Machine learning

In Cyber Security

Team Member

Anurag Pareek (89)
Animesh Parab (88)

1. Introduction

5 min

➢ What Is Machine Learning in Security?

- **Machine learning (ML)** lets computers learn without being explicitly programmed. Put another way, machine learning teaches computers to do what people do: learn by experience. Machine learning is a domain within the broader field of artificial intelligence.

- In **security**, machine learning continuously learns by analyzing data to find patterns so we can better detect malware in encrypted traffic, find insider threats, predict where “bad neighborhoods” are online to keep people safe when browsing, or protect data in the cloud by uncovering suspicious user behavior.

2. How ML Works

15 min

➢ How does machine learning work in security?

- The cyber threat landscape forces organizations to constantly track and correlate millions of external and internal data points across their infrastructure and users. It simply is not feasible to manage this volume of information with only a team of people.
- This is where machine learning shines, because it can recognize patterns and predict threats in massive data sets, all at machine speed. By automating the analysis, cyber teams can rapidly detect threats and isolate situations that need deeper human analysis.

➢ How does it work?

- The details of machine learning can seem intimidating to non-data scientists, so let's look at some key terms.
- Supervised learning calls on sets of training data, called “ground truth,” which are correct question-and-answer pairs. This training helps classifiers, the workhorses of machine learning analysis, to accurately categorize observations. It also helps algorithms, used to organize and orient classifiers, successfully analyze new data in the real world.
- An everyday example is recognizing faces in online photos: Classifiers analyze the data patterns they are trained on--not the actual noses or eyes--in order to correctly tag a unique face amongst many millions of online photos.

3. How machine learning helps security...

30 min

Find threats on a network

Machine learning detects threats by constantly monitoring the behavior of the network for anomalies. Machine learning engines process massive amounts of data in near real time to discover critical incidents. These techniques allow for the detection of insider threats, unknown malware, and policy violations.

Keep people safe when browsing

Machine learning can predict "bad neighborhoods" online to help prevent people from connecting to malicious websites. Machine learning analyzes Internet activity to automatically identify attack infrastructures staged for current and emergent threats.

Provide endpoint malware protection

Algorithms can detect never-before-seen malware that is trying to run on endpoints. It identifies new malicious files and activity based on the attributes and behaviors of known malware.

Protect data in the cloud

Machine learning can protect productivity by analyzing suspicious cloud app login activity, detecting location-based anomalies, and conducting IP reputation analysis to identify threats and risks in cloud apps and platforms.

Detect malware in encrypted traffic

Machine learning can detect malware in encrypted traffic by analyzing encrypted traffic data elements in common network telemetry. Rather than decrypting, machine learning algorithms pinpoint malicious patterns to find threats hidden with encryption.

3 Types of Machine Learning in Cybersecurity

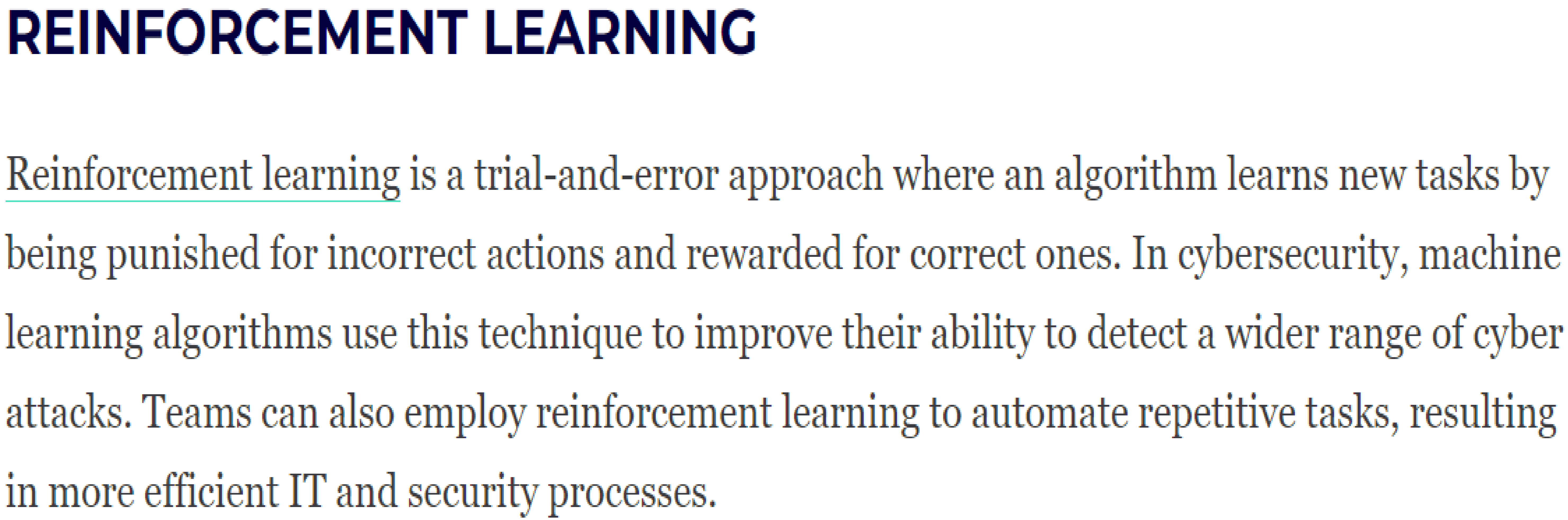
There are three types of machine learning used in cybersecurity: supervised learning, unsupervised learning and reinforcement learning.

SUPERVISED LEARNING

Supervised learning involves training an algorithm on labeled data, so it learns how to organize data based on the relationships between inputs and outputs. Human guidance is often needed to assist algorithms during training. Machine learning algorithms use supervised learning to classify data as neutral or harmful, identifying threats like denial-of-service attacks and predicting future cyber attacks.

UNSUPERVISED LEARNING

Unsupervised learning refers to an algorithm trained on unlabeled or raw data, and it labels and classifies data without human guidance. Security teams rely on unsupervised learning to train algorithms to detect new and more complicated cyber attacks, especially as hackers develop different techniques to infiltrate company defenses.



REINFORCEMENT LEARNING

Reinforcement learning is a trial-and-error approach where an algorithm learns new tasks by being punished for incorrect actions and rewarded for correct ones. In cybersecurity, machine learning algorithms use this technique to improve their ability to detect a wider range of cyber attacks. Teams can also employ reinforcement learning to automate repetitive tasks, resulting in more efficient IT and security processes.

How Is Machine Learning Used in Cybersecurity?

A subset of artificial intelligence, machine learning uses algorithms born of previous datasets and statistical analysis to make assumptions about a computer's behavior. The computer can then adjust its actions, even performing functions it wasn't programmed to do. These abilities have made machine learning a crucial cybersecurity asset.

HOW IS MACHINE LEARNING USED IN CYBERSECURITY?

- Detecting threats in early stages
- Uncovering network vulnerabilities
- Reducing IT workloads and costs

DETECTING THREATS IN EARLY STAGES

With its ability to sort through millions of files and identify potentially hazardous ones, machine learning is increasingly used to uncover threats and squash them before they can wreak havoc.

Software from Microsoft showcased this skill in 2018, when cybercrooks attempted to infect over 400,000 users with a cryptocurrency miner during a 12-hour time frame. The attack was stopped by Microsoft's Windows Defender, a software that employs multiple layers of machine learning to identify and block perceived threats. The crypto miners were shut down almost as soon as they started digging.

UNCOVERING NETWORK VULNERABILITIES

Rather than wait for cyber attacks to happen, companies are taking a more proactive approach with machine learning. Penetration testing involves simulating a cyber attack to locate weak points in a company's networks, firewalls and systems. Machine learning can execute this task and apply software patches, code fixes and other solutions to address any holes in an organization's security suite.

In addition, machine learning's ability to learn from historical data allows it to pick up on unusual software and user behavior during these kinds of training sessions. The technology then remembers how specific cyber attacks occur and can determine which ones pose the biggest threats based on a network's vulnerabilities.

Benefits of Machine Learning in Cybersecurity

With its range of applications, machine learning offers many advantages to IT and security personnel.

AUTOMATED CYBERSECURITY PROCESSES

Machine learning can learn new functions and get better at performing existing ones on its own, resulting in automated workflows. Security and IT teams can then leave basic responsibilities to machine learning while focusing their time and resources on addressing new cyber threats, fixing urgent flaws and completing other advanced tasks.

STRENGTHENED SECURITY PROCEDURES

Reviewing a company's security infrastructure, machine learning algorithms can expose weak points, recommend fixes and help teams prepare for a variety of cyber attacks. This way, security and IT teams can address threats before they even happen, establishing the procedures and systems needed to fend off more complex attacks.

ABILITY TO HANDLE LARGE DATA SETS

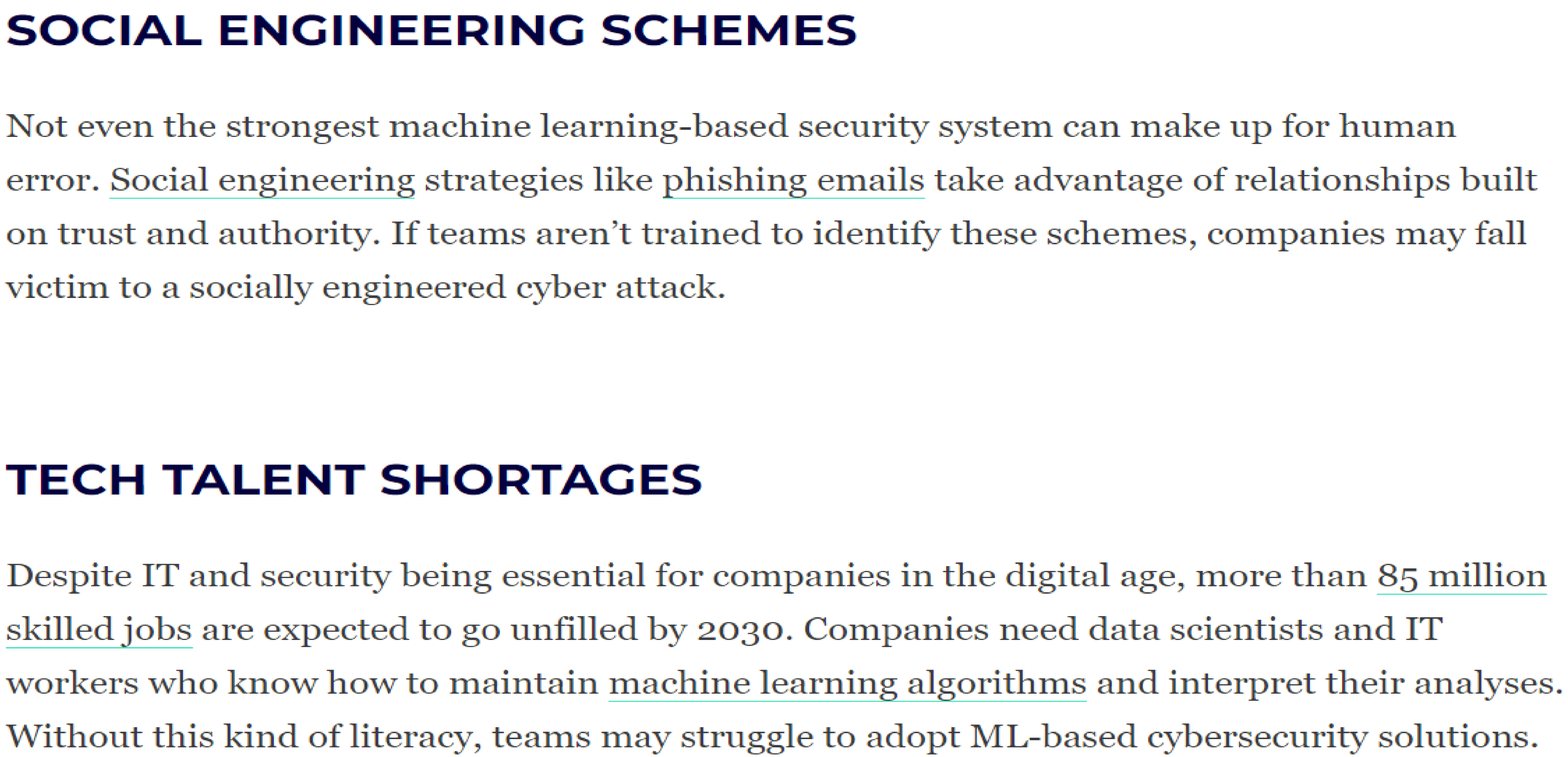
Humans may struggle to deal with large volumes of data, but machine learning can quickly process and analyze larger data sets. Algorithms can spot trends much faster than humans and alert teams of developing cyber attacks. IT and security personnel can then take immediate action, snuffing out cyber attacks in their early stages before they spread.

Machine Learning in Cybersecurity Challenges

While machine learning in cybersecurity meets various IT and security needs for businesses, the technology must continue to adapt to an ever-changing digital ecosystem. Even then, machine learning may not be able to overcome some limitations and outside factors.

INCREASING NUMBER OF CONNECTIONS

The number of connected devices is expected to reach 29 billion by 2027 as hybrid and cloud environments become more popular. Company networks are constantly adding new computers, tablets and other devices, putting pressure on machine learning to account for and protect more connections against cyber attacks.

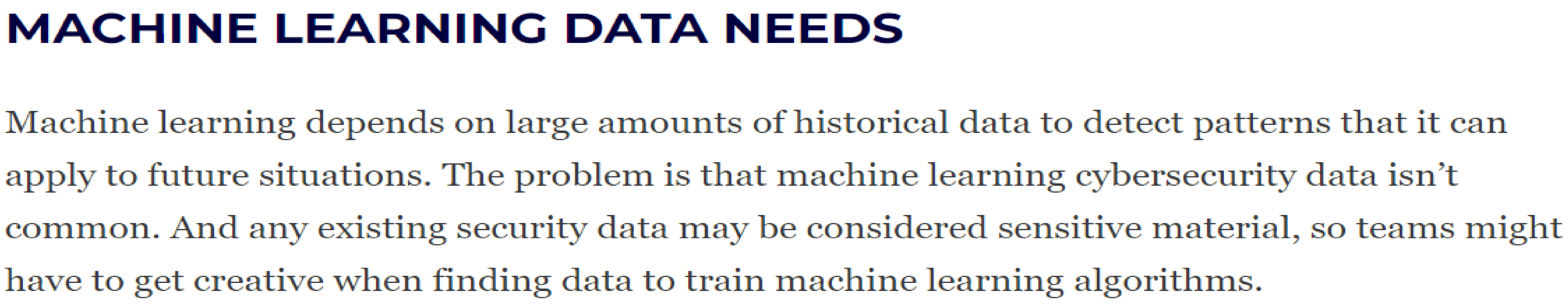


SOCIAL ENGINEERING SCHEMES

Not even the strongest machine learning-based security system can make up for human error. Social engineering strategies like phishing emails take advantage of relationships built on trust and authority. If teams aren't trained to identify these schemes, companies may fall victim to a socially engineered cyber attack.

TECH TALENT SHORTAGES

Despite IT and security being essential for companies in the digital age, more than 85 million skilled jobs are expected to go unfilled by 2030. Companies need data scientists and IT workers who know how to maintain machine learning algorithms and interpret their analyses. Without this kind of literacy, teams may struggle to adopt ML-based cybersecurity solutions.



MACHINE LEARNING DATA NEEDS

Machine learning depends on large amounts of historical data to detect patterns that it can apply to future situations. The problem is that machine learning cybersecurity data isn't common. And any existing security data may be considered sensitive material, so teams might have to get creative when finding data to train machine learning algorithms.