repeat
        Perform $a_t$ according to policy $\pi(a_t|s_t;\theta')$
        Receive reward $r_t$ and new state $s_{t+1}$
        $t \leftarrow t + 1$
        $T \leftarrow T + 1$
    until terminal $s_t$ or $t - t_{start} == t_{max}$
    $R = \begin{cases} 0 & \text{for terminal } s_t \\ V(s_t, \theta'_v) & \text{for non-terminal } s_t \text{// Bootstrap from last state} \end{cases}$
    for $i \in \{t-1, \ldots, t_{start}\}$ do
        $R \leftarrow r_i + \gamma R$
        Accumulate gradients wrt $\theta'$: $d\theta \leftarrow d\theta + \nabla_{\theta'} \log \pi(a_i|s_i;\theta')(R - V(s_i;\theta'_v))$
        Accumulate gradients wrt $\theta'_v$: $d\theta_v \leftarrow d\theta_v + \partial (R - V(s_i;\theta'_v))^2 / \partial \theta'_v$
    end for
    Perform asynchronous update of $\theta$ using $d\theta$ and of $\theta_v$ using $d\theta_v$.
until $T > T_{max}$
```

Fig. 10: Asynchronous Advantage Actor-Critic (A3C) Algorithm pseudo-code (Source: [1])

In terms of performance and training speed, the on-policy A3C algorithm looks to be the most effective asynchronous reinforcement learning algorithm.

Deep Learning (DL) is a recent area of study in machine learning. The creation of a neural network that mimics the human brain for analytical learning is the driving force behind it. It imitates the way the human brain interprets data like sights, sounds, and words.[21] DL is a machine learning technique based on data learning characterisation.[21] An observation, such as an image, can be expressed in many different forms, such as a vector representing each pixel's intensity value or, in a more abstract form, as a collection of edges, a region with a specific shape, or something similar. The learning of tasks from cases is facilitated by the use of specialized representations. DL approaches have supervised learning and unsupervised learning just like ML methods. Different learning frameworks have produced learning models that are very dissimilar.

The following are some of the distinctions between ML and DL

Data Dependencies

Deep learning and conventional machine learning differ mostly in how well they perform as the volume of data grows. Due to the fact that deep learning algorithms need a lot of data to fully understand the data, they do not perform as well when the data volumes are minimal. On the other hand, in this instance, the performance will be greater when the conventional machine-learning algorithm follows the specified principles.[21]

Hardware requirements

There are many matrix operations needed by the DL algorithm. The GPU is frequently utilized to efficiently optimize matrix operations. Therefore, the hardware required for the DL to function effectively is the GPU. Compared to conventional machine-learning algorithms, DL makes greater use of high-performance computers with GPUs.[21]

Feature Processing

Feature processing is the practice of incorporating subject expertise into a feature extractor to simplify the data and produce patterns that improve the performance of learning algorithms. The process of processing features takes time and requires expertise. In ML, the majority of an application's attributes need to be specified by a professional before being encoded as a data type. Pixel values, forms, textures, positions, and orientations are all examples of features. The precision of the characteristics extracted determines how well most ML systems function. The main distinction between DL and other machine-learning methods is the attempt to directly extract high-level characteristics from data.[21] As a result, DL lessens the effort required to create a feature extractor for each issue.

Problem-solving Method

Traditional machine learning algorithms typically divide an issue into several smaller problems, solve the smaller problems, and then combine the results to produce the desired outcome. Deep learning, on the other hand, supports direct, end-to-end issue solutions.

Execution Time

Because there are numerous parameters in a DL algorithm, it often takes a long time to train one; as a result, the training stage takes longer. The fastest DL algorithm, like ResNet, needs exactly two weeks to finish a training session, whereas ML training only takes a few seconds to a few hours. The exam time is the exact reverse, though. Running deep learning algorithms while testing takes relatively little time. As the amount of data rises, the test duration increases in comparison to some ML algorithms. Due to the short test duration of some ML algorithms, this argument does not hold true for all of them.

Interpretability

Importantly, when contrasting ML vs DL, interpretability is a crucial consideration. The performance of DL in recognizing handwritten numbers can be fairly impressive, coming close to meeting human standards. A DL algorithm won't, however, explain why it produced this outcome.[21]

Among all machine-learning algorithms, Support Vector Machine (SVM) is one of the most reliable and precise techniques. Support Vector Classification (SVC) and Support Vector Regression make up the majority of it (SVR). The idea of decision boundaries serves as the foundation for the SVC.[21] A set of instances with differing class values are divided into two groups by a decision boundary. Both binary and multi-class classifications are supported by the SVC. The separation hyperplane, which establishes the ideal separation hyperplane, is closest to the support vector. In the classification process, the places on the feature space's opposite side of the separation hyperplane and the mapping input vectors located there belong to different classes.

When there are data points that cannot be separated linearly, the SVM employs the proper kernel functions to map them onto higher dimensional spaces where they can be separated.[21]

Cyberthreats are still evolving quickly all around the world, posing a threat to the data security of systems, networks, software, and especially data. Because assaults are growing so quickly today, cybersecurity is a problem that affects us all and is increasingly important.[21] Cybercriminals target organizations that use networks to collect sensitive data, such as those in the medical, retail, and public sectors. They also target computers to steal personal information from users. Numerous cutting-edge techniques are utilized to create vulnerabilities, such as malicious URLs, CAPTCHA bypassing, email fraud and spam, network anomalies, malware, and impersonation assaults. Recently, it has become challenging to identify these assaults, even with the aid of security administrators and existing attack detection software. That is why artificial intelligence, specifically machine learning, is used in cybersecurity intelligence since it can manage vast volumes of data, evaluate it, and make conclusions more quickly than a human. Since machine learning algorithms are so effective at identifying new threats and predicting their arrival, a lot of research has been done using this combination of cybersecurity and machine learning.[21] Given the rise in attacks, cybersecurity is a major concern for any firm. Our existence is being threatened by an increase in cybersecurity attacks. Cyber analysts can receive recommendations and assistance from machine learning (ML) and artificial intelligence (AI) in identifying risks.

The need for updated research in the field of intrusion detection in computer networks is growing. Since the IPv6 protocol connects to the Internet of Things, there is a significant issue with the IP protocol's implementation in version 6 (IPv6) when it comes to network security and specifically in detecting breaches (IoT). As a result of this convergence of IPv6 and the IoT paradigm, a variety of devices—including blenders, microwaves, clothes, wearable technology, and cognitive buildings—can freely access the Internet. This poses a challenge to network security, making the search for IoT-specific intrusion detection techniques crucial. Detecting these attacks is difficult due to the rise in cyberattacks and cybercrime against cyber-physical systems (CPSs). Due to opportunities provided by machine learning (ML), in particular deep learning, it may be the worst of times, but it may also be the best of times (DL). Because of its layered environment and efficient technique for extracting relevant information from training data, DL generally outperforms ML in performance. DL models are swiftly adapted to CPS system cyberattacks.[22]

Systems that are connected to the Internet have grown significantly in size, making them more vulnerable than ever to cyberattacks. Cyberattack complexity and dynamics necessitate responsive, adaptable, and scaleable defense measures. Methods based on machine learning, most notably deep reinforcement learning (DRL), have been widely suggested to deal with these problems.[23] DRL is incredibly good at handling complicated, dynamic, and especially high-dimensional cyber protection challenges since it combines deep learning with classical RL.[23]

Therefore, machine learning-based cybersecurity systems rarely need updates and maintenance since they can stay current by learning from the acquired data.[24] However, because these algorithms rely so much on data, there may be a problem with bias in their decisions.[24] Good-quality data must therefore be provided into these systems in order for the machine learning techniques to function correctly and deliver high cybersecurity. Deep learning is a way of learning that goes into great detail and makes sure that only pertinent information is gathered and used for system development and improvement. This method of learning is made possible by comparing the new data with the old data in order to find patterns and confirm the accuracy of the data.[24] One of the most crucial aspects of the deep learning approach that aids in effective learning is network layers. However, one of the main drawbacks of these approaches is that learning takes place after a number of time-consuming steps. Additionally, these methods require more care and knowledge from the developers because they are more complex than shallow machine learning methods. However, once they are operational, deep learning technologies lessen the need for maintenance because they offer an automatic learning process that requires little human input.[24] As a result, deep learning techniques are also used to improve cybersecurity systems since they can respond quickly and effectively to online threats.

Based on their areas of expertise, the following categories of cybersecurity can be identified.[24]

Network security

These cybersecurity measures protect an organization's networks and communications from outsiders.

Application Security

With the help of this cybersecurity protocol, software and apps are protected against malicious behavior that could result in data loss.

End-user Security

This technique keeps system users informed so they may operate safely and stay safe from online risks.

Operational Security

Managing and moving a lot of data is a typical requirement for business operations. Cybersecurity ensures its safety, which is crucial to the continued operation of organizations.

Informational Security

This kind of protection prevents hackers and intruders from accessing databases for their own or other people's financial or personal advantage.

Disaster Recovery and Business Continuity

This cybersecurity practice was developed in response to a cybersecurity incident that may have resulted in data loss. Therefore, these techniques guarantee that all lost data is retrieved and that business operations return to normal.

Cybersecurity Education

A society that is trained in cybersecurity should serve as the foundation of a strong digital ecosystem.[32] As a result, the government should be in charge of any country's dream of cybersecurity education. The right plans and tactics should be implemented along with the required resources and support in order to fulfill the dream. The path to realizing the dream should also be tracked appropriately.[33] presented a framework, which is depicted in Figure 11 below, that helps to solve cybersecurity education concerns from five aspects, including strategic, tactical, preparation, delivery, and monitoring.

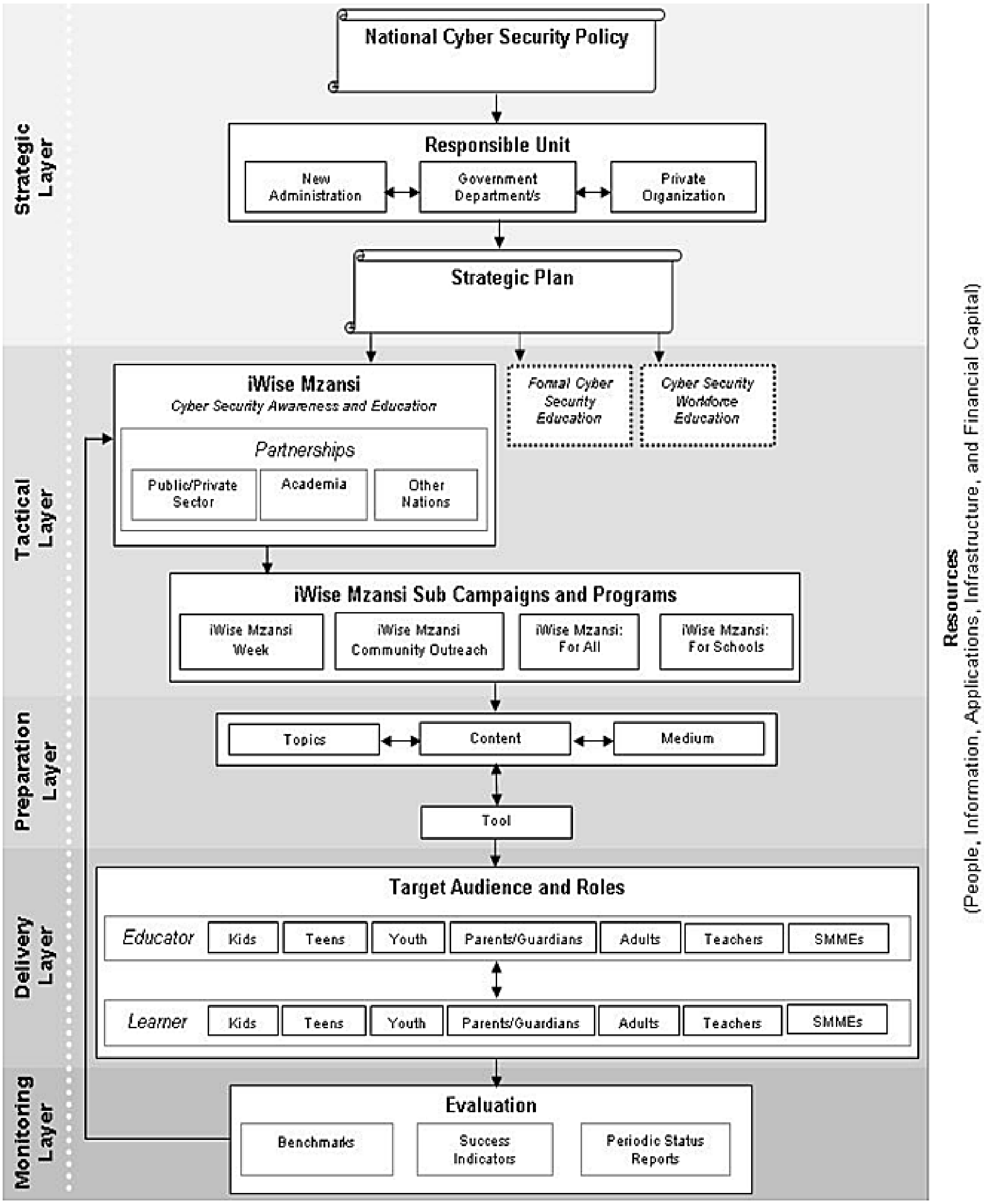


Fig. 11: Cybersecurity education framework (Source: [33])

Strategic approach to cybersecurity education

Cybersecurity is a problem that affects the entire country and should be handled at the national level. A cybersecurity strategy and strategic plan should be implemented, outlining the government's entire vision for cybersecurity education. Identifying and equipping accountable government agencies to carry out the mandate for cybersecurity education is necessary.[33] These agencies will be tasked with managing cybersecurity awareness and education operations across a range of citizen groups.

In order to improve cybersecurity instruction and curriculum, policy reviews at higher education institutions are also essential. This will motivate students to enroll in security courses and encourage financial support for cybersecurity research. Policies must be put in place to encourage and promote students' early preparation for and interest in STEM courses at the primary and secondary school levels.[34] However, it is crucial to make sure that the developed policies and strategies are put into action because nothing will happen without it.

Tactical Approach to Cybersecurity Education

Goals for cybersecurity education should be realized using certain strategies, which are crucial. There should be national campaigns to raise awareness and educate the public about cybersecurity..[35] asserts that community-based effort is necessary for the sustainable development of cybersecurity education and that university-led efforts alone will not be sufficient. Partnerships with important parties, including academics, the public and business sectors, as well as other countries, are crucial as a result. They are essential in laying the groundwork for the participants to support a country's efforts to raise cybersecurity awareness and educate its citizens. Primary and secondary school students should be the target audience for targeted cybersecurity education and awareness efforts. Cybersecurity should be taught to students in a way that is appropriate for their age group as part of the school curriculum and should be integrated into the formal education system.[33] To effectively spread cybersecurity knowledge throughout the society, schools must transform into knowledge hubs. Together, school administrators and teachers can plan cybersecurity-related school projects or programs.[36] [37]claims that student councils or clubs can be established so that everyone in the neighborhood can learn about cybersecurity in a more informal setting. On the other hand, teacher training programs should equip teachers with the skills necessary to teach students about cybersecurity and cyber ethics in a variety of levels. In order to spark interest and attract students to the sector, it is also crucial that secondary education students learn about computer science and cybersecurity careers. Therefore, it is important to strengthen educational efforts and discussions with professionals in the cybersecurity business in order to dispel misconceptions about computer science studies. In order to expand the pool of future cybersecurity specialists, it is crucial to aggressively encourage girls and women to engage in all of these cybersecurity education initiatives.[34]

Preparatory Approach to Cybersecurity Education

The preparation layer has four sections, according to.[33] Topics, content, medium, and tools are these. It is important to decide which cybersecurity themes will be covered. These may include, among many others, email security, identity theft, protecting sensitive information online, tailgating, social engineering attacks, and mobile device security. In order to give the pertinent cybersecurity education, the right content for the appropriate audience should subsequently be prepared based on the specified subjects. The appropriate communication method, which may be paper-based or electronic, will be selected. Games, films, and webpages are some of the tools employed.

According to,[38] scholarly articles or a search in the area of cyber security might turn up several games in the field that cater to various end user groups, including professionals and non-professionals. They instruct end users on how to spot events and what steps to take in the event of a cybersecurity concern. The methods used to offer cybersecurity education content in formal education are crucial to achieving the objectives. To boost understanding of cybersecurity topics at the elementary school level, video cartoons and analogies based on pre-existing mental models of the physical world may be used in some circumstances.[36] In the same vein, it's critical to maintain end users' motivation and engagement through the use of gamification techniques that employ rewards and positive reinforcement in order to increase their level of interest and involvement in cybersecurity education and increase the program's effectiveness.[38]

Cybersecurity Education Delivery

The intended audience for a certain endeavor, like a cybersecurity education campaign, is specified by the delivery layer. Define different target audiences, such as teenagers, parents, teachers, or small to medium businesses (SMEs). Since cybersecurity is a collective responsibility of everyone who interacts with cyberspace and derives value and benefits from it, the audiences are expected to play the roles of a learner who becomes educated on cybersecurity issues and that of an educator who imparts cybersecurity knowledge to others.[33] More importantly, to guarantee that cybersecurity education is effective, reinforcing of cybersecurity concepts or lessons learned should be done on a regular basis by going over previous material.[38]

Monitoring Cybersecurity Education

The layer examines the scheme's progress in realizing the government's vision. It is important to track and assess the development of cybersecurity education programs to determine their efficacy. In that regard, criteria must be specified, as well as success indicators. The tactical and preparation layers' operations should be informed by evaluation feedback in order to continuously develop cybersecurity initiatives and fulfill the goal of cybersecurity education.[33] Tracking and reporting the progress accomplished should be done consistently.[38]

Resources

The value chain for cybersecurity education has resources as a critical enabler.[39] Any cybersecurity education initiative's success or failure is determined by the availability of resources. The framework's many layers call for various resources. These could include the knowledge and human resources needed to carry out specific tasks, as well as the ICT hardware and software. To deploy cybersecurity laboratories and establish cybersecurity research and development centers, financial resources might be needed. Additionally, in order to accomplish the aims of cybersecurity education, private funding, scholarships for cybersecurity education, and grants for cybersecurity research and development should be used.[34]

Cybersecurity Training

The exponential rise in cyberattacks has made it clear that society needs effective cybersecurity training in order to fully equip people for the ability to recognize, flag, prevent, and neutralize harmful efforts at any attack. A change in user behavior should be the end goal of effective cybersecurity training programs, which should keep participants interested and provide hands-on activities,[40,41] and.[42] A redesigned Analysis, Design, Development, Implementation, and Evaluation (ADDIE) paradigm built on fast prototyping was developed by[42] to guarantee effective cybersecurity training. Figure 12 below shows the ADDIE model.

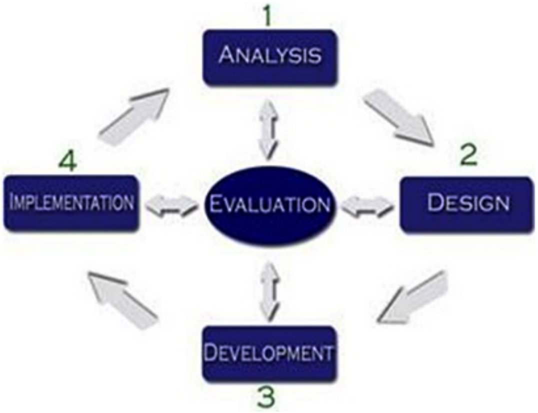


Fig. 12: A Revised Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model, based on rapid prototyping.

Analysis Stage

The analysis step looks at the requirements for cybersecurity training. Clarification of the training's objectives and results is done while keeping in mind the target audience, the mode of instruction, and the subject matter. The issue of the necessary training resources is also taken into account.[42]

Design Stage

The development of a cybersecurity training program should be done in accordance with the objectives and specifications of the program as well as the findings of the analysis phase. The audience is taken into account when choosing how cybersecurity training is delivered.[42]

Development Stage

At this step, the produced training components are validated using internal reviews or test cases.[42]

Implementation Stage

The setting for cybersecurity training is currently set up, and training exercises are scheduled. Training is started after engaging the necessary participants.[42]

Evaluation Stage

As a technique of getting feedback and improving training efficacy, evaluation is done throughout the training lifecycle to look at user happiness, the caliber of the training resources, and whether any corrective actions are required.[42]

Numerous cybersecurity training methods are available, some of which are listed in Table 1 below, according to.[42]

The target audience, the availability of infrastructure and connectivity, interactivity, and the availability of funding are just a few of the numerous elements that should be taken into account while using cybersecurity training approaches.

Table 1: Cybersecurity Training Methods

Cybersecurity Training Method	Examples
a) Conventional means	• Training sessions on-site • Training in the classroom • Presentations , • Conferences, • Symposiums
b) Online/Virtual and Software-based	• Online/Virtual courses and classes; • Cloud-based training; • Web-accessible training material and software;
c) Game-based (Online/Offline)	• Games for cybersecurity training delivered online or offline as downloadable applications
d) Video-based	• Educational videos
e) Simulation and virtualization based techniques	• Testbeds, • Simulation platforms, • Simulated Laboratory exercises

Research Methodology

25 defines research methodology as a framework for conducting research. A research methodology is also defined by[26] as a description of the study design, data collection, methodologies, sample techniques, fieldwork processes, and data analysis contained inside the body of a research report. The paper's goal was to determine whether the Reinforcement Learning (RL) paradigm is appropriate for cybersecurity training and instruction. This research adopted a pragmatism research philosophy that accommodate qualitative methods and quantitative approaches. In a pragmatist research philosophy, there is a combination of the use of facts and figures as well as subjective elements such as freely expressed opinions and insights. The study's methodology was mainly positivism, which emphasizes quantitative methods for identifying the RL paradigm for cybersecurity teaching and training. Ontology is a belief system that represents a person's understanding of the nature of a reality, according to.[26] The study took positivist views of reality into account. The philosophy of knowledge known as epistemology is concerned with the reliability, modes of attaining knowledge, and depth of knowledge.[25] As a result, although offering improved insights into realism, the research philosophy employed might not be ontologically dogmatic.[25] According to positivists, the task of the researcher is to seek and highlight the truth, which exists and can only be learned via thorough investigation. Since interpretivists prefer humanistic qualitative methods, the positivist paradigm largely used in this work emphasizes

scientific quantitative methods instead. Due to the investigation's nature, a quantitative research design (positivism) was deemed appropriate for the RL scientific issues. In order to maximize the research's depth and breadth of knowledge regarding how to use RL in cybersecurity education and training, the study employed a positivist approach strategy. The research design was an experiment with a primary goal of using Python to build the RL Q-Learning and A3C algorithms. However, a survey research design was used on cybersecurity education and training as exemplified through Zimbabwean commercial banks.

A research method known as descriptive research aims to characterize the characteristics of the population being examined or the phenomenon being investigated.[27] Researchers can communicate or present a picture of the phenomenon they are studying by using descriptive designs.[27] When measuring and testing several samples, which are needed for more quantitative experiments, is not credible, descriptive research is the best option. Although descriptive research results cannot be used to definitively prove or disprove a hypothesis, they can be a useful tool in a variety of scientific fields if the limitations are recognized. Exploratory research is the process of analyzing an issue that has not before been fully researched or analyzed.[28] Exploratory research is typically carried out to get a deeper understanding of a current issue, although it typically yields inconclusive findings. In order to become familiar with an existing occurrence and gain fresh insight to define a more specific problem, researchers employ exploratory research.[29] The research is started with a broad concept, and the findings are then used to discover problems that are connected to the research topic. When there is little information available, explanation research focuses on why something happens. It enables the researcher to gain a deeper grasp of each research topic, determine how or why a particular phenomenon occurs, and predict upcoming occurrences.[31] Investigating the causes or mechanisms of a phenomenon can benefit from explanation research. Explanatory research is perfect for analyzing trends and developing ideas that serve as a roadmap for additional investigation.[30] An in-depth knowledge of an association between variables is produced via explanatory research. The study is exploratory, descriptive, and explanatory in nature. The abductive method was used in this study because pragmatism is frequently associated with abductive reasoning. Due to its reasoning that alternates between induction and deduction, this strategy was the best. The survey design was employed by the researcher with regards to cybersecurity education and training in commercial banks because it provides the answers to the research's foundational questions, closes the knowledge gap in the area, and creates benchmarks for future evaluations. The survey design was chosen because it offers a highly cost-effective means to collect copious amounts of data to answer the questions of who, what, where, when, and the methodology used for the study.

A survey was conducted on the cybersecurity education and training as exemplified by Zimbabwean commercial banks. The research also sought to investigate the adoption of ML in combating cyber security threats amongst banks in Zimbabwe to demonstrate the importance of cybersecurity education and training. The researcher chose the mixed methods approach because it is likely to yield insightful results into the subject being investigated that cannot be fully comprehended by relying simply on qualitative or quantitative procedures. Combining two approaches could be preferable to utilizing just one. Both closed and open questions were included in the questionnaire's question format. Closed questions allowed respondents to select from a set of possibilities the one that best expressed their viewpoint. Open questions required respondents to provide a response in order to respond. The researcher attached the survey as a Google forms document to the email. The primary benefit of using emails was that the responses were rapid. The primary drawback of doing surveys by email is that very few people answer to it because they rarely check their email accounts. To analyze the qualitative data acquired throughout the investigation, content analysis was performed. These questionnaire questions in this study were open-ended. The study population encompassed employees and customers from five commercial banks (Stanbic Bank, ZB Bank, CBZ, BancABC and Nedbank Zimbabwe). Employees and customers had valid information with respect to digital transformation processes and cybercriminal activities in commercial banks from their experience and from following trends in the commercial banking sector. The breakdown of the study population is presented in Table 2.

Table 2: Target Population

Description	Population Size
IT and Digital banking employees	1000
Risk and Compliance	2100
Information Systems and Business development	2800
Retail banking	2600
High valued customers	3000
Total	11,500

Source: Mid-Year Zimbabwean Commercial Banking Sector Survey, 2022

Table 2 shows that the target population of this study was 11,500, which was the total number of banking employees and customers in the selected commercial banks in Harare as at 30 September 2022. Based on the Krejcie and Morgan (1970) formula, the sample size for this study was 370. Table 3 below shows the breakdown of the sample size for this study.

Table 3: Sample Size

Population	Size	Instrument
IT and Digital banking employees	32	Questionnaire
Risk and Compliance	68	Questionnaire
Information Systems & Business development	90	Questionnaire
Retail banking	84	Questionnaire
High valued customers	97	Interview
Total	370	

Table 3 shows the breakdown of the sample size for this study. The sample size was 370 employees and customers in commercial banks. To have a fair representation of the research participants from the five banks, the researcher determined the number of the participants to make up the sample by multiplying the proportion of each category with the sample size. This research employed sampling techniques of stratified and simple random. This research used a structured questionnaire and unstructured interviews for bank employees and customers respectively. This research used self-administered structured research questionnaires to collect data from the employees in banks. The questionnaire used in this research was made up of closed questions to gather quantitative data. A Likert scale was applied for statistical data analysis. The structured questions have end points of “Strongly Disagree” and “Strongly Agree”. The study also used the interview guide as a qualitative data collection instrument on high valued customers. In this research, the researcher interviewed high valued customers face to face when they visit the branch and a few were arranged through Zoom Meeting platform.

Machine Learning (ML) is primarily split into four groups based on the learning procedures, including supervised, unsupervised, semi-supervised, and reinforcement learning approaches. Actions per state, agents, and states (S) are all components of reinforcement learning (A). The goal of the RL algorithm Q-Learning is to learn the policy that instructs agents on what action to take in specific circumstances.[2] The Q-learning algorithm was implemented using the Python programming language, where each thread interacts with its own copy of the environment and at each step, computes the gradient of the Q-learning loss. The Asynchronous Advantage Actor-Critic (A3C) Algorithm is much faster, simpler, and scores higher on Deep Reinforcement Learning tasks.[2] While reinforcement learning (RL), a goal-oriented machine learning (ML) technique, finds use in routine real-world tasks, supervised and unsupervised machine learning are currently far more widely used in organizations. RL's target industries include robots, dialogue systems, autonomous driving, personalisation, industrial automation, preventive maintenance, and medicine.

The Q-Learning Algorithm implementation done in Python is shown on Figure 11.

The A3C Algorithm was implemented in Python using four agents. The visualization of the A3C Algorithm implementation is shown on Figure 12 below.

In carrying out this research, the researcher adhered to the ethical principles of anonymity, confidentiality and empathy. The research adhered to the following ethical standards in conducting the study which allowed the study to have sound data collection.

- Respondents of the study were to participate involuntarily and were notified that there would be no remuneration for their time expended and services offered. They were further informed of their right to withdraw from the study at any given stage if they so wished to do so. This was done in an attempt of ascertain that the data collected was of free will and individuals were not compelled into responding to the questionnaire and participating in interviews.
- Respondents participated on the basis of informed consent, participated freely and were very much aware of the aims of the study and the implications of participating. They were not coerced in any way to participate in the study. The researchers ensured the privacy and anonymity of respondents was respected and maintained by issuing questionnaires and interviews with numbers for identification rather than using names and or other details of respondents.
- The ethical consideration for the study ensured adequate level of confidentiality of the research data to be ensured, anonymity of individuals participating in the research, honest communication, transparency and avoidance of misleading information.

Key Findings, Data Analysis and Interpretation

Differentiation between Reinforcement Learning (RL) and other Machine Learning (ML) taxonomies (supervised, semi-supervised and unsupervised).

Supervised machine learning is a type of machine learning that requires humans to give input and required output. Algorithms are not supervised using training datasets when utilizing the machine learning technique known as unsupervised learning. The common uses of supervised learning include speech recognition, fraud detection, medical diagnosis and image segmentation. Training and testing are the two crucial phases of the supervised learning process. In the training stage, training data is taken into consideration as the input, and the learning algorithm studies the structures to build the learning model.

Unsupervised learning is a subset of machine learning in which models are taught with unlabeled datasets and then let to act on the data unchecked. It aims to find the underlying structure of a dataset, classifying the data into groups based on similarities, and representing the dataset in a compressed format.

- Models are not supervised using training datasets when utilizing the machine learning technique known as unsupervised learning.
- As a result, unsupervised learning is also described as a subset of machine learning in which models are taught with unlabeled datasets and then let to act on the data unchecked.
- In contrast to supervised learning approaches, unsupervised learning approaches do not require any sort of training procedure.
- Again, unlike supervised learning, where we have the input data but no corresponding output data, unsupervised learning cannot be applied immediately to a regression or classification task.
- Compared to supervised learning, unsupervised learning algorithms enable more complicated tasks.

Semi-supervised learning aims to make effective use of all available data. Instead of using solely labeled data as in supervised learning, it uses a combination of labeled and unlabeled datasets during the training phase. Although semi-supervised learning combines supervised and unsupervised learning and uses data with a few labels, the majority of the data it uses is unlabeled. To reduce the drawbacks of supervised learning and unsupervised learning methods, one is advised to use semi-supervised learning.

Actions per state, agents, and states are all components of Reinforcement Learning (RL). RL algorithms don't know the precise Markov decision processes, in contrast to traditional dynamic programming techniques. The goal of the RL algorithm Q-Learning is to learn the policy that instructs agents on what action to take in specific circumstances. This policy is optimal and provides all the steps required to accomplish a task while maximizing the incentive gain. The Asynchronous Advantage Actor-Critic (A3C) Algorithm is much faster, simpler, and scores higher on Deep Reinforcement Learning tasks. In contrast to DQN, which uses a single neural network to represent a single agent that interacts with a single environment, A3C makes use of various iterations of the aforementioned to learn more effectively. In terms of performance and training speed, the on-policy A3C algorithm looks to be the most effective asynchronous reinforcement learning algorithm.

The impact of ML on cybersecurity training and education.

Cybersecurity is a problem that affects the entire country and should be handled at the national level. Identifying and equipping accountable government agencies to carry out the mandate for cybersecurity education is necessary. Policies must be put in place to encourage and promote students' early preparation for and interest in STEM courses. Goals for cybersecurity education should be realized using certain strategies, which are crucial. There should be national campaigns to raise awareness and educate the public about cybersecurity. Cybersecurity should be taught to students in a way that is appropriate for their age group as part of the school curriculum. It is important to decide which cybersecurity themes will be covered. These may include, among many others, email security, identity theft, protecting sensitive information online, tailgating, social engineering attacks, and mobile device security. The appropriate communication method, which may be paper-based or electronic, will be selected. The value chain for cybersecurity education has resources as a critical enabler. These could include the knowledge and human resources needed to carry out specific tasks, as well as ICT hardware and software. Any cybersecurity education initiative's success or failure is determined by the availability of resources. The exponential rise in cyberattacks has made it clear that society needs effective cybersecurity training. A redesigned Analysis, Design, Development, Implementation, and Evaluation (ADDIE) paradigm built on fast prototyping was developed by.[42]

Examination of ML use in cybersecurity education and training in commercial banks in Zimbabwe. Types of cybersecurity threats affecting banks in Zimbabwe.

The lack of a cybersecurity culture among Zimbabwean enterprises, including banks, has been shown by,[2] leaving them vulnerable to assaults like phishing, hacking, harmful software, identity theft, and card fraud. This goal is to comprehend whether these security threats exist from the banks' point of view. The information in Table 4 demonstrates that all of these cybersecurity threats had an impact on Zimbabwean banks.

Table 4: Cybersecurity threats in Zimbabwean banks

Cybersecurity threat	%
Phishing	77%
Hacking	69%
Malicious software	100%
Identity theft	86%
Card fraud	100%

Source: Survey data (2022)

According to the statistics gathered, phishing poses a security risk to 47% of banks. They also discovered that 69% of banks in Zimbabwe had already been compromised. The ability of financial organizations to steal consumers' sensitive information and use it for evil purposes has made them one of the top targets for hackers. Additionally, banks acknowledge being impacted by cybersecurity dangers like rogue malware. This is demonstrated by the fact that all banks rated this threat as being prominent in their organizations. These results are in line with the theory that banks, by virtue of their status as financial organizations, are susceptible to Trojans and other malware that passes for genuine software. The fact that 100% of the banks selected card fraud as a concern to deal with further proves that all the institutions acknowledged having been exposed to this cyber security threat. The fact that 86% of respondents said that identity theft is a threat to banks in Zimbabwe shows that it is. This is in line with research from other countries that identifies identity theft as the root cause of all other cybercrime in financial institutions. Considering that the principal intent of the offender is to commit fraud by obtaining personal information from the victim. Because of this, some banks may have been victims of other cybercrime, but they might have mistaken it for identity theft.

Role of ML in combating cybersecurity threats in Zimbabwean banks.

According to,[1] traditional cyber security methods are ineffective in the face of the exponentially increasing quantity of cyberthreats; as a result, ML has been included into cybersecurity. The part aimed to comprehend the position of machine learning (ML) in thwarting cybersecurity threats in Zimbabwe, as well as system issues and other difficulties that have prevented banks from properly utilizing the advantages of this facility. The effectiveness of ML in preventing cyberattacks from a risk management perspective is displayed in Table 5 below.

Table 5: Effectiveness of ML (in risk management systems) in combating cybersecurity threats

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	7	43.8	43.8	43.8
	No	9	56.2	56.2	100.0
	Total	16	100.0	100.0	

Source: Survey data (2022)

Data gathered revealed that while the majority of banks (56.2%) were not satisfied that these technologies can prevent cyberattacks, 43.8% of banks were convinced that the use of ML in risk management helps to prevent cyberattacks. Some banks' lack of faith in ML may be due to a heavy dependence on conventional cybersecurity. This might also be the result of banks not properly implementing ML systems because of a lack of funding. The majority of systems, according to banks, are complicated and necessitate frequent changes, which cost money.

Table 6 presents the extent in which ML has helped reduce the cybersecurity threats and in using the mean score to interpret the Likert scale used, the researcher used the scale below to interpret.

1. From 1 to 1.80 represents (not at all).
2. From 1.81 to 2.60 represents (not much).
3. From 2.61 to 3.40 represents (Neutral).
4. From 3:41 to 4:20 represents (Somewhat combated).
5. From 4:21 to 5:00 represents (Very much combated)

Table 6: ML in combating cybersecurity threats in banks

	N	Mean
Phishing	16	4.31
Hacking	16	5.00
Malicious software	16	3.63
Identity theft	16	3.49
Card fraud	16	2.00
Valid N (listwise)	16	

Source: Survey data (2022)

According to the survey's findings, ML has been implemented into risk management processes, and card fraud and other cybersecurity concerns like phishing, hacking, harmful software, and identity theft are common in Zimbabwean institutions. The findings from banks regarding the application of ML to counteract these dangers are shown in Table 6. The results demonstrated that hacking was one of the cyber threats that had been extremely successfully fought, as indicated by a mean score of 5, and phishing was another threat that had been successfully fought. There were still moments when AI and ML failed to protect against malicious software (Mean score-3.63) and identity theft (Mean score-3.49) in some banks. Card fraud had a mean score of 2, indicating that the majority of Zimbabwean banks have had little success battling this cyber threat. These findings demonstrate that card fraud, malicious software, and identity theft continue to be the most common bank threats, even in the presence of ML. The effectiveness of ML in underdeveloped countries has been cited as being specifically impacted by three issues: a lack of awareness, weak rules, and implementation costs. Data hurdles, the black box aspect of the models, validation challenges, model testing and outcomes analysis challenges, as well as issues with models created by suppliers, were additional system obstacles for banks. However, as shown on Table 7, lack of awareness has had an effect on the implementation of ML in combating cyber threats.

Table 7: Lack of awareness has had an effect on the implementation of ML as cybersecurity threat managers.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	True	11	68.8	68.8	68.8
	False	5	31.3	31.3	100.0
	Total	16	100.0	100.0	

Source: Survey data (2022)

Banks were questioned about whether a lack of awareness affected the application of ML to counter cyber security threats. According to the data collected, 68.8% of the banks concurred that the use of ML to tackle cyber security concerns was hindered by a lack of awareness.[2] concurred and argued that due to a lack of personnel with the necessary training and experience, training facilities, suppliers of cyber assessment and penetration analysis, cybersecurity technology framework, security culture and financial resources, and cybersecurity management and monitoring frequently do not exist in developing countries.

The findings also showed that all banks believed weak laws and regulations hindered the successful use of machine learning as a risk management tool. In Table 8, this is displayed. Moving a lot of consumers to mobile banking has exposed banks to threats they were ill-equipped to handle due to changes in the nation's financial legislation. In poor countries, it is typical for the majority of governments to lack effective cybercrime legislation.

Table 8: Ineffective legislation and policies have affected the efficient implementation of ML.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	True	16	100.0	100.0	100.0

Source: Survey data (2022)

Table 9: The cost of implementing AI and ML have affected the proper use of the systems in combating cybersecurity threats.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	True	16	100.0	100.0	100.0

Source: Survey data (2022)

Banks acknowledged that the expenses associated with ML implementation have limited how well the systems can be used to tackle online threats. Table 9 demonstrates that all banks accepted the assertion. Savings and credit cooperatives, microfinance institutions, and financial cooperatives are the most susceptible to attacks in developing nations since their systems are underdeveloped and do not have enough safeguards and precautions.

Table 10: System challenges affecting banks in using ML to combat cybersecurity threats

	N	Mean
Financial data gathering challenges	16	2.31
The inability of the system to adequately explain the output	16	4.32
Validation challenges of the models	16	3.69
Model testing and outcome analysis problems	16	4.00
Problems with vendor-designed models	16	3.31
Valid N (listwise)	16	

Source: Survey data (2022)

Ineffective legislation, high expenses, and a lack of understanding are not the only difficulties banks have had to deal with when attempting to integrate ML into their risk management. The results showed that, as shown on Table 10, among the system challenges with the highest effect on banks' ability to effectively implement ML in defending against cyber security threats were: The system's inability to adequately explain the output, which had a mean score of 4.32 and denotes that it has a very high effect on how effectively banks use ML in risk management. This is in line with a literature review that concluded that appropriate logical justification of the output is required to inspire the decision-makers with sufficient confidence and that simple outputs are insufficient in many important financial applications. Model testing and results analysis yielded a mean score of 4.00, indicating that this issue has a significant impact on banks' capacity to apply machine learning to counter cybersecurity threats. It has been demonstrated in the past that banks lack the financial resources necessary to guarantee the effectiveness of such trainings. Validation of the models was another difficulty that had a significant impact on banks' use of ML to tackle cybersecurity risks. A mean score of 3.69 demonstrates that this challenge had a significant impact on banks. Due to their increased complexity and opaque nature in operation, ML models pose some significant obstacles to risk management and validation.

The effectiveness of models created by manufacturers in defending against cybersecurity threats seems to be moderate. This is demonstrated by a mean score of 3.31, which indicates that although this challenge did have an impact, it was only moderate and not as significant as the ones previously discussed. This also implies that, due to proprietary restrictions, testing the models provided by manufacturers can be challenging in many real-world situations. Banks also claimed that despite having measures in place, clients continue to be impacted because most customers are frequently vulnerable to assaults due to ignorance of security threats. The collection of financial data did not appear to have much of an impact on banks, as indicated by their mean score of 2.31. This indicates that ML's capacity to combat cybersecurity risks is unaffected by minor financial difficulties. Financial data is typically difficult to synthesize, which is a barrier to the effective application of ML to risk management.

Cybercriminal Activities in Commercial Banks in Zimbabwe

Descriptive statistics are presented on Table 11 to show the cybercriminal activities in commercial banks in Zimbabwe. The mean score reflects the position of respondents and the standard deviation shows the reliability of the mean score.

Table 11: Cybercriminal Activities in Commercial Banks in Zimbabwe

Cybercriminal activities	Mean	Std. deviation
The bank I work for has dealt with cases of card cloning where a customer lost their account balances to fraudsters.	3.55	1.28
The workers in the bank I work for have reported instances when they were tricked into handing over sensitive information.	3.06	1.55
The workers in the bank I work for have reported instances when they were tricked into installing malware.	3.54	1.45
The bank I work for has dealt with cases of hacking at work where a customer lost money in their account to fraudsters.	3.58	1.98
The workers of the bank I work for were once promised monetary benefit if they agree to transfer money using personal accounts.	4.13	1.21
The employees of the bank I work for have unlimited access to confidential and private customer information.	3.25	1.44

Source: Primary data (2022)

Table 12: Cybersecurity Strategies required during Digital Transformation

Cybersecurity strategies	Mean	Std. deviation
The bank I work needs to prepare for digital transformation processes.	3.94	1.00
The bank I work can reduce risk by conducting fraud and cybercriminal activities awareness	3.51	1.40
The bank I work needs effective anti-virus softwares for cybercrime free digital transformation processes.	3.54	1.45
The bank I work should develop a data breach response plan during digital transformation period.	3.58	1.98
The bank I work should engage a cybersecurity attorney during digital transformation period.	4.13	1.21
The bank I work should invest in a robust cyber insurance policy during digital transformation period.	3.25	1.44
The bank I work should foster a top-down cybersecure culture during digital transformation period.	3.32	1.51
The bank I work should treat data like revenue during digital transformation period.	3.55	1.28
The bank I work should make use team-driven approach during digital transformation period.	2.02	0.99
The bank I work should intensify Euro Mastercard Visa Technology use during digital transformation period.	3.95	1.00

Source: Primary data (2022)

Cybersecurity Strategies required for Digital Transformation Processes in Zimbabwe

Here, utilizing descriptive information from Table 12 that demonstrate the cybercriminal activities in commercial banks in Zimbabwe, we offer the cybersecurity methods needed for the digital transformation processes in Zimbabwe. The standard deviation demonstrates the reliability of the mean score whereas the mean score itself indicates the respondents' positions. The Kolmogorov-Smirnov (K-S) values also demonstrate the normalcy of the replies.

The findings of this study demonstrated that ML is utilized in the banking risk management procedures of Zimbabwean institutions. All banks utilize machine learning (ML) to manage credit risk, operational risk, liquidity risk, and reputational risk, according to the survey's findings. The results also showed that, despite the usage of ML in banks, some parts of risk management are ineffective. Since ML has proven to be an excellent tool for helping banks manage credit risk, it is safe to infer that they are equally adept at handling operational risk management. ML performed worse when it comes to managing liquidity and reputational risk.

Explore the types of cyber security threats affecting banks in Zimbabwe.

The study found that cyberthreats such card fraud, fishing, hacking, malware, and data theft have an impact on Zimbabwe's bank. Card fraud and malicious software are the most pervasive risks, and all banks are vulnerable to them.

Examine the role of artificial intelligence and machine learning in combating cybersecurity threats in Zimbabwean banks.

The findings indicated that banks did not believe machine learning was a viable defense against cybersecurity risks. The study's findings demonstrated that ML was utilized to battle phishing and hacking, but that it only partially countered the threats posed by dangerous software and identity theft. One of these was card fraud, which ML was unable to stop. The findings indicated that one of the causes of this inefficiency was a lack of knowledge in the application of ML. The results also demonstrated that the efficient deployment of ML in Zimbabwean banks has been hampered by poor legislation, policies, and implementation costs. The outcomes also demonstrated that there were systemic issues that hindered the effective application of ML in addressing these cybersecurity concerns.

Develop a framework specific to Zimbabwean banks for combating cybersecurity threats using ML.

The effective application of ML in defending against cybersecurity threats was greatly hampered by issues including the systems' failure to appropriately explain the output, validation of the model, model testing and outcome analysis, and vendor-designed models.

Education and awareness of Cybersecurity

Which of the following have raised your level of education and awareness of cybersecurity risks or issues in the organization?

The level of education and awareness of cybersecurity risks is illustrated on Figure 13.

The survey's findings showed that conducting cybersecurity training sessions and sending awareness messages to every employee were very successful in boosting the organization's level of cybersecurity knowledge and awareness. On the other hand, users learned the significance and relevance of data protection in the context of protecting data in the case of computer theft as a result of the adoption of data protection softwares like bitlocker on end user workstations. Conferences, seminars, and symposiums held as summer or winter schools were also mentioned as a way for people to learn about cybersecurity challenges, while some people learned about cybersecurity dangers or incidents as a result of some system outages brought on by ransomware.

Which of the following do you intend to implement in the short-term to maintain or improve your cyber security posture?

The proposed Cybserecurity Plans are shown on Figure 14. A crucial task to be completed in the near future to strengthen the organization's cybersecurity posture was cybersecurity training for both ICT and non-ICT workers. However, in order for the training to be successful, other measures including executive management support, the adoption of cybersecurity technologies, and the hiring of more cybersecurity personnel have to be undertaken. Plans to encourage personnel to practice good cyber hygiene and deter staff from engaging in destructive actions included reward or discipline schemes.

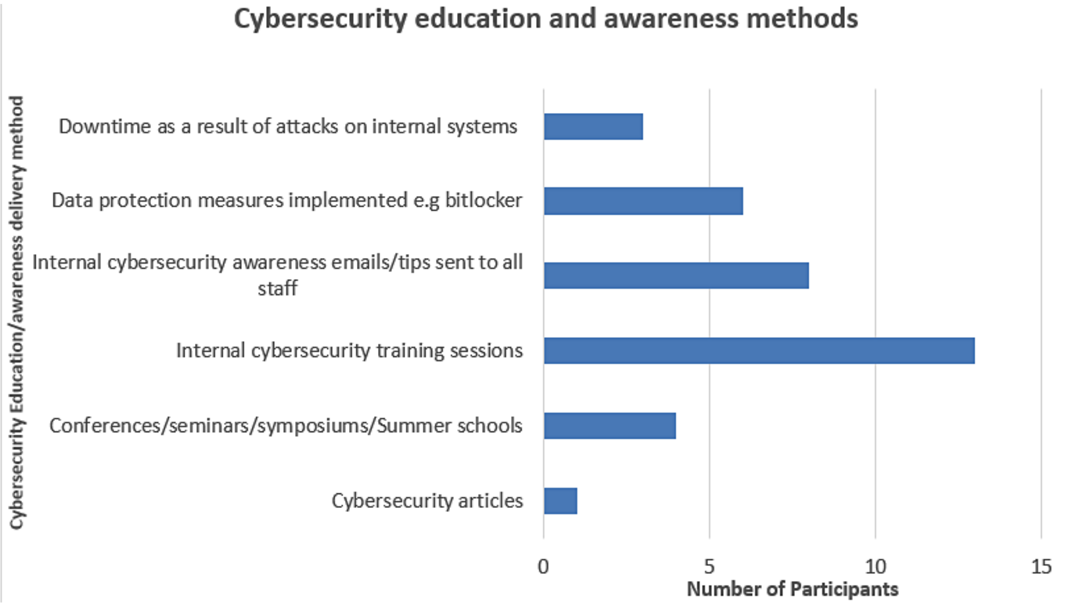


Fig. 13: Level of Education and Awareness of Cybersecurity Risks.

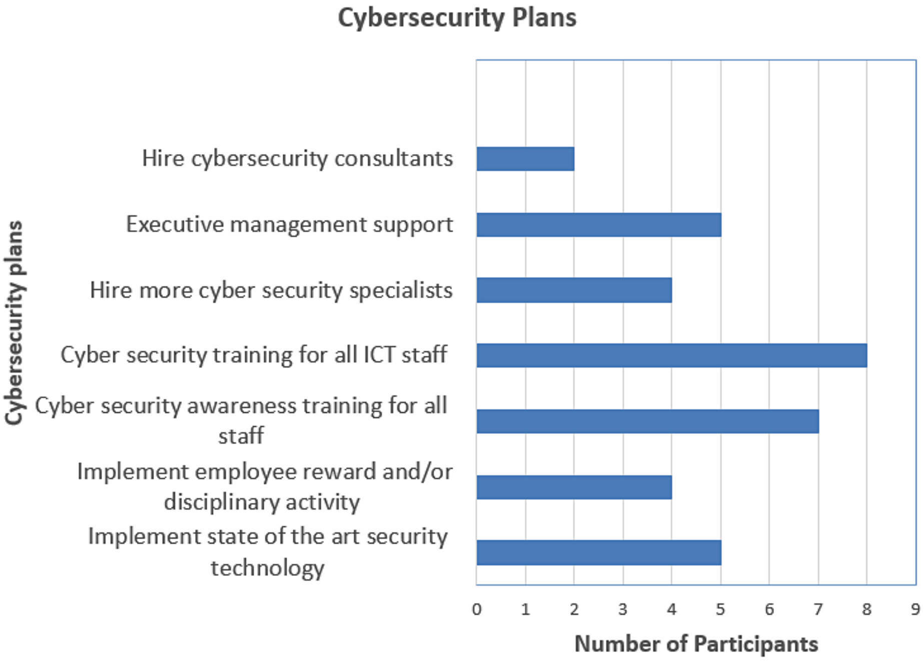


Fig. 14: Cybersecurity Plans

Evaluation of the performance of the Q-learning algorithm and the A3C method for reinforcement learning.

The Asynchronous Advantage Actor-Critic (A3C) Algorithm is much faster, simpler, and scores higher on Deep Reinforcement Learning tasks.[2] A3C completely blows most algorithms like Deep Q Networks (DQN). The Asynchronous Advantage Actor-Critic (A3C) Algorithm pseudo-code was implemented in Python. A3C is faster, simpler, and scores higher on Deep Reinforcement Learning tasks. It uses a neural network that has one softmax output for the policy π(at| st; θ) and one linear output for value function V(st; v). The main distinction between DL and other machine-learning methods is the attempt to directly extract high-level characteristics from data. As a result, DL lessens the effort required to create a feature extractor for each issue. The fastest DL algorithm, like ResNet, needs exactly two weeks to finish a training session, whereas ML training only takes seconds to hours.

Support Vector Machine (SVM) is one of the most reliable and precise techniques. SVM employs the proper kernel functions to map data points onto higher dimensional spaces. Since machine learning algorithms are so effective at identifying new threats, a lot of research has been done using this combination. The Internet of Things (IoT) poses a challenge to network security, making the search for IoT-specific intrusion detection techniques crucial. Detecting these attacks is difficult due to the rise in cyberattacks and cybercrime against cyber-physical systems (CPSs). Machine learning-based cybersecurity systems rarely need updates and maintenance since they can stay current by learning from acquired data. Deep learning technologies offer an automatic learning process that requires little human input. These methods require more care and knowledge from the developers because they are more complex than shallow machine learning methods. Cybersecurity ensures data's safety, which is crucial to the continued operation of organizations.

Comparative performance of the A3C Algorithm with other RL algorithms are shown on Figure 15.

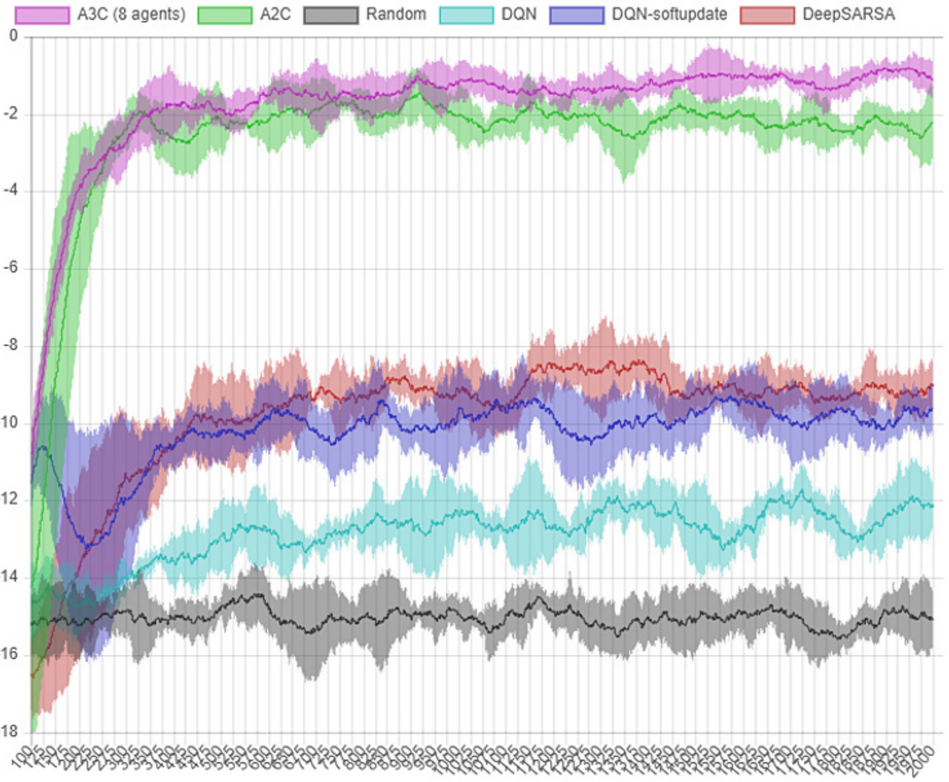


Fig. 15: Comparative performance of RL A3C Algorithm

Identification of the Reinforcement Learning paradigm that may affect cybersecurity training and education.

Systems that are connected to the Internet have grown significantly in size, making them more vulnerable than ever to cyberattacks. Cyber threats are complex and dynamic, thus defenses must be quick to react, flexible, and scalable.[43] Methods based on machine learning, most notably deep reinforcement learning (DRL), have been widely suggested to deal with these problems. DRL is incredibly good at tackling complicated, dynamic, and especially high-dimensional cyber protection challenges since it combines deep learning with classical RL.[43] As a subset of machine learning (ML), reinforcement learning (RL) is the most similar to how humans learn since it may get knowledge from its own exploration and exploitation of uncharted territory. RL is highly adaptive and helpful in real-time and adversarial contexts because it can model an autonomous agent to conduct consecutive actions ideally without or with minimal prior information of the environment.[43] Deep learning has been included into RL approaches and has given them the ability to solve a wide variety of complex issues thanks to the strength of function approximation and representation learning. Because cyberattacks are becoming more complex, quick, and pervasive, the combination of deep learning and RL suggests high applicability for cybersecurity applications. Recently, DRL techniques have been used to address a variety of issues in the Internet of Things (IoT) space.

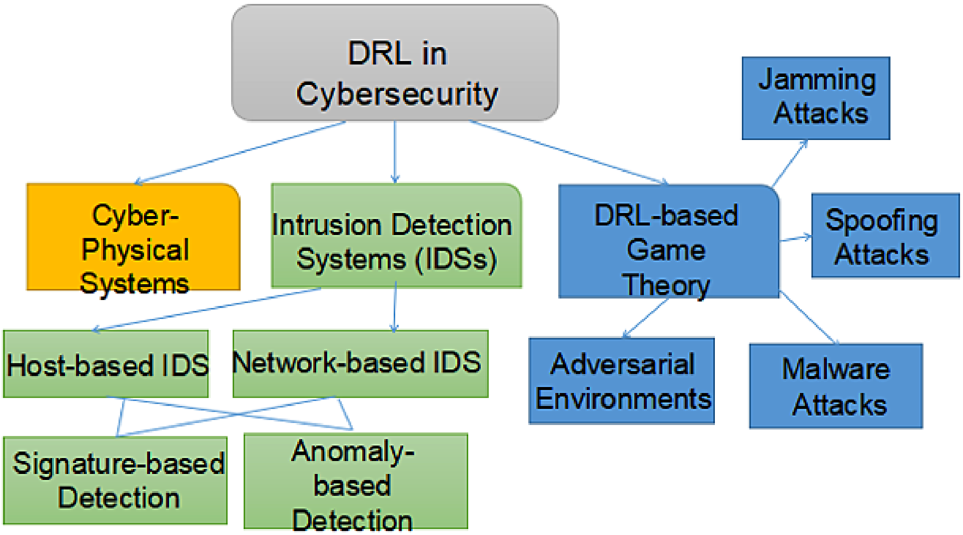


Fig. 16: Different sub-sections of DRL in Cybsersecurity (Source [43]).

According to the results of a survey conducted by,[43] applications of DRL in cyber settings may generally be divided into two categories: increasing and optimizing the networking and communications capabilities of IoT applications, and fighting against cyberattacks. Contrary to the other well-known field of machine learning, supervised methods learning from examples, RL defines an agent by allowing it to generate its own learning experiences through direct interaction with the environment. Concepts of *state, action, and reward* are used to describe RL. To store expected rewards (Q-values) for actions given a set of states, Q-learning requires the use of a lookup table or Q-table. When the state and action spaces grow, this necessitates a considerable amount of memory. Since continuous state or action spaces are frequently present in real-world issues, Q-learning is ineffective for solving them. Fortunately, deep learning has become a potent tool that is a fantastic addition to conventional RL methods. Deep learning techniques typically employ function approximation and representation learning, which enable them to efficiently learn a condensed low-dimensional representation of unstructured, high-dimensional data.[43] To help Q-learning handle high-dimensional sensory inputs, the deep Q-network (DQN) takes use of a deep neural network (DNN). The hierarchy of A3C's structure shown on Figure 9 is made up sssof both individual learners and a master learning agent (global) (workers). DNNs simulate both the master agent and individual learners, with each having two outputs: one for the actor and one for the critic. In recent years, the number of linked IoT devices has increased, which has significantly increased both the frequency and complexity of cyber attacks. According to,[43] the emergence of deep learning and its integration with RL have produced a class of DRL methods that are able to identify and counteract sophisticated cyberattacks, including intrusions into host computers or networks, distributed denial-of-service attacks, attacks using malware, spoofing, jamming, and injected false data into cyber-physical systems, deception attacks against autonomous systems, etc. Figure 16 shows the different sub-sections of DRL in cybersecurity.

The cybersecurity research community has given investigations into protection strategies for cyber-physical systems (CPS) against cyber threats a great deal of attention and interest. Internet integration makes it possible for computer-based algorithms to control the CPS mechanism. Through the shared network, this approach offers effective administration of distributed physical systems.[43] After then, the CPS defense problem is modeled as a two-player, zero-sum game in which each player's utilities are added up to 0. An actor-critic DRL algorithm serves as the representation for the defender. A certain amount of correctness is necessary for applications of CPS in important safety areas including autonomous vehicles, chemical processes, automatic pilot avionics, and smart grid. Future smart cities will require autonomous vehicles (AVs) to operate with a powerful processing unit of intra-vehicle sensors, such as cameras, radar, roadside smart sensors, and inter-vehicle beaconing. Such reliance leaves AVs open to cyberphysical attacks that manipulate sensory input in an effort to take over the devices, increasing the likelihood of accidents or decreasing traffic flow, for example. Security professionals typically need to observe and analyse audit data, such as application traces, network traffic flow, and user command data, to distinguish between normal and anomalous behaviour in order to discover intrusions. A software or hardware platform known as an intrusion detection system (IDS) is installed on host computers or network equipment in order to monitor audit data in order to identify and notify the administrator of any suspicious or malicious activity. A system for intrusion detection and prevention might be able to take the necessary steps right away to lessen the effects of the malicious activity.

Antivirus software, firewalls, and intrusion detection systems are examples of traditional cybersecurity measures that are typically passive, unilateral, and lagging behind dynamic attacks. Because there are many different cyber components in cyberspace, reliable cybersecurity must take into account how these components interact. Particularly, the choices made by other components are somewhat influenced by the security policy that has been applied to a component. As a result, when the system is huge, the decision space grows significantly and contains a lot of what-if situations. Because it can evaluate a variety of scenarios to determine the optimum course of action for each actor, game theory has been shown to be helpful in resolving such complex issues.[43] Table 13 below shows the summary of typical DRL applications in cybersecurity, indicating the appropriate algorithms used by.[43]

AI can aid in the defense against cyberattacks, but it can also enable risky attacks, such as aggressive AI. AI can be used by hackers to enhance their attacks and make them more sophisticated in order to evade security measures and breach networks or computer systems. For instance, hackers may use algorithms to track users' typical habits and leverage those patterns to create untraceable attack methods. Large-scale phishing assaults can be carried out using convincing false communications that are created by machine learning-based systems that can resemble people. Similarly, hackers can rig elections or manipulate financial markets by disseminating misleading information using very realistic fake video or audio communications created using AI advancements (also known as deepfakes).[43]

Table 13: Summary of DRL applications in Cybersecurity with their respective algorithms.

Applications	Goals/ Objectives	Algorithms
Robustness-guided falsification of CPS	Find falsifying inputs (counter examples) for CPS	Double deep Q-network (DQN) and A3C
Security and safety in autonomous vehicle systems	Maximize the robustness of AV dynamics control to cyberphysical attacks that inject faulty data to sensor readings.	Q-learning with Long short term memory (LSTM)
Increasing robustness of the autonomous system against adversarial attacks	Devise filtering schemes to detect corrupted measurements (deception attack) and mitigate the effects of adversarial errors.	Trust region policy optimization (TRPO)
Secure offloading in mobile edge caching	Learn a policy for a mobile device to securely offload data to edge nodes. against jamming and smart attacks	DQN with hotbooting transfer learning technique
Anti-jamming communication scheme for CRN	Derive an optimal frequency hopping policy for CRN SUs to defeat smart jammers based on a frequency-spatial anti-jamming game.	DQN that employs CNN
Anti-jamming communication method, improving the previous work	Propose a smart anti-jamming Scheme with two main differences: spectrum waterfall is used as the state, and jammers can have different channel-slot transmission structure with users.	DQN with recursive CNN due to recursion characteristic of spectrum waterfall.
Spoofing detection in wireless networks	Select the optimal authentication threshold.	Q-learning and Dyna-Q
Mobile offloading for cloud-based malware detection	Improve malware detection accuracy and speed.	Hotbooting Q-learning and DQN.
Autonomous defense in SDN	Tackle the poisoning attacks that manipulate states or flip reward signals during the training process of RL-based defense agents.	Double DQN and A3C
Secure mobile crowdsensing (MCS) system	Optimize payment policy to improve the sensing performance against faked sensing attacks by formulating a Stackelberg game.	DQN
Automated URL based phishing detection	Detect malicious websites (URLs)	DQN

Conclusion

Reinforcement learning (RL) is a type of learning in which long-term goals related to the environment are attained through learning from interactions with the environment. RL is hypothesis-based and goal-oriented when action sequences, observations, and rewards are used as inputs. Reinforcement learning (RL) is thought to be one of the most successful and efficient learning approaches that may be utilized for cybersecurity education and training, despite the fact that there are many other forms of learning. Compared to supervised, semi-supervised, and unsupervised learning systems, RL has numerous benefits. This is because RL algorithms commonly make use of dynamic programming approaches, giving them the advantage of using the knowledge they have already acquired while also exploring new potential courses of action without depending on predetermined knowledge. These characteristics are crucial because they demonstrate that RL may be employed successfully in the current environment, where cyberthreats change on a daily basis. It was crucial for this research to find an RL paradigm that is appropriate for cybersecurity teaching and training because cybersecurity impacts everyone. For the objectives of cybersecurity education and training, a reinforcement-learning paradigm was found for this study and is regarded as being quite effective. The majority of DRL algorithms currently being utilized for cyber protection are model-free methods, which are sample inefficient due to their high training data requirements. In actual cybersecurity practice, obtaining these data might be challenging. When training data is scarce, model-based DRL approaches are preferable to model-free methods because it is often simple to obtain data in a scalable manner. Therefore, an interesting future study would involve investigating model-based DRL techniques or combining model-based and model-free techniques for cybersecurity. Cyber security professionals no longer manually review a large number of attack data to identify and prevent cyberattacks with the use of AI algorithms. The security staff cannot handle the volume alone, thus this offers several benefits. Defense plans that are AI-enabled can be automated, deployed quickly, and effectively, yet these systems alone cannot generate original defenses in the face of fresh threats. Additionally, human enemies are always responsible for cybercrime or cyberwarfare. Therefore, human intelligence combined with technology is essential for cyber protection.

A survey was conducted on the cybersecurity education and training as exemplified by Zimbabwean commercial banks. The study population encompassed employees and customers from five commercial banks (Stanbic Bank, ZB Bank, CBZ, BancABC and Nedbank Zimbabwe), where the sample size was 370. The lack of a cybersecurity culture among Zimbabwean enterprises, including banks, leaves them vulnerable to assaults like phishing, hacking, harmful software, identity theft, and card fraud. The ability of financial organizations to steal consumers' sensitive information has made them one of the top targets for hackers. Machine learning (ML) has been used to thwart cybersecurity threats in Zimbabwe's banks. The effectiveness of ML in underdeveloped countries has been specifically impacted by three issues: a lack of awareness, weak rules, and implementation costs. Data hurdles, the black box aspect of the models, validation challenges, model testing and outcomes analysis challenges, as well as issues with models created by suppliers were additional system obstacles for banks.

Systems connected to the Internet are more vulnerable than ever to cyberattacks. Methods based on machine learning, most notably deep reinforcement learning (DRL), have been widely suggested to deal with these problems. DRL techniques have been used to address a variety of issues in the Internet of Things (IoT) space. Agents are defined by their interactions with the environment. DRL can be used to fight against cyberattacks, such as intrusions into host computers or networks and fake data in IoT devices.

Acknowledgement
The author would like to thank, Machine Learning Woxsen School of Business, Woxsen University, Hyderabad, India. for their guidance and support to complete this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest
The authors do not have any conflict of interest.

References

1. Kammann, L. (2018). *Digitalisierung im Versicherungsvertrieb: Eine Untersuchung der rechtlichen Grenzen und Möglichkeiten unter besonderer Berücksichtigung der Versicherungsvergleichsportale.* VVW GmbH.
2. Kabanda, G., (2022), *"Face Recognition in Machine Learning: A Framework for Dimensionality Reduction Algorithms",*

International Journal of Advanced Networking and Applications (IJANA), Volume: 14, Issue: 02, September-October, 2022, Pages: 5396-5407 (2022), ISSN: 0975-0290, http://www.ijana.in/, https://www.ijana.in/papers/V14I2-11.pdf.

3. Saravanan, R., & Sujatha, P. (2018, June). A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In 2018 *Second International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 945-949). IEEE.
4. Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine learning from theory to algorithms: an overview. *In Journal of physics*: conference series (Vol. 1142, No. 1, p. 012012). IOP Publishing.
5. Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence,* 2, 100017.
6. Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning applications-a vision. *Global Transitions Proceedings,* 2(1), 24-28
7. Oh, D. Y., & Yun, I. D. (2018). Residual error based anomaly detection using auto-encoder in SMD machine sound. *Sensors,* 18(5), 1308.
8. Malekloo, A., Ozer, E., AlHamaydeh, M., & Girolami, M. (2022). Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring,* 21(4), 1906-1955.
9. Divya, K. S., Bhargavi, P., & Jyothi, S. (2018). Machine learning algorithms in big data analytics. *Int. J. Comput. Sci.* Eng, 6(1), 63-70.
10. Rochan, M. (2020). Efficient deep learning models for video abstraction.
11. Sáray, S., Rössert, C. A., Appukuttan, S., Migliore, R., Vitale, P., Lupascu, C. A., ... & Káli, S. (2021). HippoUnit: A software tool for the automated testing and systematic comparison of detailed models of hippocampal neurons based on electrophysiological data. *PLoS computational biology,* 17(1), e1008114.
12. Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. *In Emerging technology in modelling and graphics* (pp. 99-111). Springer, Singapore.
13. Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering,* 27(4), 1071-1092.
14. Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015, June). Unsupervised learning of video representations using lstms. *In International conference on machine learning* (pp. 843-852). PMLR.
15. Lemenkova, P. (2018, November). Hierarchical cluster analysis by R language for pattern recognition in the bathymetric data frame: a Case study of the Mariana Trench, Pacific Ocean. In *Virtual Simulation, Prototyping and Industrial Design. Proceedings of 5th International Scientific-Practical Conference* (Vol. 2, No. 5, pp. 147-152).
16. Nozari, H., & Sadeghi, M. E. (2021). Artificial intelligence and Machine Learning for Real-world problems (A survey). *International Journal of Innovation in Engineering,* 1(3), 38-47.
17. Mooney, S. J., & Pejaver, V. (2018). Big data in public health: terminology, machine learning, and privacy. *Annual review of public health,* 39, 95.
18. Alpaydin, E. (2020). *Introduction to machine learning.* MIT press.
19. Iscen, A., Tolias, G., Avrithis, Y., & Chum, O. (2019). Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5070-5079).
20. Jeong, J., Lee, S., Kim, J., & Kwak, N. (2019). Consistency-based semi-supervised learning for object detection. *Advances in neural information processing systems,* 32.
21. Y. Xin et al., "Machine Learning and Deep Learning Methods for Cybersecurity," in IEEE Access, vol. 6, pp. 35365-35381, 2018, doi: 10.1109/ACCESS.2018.2836950.
22. J. Zhang, L. Pan, Q. -L. Han, C. Chen, S. Wen and Y. Xiang, "Deep Learning Based Attack Detection for Cyber-Physical System Cybersecurity: A Survey," in IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 3, pp. 377-391, March 2022, doi: 10.1109/JAS.2021.1004261.
23. T. T. Nguyen and V. J. Reddi, "Deep Reinforcement Learning for Cyber Security," in IEEE Transactions on Neural Networks

and Learning Systems, doi: 10.1109/ TNNLS.2021.3121870.

24. Alghamdi, M. I. (2020). Survey on Applications of Deep Learning and Machine Learning Techniques for Cyber Security. *International Journal of Interactive Mobile Technologies,* 14(16).
25. Kivunja, ,. C. & Kuyini, B. A., 2017. Understanding and Applying Research Paradigms in Educational Contexts. I*nternational Journal of Higher Education,* 6(26).
26. Mohajan, 2017. Qualitative Research Methodology in Social Sciences and Related Subjects. Journal of Economic Development, Environment and People, Volume 7, pp. 23-48.
27. Siedlecki, S.L. (2020), "Understanding Descriptive Research Designs and Methods", *Clinical Nurse Specialist,* Lippincott Williams and Wilkins, Vol. 34 No. 1, pp. 8–12.
28. Kumar, M. (2022), "Classification of Research Design: Descriptive, Diagnostic, Exploratory and Experimental".
29. Casula, M., Rangarajan, N. and Shields, P. (2021), "The potential of working hypotheses for deductive exploratory research", *Quality and Quantity,* Springer Science and Business Media B.V., Vol. 55 No. 5, pp. 1703–1725.
30. Toyon, M.A.S. (2021), "Explanatory sequential design of mixed methods research: Phases and challenges", *International Journal of Research in Business and Social Science* (2147- 4478), Center for Strategic Studies in Business and Finance SSBFNET, Vol. 10 No. 5, pp. 253–260.
31. Dileep, P. K., Tröger, J. A., Hartmann, S., & Ziegmann, G. (2022). Three-dimensional shear angle determination with application to shear-frame test. *Composite Structures,* 285, 115134.
32. Dawson, M. (2020). National Cybersecurity Education: Bridging Defense To Offense. *Land Forces Academy Review* Vol. XXV, No 1(97), 2020
33. Kortjan, N., & Solms, R. Von. (2014). A Conceptual Framework for Cyber-Security Awareness and Education in SA. *South African Computer Journal,* 52, 29-41.
34. Catota, F., E., Morgan, G., Sicker, D., C. (2019). Cybersecurity education in a developing nation: the Ecuadorian environment. *Journal of Cybersecurity,* 2019, 1–19 doi: 10.1093/ cybsec/tyz001
35. South African Government Gazette, (2015). National Cybersecurity Policy Framework for South Africa.
36. Rahman, N. A. A, Sairi, I. H., Zizi, N. A. M., and Khalid, F. (2020) The Importance of Cybersecurity Education in School. *International Journal of Information and Education Technology,* Vol. 10, No. 5, May 2020
37. Nakama, D., and Paullet, K. (2019). The urgency for cybersecurity education: The impact of early college innovation in Hawaii rural communities. *Information System Education Journal,* vol. 16, no. 4, pp. 41-52, 2019
38. Khader, M., Karam, M.,Fares, H. (2021). Cybersecurity Awareness Framework for Academia. Information 2021, 12, 417. *https:// doi.org/ 10.3390/info12100417*
39. Mutemwa, M., Masango, M. G., & Gcaza, N. (2021, December). Managing the Shift in the Enterprise Perimeter in order to delay a Cybersecurity Breach. In *Proceedings of the International Conference on Artificial Intelligence and its Applications* (pp. 1-10).
40. Aldawood H, Skinner G (2019). Reviewing Cyber Security Social Engineering Training and Awareness Programs—Pitfalls and Ongoing Issues. *Future Internet* 2019, 11(3), 73; https:// doi.org/10.3390/fi11030073
41. Bada, M., Sasse, A. M., Nurse, J. R. C. (2019). Cyber Security Awareness Campaigns: Why do they fail to change behaviour?
42. Chowdhury, N., and Gkioulos, V. (2021). Cyber security training for critical infrastructure protection: A literature review. *Computer Science Review .Volume* 40, May 2021. https:// doi.org/10.1016/j.cosrev.2021.100361
43. Nguyen, T. T., & Reddi, V. J. (2021). Deep reinforcement learning for cyber security. *IEEE Transactions on Neural Networks and Learning Systems.*