**Cyber Security**

Cyber security is the practice of protecting digital devices, networks, and sensitive data from cyber threats such as hacking, malware, and phishing attacks.” It involves a range of strategies, technologies, and best practices designed to safeguard computers, networks, and data from cyber attacks.

Cyber Security involves using specialized tools to detect and remove harmful software while also learning to identify and avoid online scams. Practicing good cybersecurity habits helps keep your data private and ensures a safe online experience. It’s also referred to as Information Security (INFOSEC), Information Assurance (IA), or System Security.

**What is Cyber Security? (Definition & Importance)**

Cybersecurity is all about protecting your computer, phone, or any digital device from hackers and online threats. It keeps your personal information, bank details, files, and online activity safe from being stolen, damaged, or misused. By acquiring knowledge of cyber-attacks and cyber security we can secure and defend ourselves from various cyber-attacks like phishing and DDoS attacks.

**Common Cyber Attacks**

| **Attack Type** | **What It Means** |
| --- | --- |
| [**Phishing**](https://www.geeksforgeeks.org/what-is-phishing/) | Phishing is a cyber attack where hackers trick users into revealing sensitive data like passwords, banking details, or session tokens through fake emails, messages, or websites. It uses [social engineering](https://www.geeksforgeeks.org/social-engineering-the-art-of-virtual-exploitation/) to impersonate trusted sources and often includes malicious links or attachments to steal information. |
| [**DDoS (Distributed Denial of Service)**](https://www.geeksforgeeks.org/what-is-ddosdistributed-denial-of-service/) | Hackers flood a website or server with too much traffic, so it slows down or crashes. It’s like a traffic jam that blocks real users from getting in. |

**Encryption**

One crucial aspect of cybersecurity is [Encryption](https://www.geeksforgeeks.org/what-is-data-encryption/), which ensures that sensitive information remains private and readable only to authorized users. This is especially important for financial transactions, personal communications, and corporate databases to prevent data theft and unauthorized access

In short, cybersecurity keeps your online world safe and secure. It ensures that sensitive information remains confidential, intact, and accessible only to authorized users. Whether it’s securing personal information, financial transactions, or corporate databases.

**Types of Cybersecurity**

There are seven types of cyber security, each explained below in detail with uses and functions:

1. **Network Security (Protecting Networks from Unauthorized Access)**

It focuses on securing computer networks from unauthorized access, data breaches, and other network-based threats. This involves implementing technologies such as [Firewalls](https://www.geeksforgeeks.org/introduction-of-firewall-in-computer-network/), [Intrusion detection systems (IDS)](https://www.geeksforgeeks.org/intrusion-detection-system-ids/), Virtual private networks (VPNs), and Network segmentation as well as deploying [antivirus software](https://www.geeksforgeeks.org/what-is-antivirus-software/)

* Using public Wi-Fi in locations like cafes and malls poses significant security risks. Malicious actors on the same network can potentially intercept your online activity, including sensitive information. If you use payment gateways on these unsecured networks, your financial data could be compromised because these open networks don’t have proper security layers, which means anyone—even hackers—can watch what you’re doing online.
* So, use a secure private network or VPN to protect your internal network from outside threats.

**2. Application Security (Ensuring Secure Software and Apps)**

Concerned with securing software applications and preventing vulnerabilities that could be exploited by attackers. It involves secure coding practices, regular software updates and patches, and application-level firewalls.

* Most of the Apps that we use on our cell phones are Secured and work under the rules and regulations of the Google Play Store.
* There are 3.553 million applications in Google Play, Apple App Store has 1.642 million, and Amazon App Store has 483 million available for users to download. With so many choices, it’s easy to assume all apps are safe—but that’s not true.
* Some apps pretend to be secure, but once installed, they collect personal data and secretly share it with third-party companies.
* The app must be installed from a trustworthy platform, not from some 3rd party website in the form of an APK (Android Application Package).
  1. **Information or Data Security (Safeguarding Sensitive Data)**

Focuses on protecting sensitive information from unauthorized access, disclosure, alteration, or destruction. It includes Encryption, Access controls, Data classification, and Data loss prevention (DLP) measures.

* Incident response refers to the process of detecting, analyzing, and responding to security incidents promptly.
* Promoting security awareness among users is essential for maintaining information security. It involves educating individuals about common security risks, best practices for handling sensitive information, and how to identify and respond to potential threats like phishing attacks or social engineering attempts.
* Encryption is the process of converting information into an unreadable format (ciphertext) to protect it from unauthorized access.
  1. **Cloud Security (Defending Cloud Storage and Applications)**

It involves securing data, applications, and infrastructure hosted on cloud platforms, and ensuring appropriate access controls, data protection, and compliance. It uses various cloud service providers such as [AWS](https://www.geeksforgeeks.org/aws-tutorial/), [Azure](https://www.geeksforgeeks.org/microsoft-azure/), [Google Cloud](https://www.geeksforgeeks.org/google-cloud-platform-tutorial/), etc., to ensure security against multiple threats.

* Cloud-based data storage has become a popular option over the last decade. It enhances privacy if configured and managed correctly and saves data on the cloud, making it accessible from any device with proper authentication.
* These platforms offer free tiers for limited usage, and users must pay for additional storage or services
* It is a cloud service provider that offers a wide range of services, including storage, computing, and security tools.
  1. **Endpoint Security (Protecting Devices like Laptops & Phones)**

Refers to securing individual devices such as computers, laptops, smartphones, and IoT devices. It includes antivirus software, intrusion prevention systems (IPS), device encryption, and regular software updates.

* Antivirus and [Anti-malware](https://www.geeksforgeeks.org/what-is-anti-malware/) software that scans and detects malicious software, such as [Viruses](https://www.geeksforgeeks.org/what-is-a-computer-virus/), [Worms](https://www.geeksforgeeks.org/what-is-computer-worm/), Trojans, and Ransomware. These tools identify and eliminate or quarantine malicious files, protecting the endpoint and the network from potential harm.
* Firewalls are essential components of endpoint security. They monitor and control incoming and outgoing network traffic, filtering out potentially malicious data packets.
* Keeping software and operating systems up to date with the latest security patches and updates is crucial for endpoint security.
  1. **Operational Security (Managing Internal Security Protocols)**

Refers to the processes and policies organizations implement to protect sensitive data from internal threats and human errors. It involves access controls, risk management, employee training, and monitoring activities to prevent data leaks and security breaches.

* Access Controls ensure that only authorized personnel can access critical systems and sensitive information. This includes role-based access, [multi-factor authentication (MFA)](https://www.geeksforgeeks.org/multifactor-authentication/), and least privilege principles.
* Risk Management involves identifying, analyzing, and mitigating security risks within an organization. It includes regular security assessments, [vulnerability testing](https://www.geeksforgeeks.org/vulnerability-testing/), and compliance audits.
* Employee Training is crucial for preventing insider threats and social engineering attacks. Organizations conduct cybersecurity awareness programs to educate employees on phishing scams, password security, and data handling best practices.
* Monitoring & Incident Response includes tracking user activity, detecting suspicious behavior, and responding to security incidents in real time. Security Information and Event Management (SIEM) tools help organizations analyze and mitigate threats effectively.
  1. **Internet of Things (IoT) Security**

Refers to protecting internet-connected devices such as smart home gadgets, industrial sensors, medical equipment, and wearable technology from [cyber threats](https://www.geeksforgeeks.org/cybersecurity-threats/). IoT security ensures that these devices do not become entry points for hackers to exploit networks and steal sensitive data.

* Device Authentication & Encryption ensures that only authorized devices can connect to networks. Encryption protects data transmitted between IoT devices and servers from interception.
* Firmware & Software Updates are crucial to patch security vulnerabilities. Regular updates help prevent exploitation by [cybercriminals](https://www.geeksforgeeks.org/cyber-criminals-and-its-types/) who target outdated IoT firmware.
* Network Segmentation isolates IoT devices from critical systems, reducing the risk of widespread attacks if one device is compromised. This approach limits unauthorized access and lateral movement within a network.
* IoT Security Standards & Compliance include implementing industry security frameworks like [Zero Trust Architecture (ZTA)](https://www.geeksforgeeks.org/zero-trust-architecture-in-security/) and following best practices such as strong password policies, secure APIs, and endpoint protection to enhance IoT device security.

**Why is Cybersecurity is Important?**

Cyber Security is important because the government, corporations, and medical organizations, collect military, financial, process, and store unprecedented amounts of data on a computer and other properties like personal information, and this private information exposure could have negative consequences.

In 1972, when the internet was just starting (called ARPANET at the time), a test virus named Creeper was created—and then another program called Reaper was made to remove it. This early experiment showed why digital security was needed and helped start what we now call cybersecurity.

**Rising Cyber Threats: How Hackers Exploit Weak Security**

Cybercriminals are constantly finding new ways to exploit vulnerabilities in systems, networks, and personal devices. Weak passwords, outdated software, and unsecured networks create easy entry points for hackers. They use sophisticated methods like phishing emails, ransomware, and social engineering to steal sensitive data, disrupt operations, and demand ransoms.

With the rise of AI-driven cyber threats, even automated bots can breach security systems, making cybersecurity more critical than ever.

For Example – If we shop from any online shopping website and share information like email ID, address, and credit card details as well as save on that website to enable a faster and hassle-free shopping experience, then the required information is stored on a server one day we receive an email which state that the eligibility for a special discount voucher from XXXXX (hacker use famous website Name like Flipkart, Amazon, etc.) website to receive the coupon code, and we will be asked to fill the details then we will use saved card account credentials. Then our data will be shared because we think it was just an account for the verification step, and then they can wipe a substantial amount of money from our account.

**Consequences of Cyber Attacks for Businesses & Individuals**

A successful cyber attack can have devastating effects, both financially and reputationally. For businesses, a data breach can lead to massive financial losses, legal penalties, and loss of customer trust. Small businesses are especially vulnerable, as they often lack robust security measures. Individuals, on the other hand, face risks like identity theft, financial fraud, and personal data leaks.

[Cyber attacks](https://www.geeksforgeeks.org/what-is-a-cyber-attack/) can wipe out bank accounts, expose private information, and even lock users out of their own devices unless a ransom is paid. The consequences can be long-lasting, leading to emotional distress and financial instability.

**Major Cybersecurity Threats & Attacks**

Hackers use advanced techniques to find weaknesses in systems, steal or change data, and break into networks without permission. Below are the most common cybersecurity threats that target businesses, cloud storage, and personal devices:

[***Types of Cyber Attacks***](https://www.geeksforgeeks.org/types-of-cyber-attacks/)

* 1. **Malware Attacks (Viruses, Trojans, Rootkits, and Spyware)**

[Malware](https://www.geeksforgeeks.org/malware-and-its-types/) is a type of harmful software created to enter, attack, and compromise systems. It includes trojans (which look like real software but are harmful), rootkits (which hide deep inside a system to take control), and spyware (which secretly steals data).

Hackers use payload obfuscation (hides the malicious code making it harder for security software to identify), polymorphic techniques (changing malware code to avoid detection), and zero-day exploits (attacking unknown security flaws) to bypass intrusion detection systems (IDS) and endpoint protection platforms (EPP).

* 1. **Phishing & Spear Phishing Attacks**

Phishing uses tricks and manipulation to steal login details, session tokens, and financial information. [Spear phishing](https://www.geeksforgeeks.org/spear-phishing-attack/) is a more targeted version that uses open-source intelligence (OSINT) to create personalized fake messages.

Hackers use domain spoofing (making fake websites look real), homograph attacks (using similar-looking characters in URLs), and malicious macros (harmful scripts hidden in email attachments) to bypass security and trick users into revealing sensitive data.

* 1. **Ransomware Attacks (Cryptographic File Encryption)**

[Ransomware](https://www.geeksforgeeks.org/ransomware-explained-how-it-works-and-how-to-prevent-it/) locks important system files by encrypting them using [asymmetric cryptography](https://www.geeksforgeeks.org/asymmetric-key-cryptography/) (like [RSA](https://www.geeksforgeeks.org/rsa-algorithm-cryptography/), ECC) or hybrid encryption (AES-RSA). It then demands a ransom, usually in cryptocurrency, to unlock the data.

More advanced types, like double extortion ransomware, first steal sensitive data before encrypting it. Hackers then threaten to leak the stolen data on dark web sites if the ransom isn’t paid.

* 1. **Distributed Denial-of-Service (DDoS) Attacks**

DDoS attacks overload a network by flooding it with massive amounts of traffic at different levels—volumetric, protocol, or application-layer—causing servers to crash and making services unavailable.

Hackers use [botnets](https://www.geeksforgeeks.org/what-is-botnet/) (networks of infected devices), amplification techniques (like DNS reflection and NTP amplification) to increase attack size, and [HTTP flood requests](https://www.geeksforgeeks.org/http-flood-attack/) to overwhelm websites. These methods help attackers bypass rate-limiting defenses and take down their targets.

* 1. **SQL Injection (SQLi) & NoSQL Injection**

[SQL injection attacks](https://www.geeksforgeeks.org/sql-injection/) take advantage of weak web application queries by inserting malicious SQL code to modify database records, steal login credentials, or run admin-level commands.

NoSQL injection targets document-based databases like MongoDB and Firebase by altering query parameters, allowing attackers to bypass authentication and gain unauthorized access to sensitive data.

* 1. **Zero-Day Exploits & Advanced Persistent Threats (APT)**

[Zero-day exploits](https://www.geeksforgeeks.org/zero-day-exploit-cyber-security-attack/)take advantage of unknown software vulnerabilities before developers release security patches, making them highly dangerous.

Advanced Persistent Threats (APTs) use multi-stage attack techniques to stay hidden in a system for a long time. These include lateral movement (such as pass-the-hash attacks to spread across networks), privilege escalation (gaining higher access rights), and persistence mechanisms (like rootkits and fileless malware) to maintain control over compromised networks.

* 1. **Man-in-the-Middle (MITM) Attacks**

[Man-in-the-Middle (MITM) attacks](https://www.geeksforgeeks.org/how-to-prevent-man-in-the-middle-attack/) secretly intercept and modify data exchanged between two parties by exploiting weak encryption or unsecured communication channels.

Hackers use techniques like [SSL](https://www.geeksforgeeks.org/working-of-ssl/) stripping (removing HTTPS encryption), rogue access points (fake Wi-Fi networks), and ARP poisoning (manipulating network traffic) to steal login credentials, hijack user sessions, or inject malicious code into data transmissions.

* 1. **Insider Threats & Privilege Misuse**

[Insider threats](https://www.geeksforgeeks.org/what-is-insider-attack/) occur when compromised or malicious employees with high-level access steal sensitive data, modify access permissions, or install [backdoors](https://www.geeksforgeeks.org/what-is-a-backdoor-attack/) to bypass security. Since insiders already have authorized access, they can easily evade traditional perimeter security measures.

To detect and prevent insider threats, organizations use [User and Entity Behavior Analytics (UEBA)](https://www.geeksforgeeks.org/what-is-user-and-entity-behavior-analytics-ueba/) to track suspicious activities, [Privilege Access Management (PAM)](https://www.geeksforgeeks.org/privileged-access-management-pam-for-linux-and-unix/) to restrict sensitive access, and zero-trust architectures to ensure continuous verification.

**Cyber Security Tools**

Cybersecurity tools help protect systems, networks, and data from cyber threats by detecting, preventing, and responding to attacks. These tools play a crucial role in securing infrastructure, identifying vulnerabilities, and mitigating risks.

**1. Rise of AI and Machine Learning**: More cybersecurity tools are using artificial intelligence (AI) and machine learning to detect and respond to threats faster than humans can. [AI in cybersecurity](https://www.geeksforgeeks.org/ai-in-cybersecurity/) helps recognize patterns, block suspicious behavior, and even predict future threats—making it one of the most powerful tools to protect sensitive information.

**2. Increase in**[**Ransomware Attacks**](https://www.geeksforgeeks.org/rise-of-ransomware-attacks-during-covid-19-pandemic/): Ransomware, where hackers lock you out of your data until you pay a ransom, is becoming more common. Companies and individuals alike need to back up their data regularly and invest in security measures to avoid falling victim to these attacks.

**3. Cloud Security**: As more businesses move their data to the cloud, ensuring this data is secure is a top priority. This includes using strong authentication methods and regularly updating security protocols to protect against breaches.

**4. Internet of Things (IoT) Vulnerabilities**: With more devices connected to the internet, like smart home gadgets and wearable tech, there’s an increased risk of cyberattacks. Ensuring these devices have updated security features is crucial.

**5. Zero Trust Security:** This approach assumes that threats could come from inside or outside the network, so it constantly verifies and monitors all access requests. It’s becoming a standard practice to ensure a higher level of security.

**6. Cybersecurity Skills Gap:** There is a growing need for skilled cybersecurity professionals. As cyber threats become more sophisticated, the demand for experts who can protect against these threats is higher than ever.

**7. Regulatory Compliance:** New regulations are being introduced worldwide to protect personal data. Companies must stay informed about these laws to ensure they comply and avoid hefty fines.

Cybersecurity Best Practices

There are several steps you can take to protect yourself from cyber threats, including:

* Use strong passwords: Use unique and complex passwords for all of your accounts, and consider using a password manager to store and manage your passwords.
* Keep your software up to date: Keep your operating system, software applications, and security software up to date with the latest security patches and updates.
* Enable two-factor authentication: Enable two-factor authentication on all of your accounts to add an extra layer of security.
* Be aware of suspicious emails: Be cautious of unsolicited emails, particularly those that ask for personal or financial information or contain suspicious links or attachments.
* Educate yourself: Stay informed about the latest cybersecurity threats and best practices by reading cybersecurity blogs and attending cybersecurity training programs.

**Challenges of Cybersecurity**

* Constantly Evolving Threat Landscape: Cyber threats are constantly evolving, and attackers are becoming increasingly sophisticated. This makes it challenging for cybersecurity professionals to keep up with the latest threats and implement effective measures to protect against them.
* Lack of Skilled Professionals: There is a shortage of skilled cybersecurity professionals, which makes it difficult for organizations to find and hire qualified staff to manage their cybersecurity programs.
* Limited Budgets: Cybersecurity can be expensive, and many organizations have limited budgets to allocate toward cybersecurity initiatives. This can result in a lack of resources and infrastructure to effectively protect against cyber threats.
* Insider Threats: Insider threats can be just as damaging as external threats. Employees or contractors who have access to sensitive information can intentionally or unintentionally compromise [data security](https://www.geeksforgeeks.org/data-security/).
* Complexity of Technology: With the rise of [cloud computing](https://www.geeksforgeeks.org/cloud-computing/), IoT, and other technologies, the complexity of IT infrastructure has increased significantly. This complexity makes it challenging to identify and address vulnerabilities and implement effective cybersecurity measures.

**Strategies for Addressing Cybersecurity Challenges**

* Comprehensive Risk Assessment: A comprehensive [risk assessment](https://www.geeksforgeeks.org/risk-assessment/) can help organizations identify potential vulnerabilities and prioritize cybersecurity initiatives based on their impact and likelihood.
* Cybersecurity Training and Awareness: Cybersecurity training and awareness programs can help employees understand the risks and best practices for protecting against cyber threats.
* Collaboration and Information Sharing: Collaboration and information sharing between organizations, industries, and government agencies can help improve cybersecurity strategies and response to cyber threats.
* Cybersecurity Automation: Cybersecurity automation can help organizations identify and respond to threats in real time, reducing the risk of data breaches and other cyber-attacks.
* Continuous Monitoring: Continuous monitoring of IT infrastructure and data can help identify potential threats and vulnerabilities, allowing for proactive measures to be taken to prevent attacks.

**ML Security**

Machine Learning (ML) security focuses on safeguarding ML algorithms and systems from various risks and attacks, including adversarial attacks, data poisoning, and privacy breaches. It's a rapidly growing field crucial for ensuring the reliability, robustness, and trustworthiness of ML-powered applications, especially in cybersecurity where ML is used for threat detection and risk assessment.

Key aspects of ML security include:

* **Protecting ML models from attacks:**

Adversarial attacks involve deliberately manipulating input data to cause ML models to make incorrect predictions. Data poisoning involves introducing malicious data into the training set to corrupt the model.

* **Ensuring data privacy:**

ML models can be vulnerable to privacy attacks, where attackers can infer sensitive information about individuals from the data used to train the model or from the model's predictions.

* **Securing the ML development pipeline:**

This includes protecting the data used to train ML models, the code used to build and deploy them, and the infrastructure used to run them.

* **Implementing security measures throughout the ML lifecycle:**

This includes designing security into the model from the outset, testing for vulnerabilities, and monitoring the model's performance in production.

* **Using ML to defend against security threats:**

ML can be used to detect and prevent various security threats, such as cyberattacks and fraud.

Specific examples of ML in cybersecurity include:

* **Threat detection**: ML algorithms can be trained to identify suspicious network traffic, anomalous user behavior, or malicious code.
* **Fraud detection**: ML can be used to detect fraudulent transactions or other malicious activities.
* **Malware analysis**: ML models can be used to classify and analyze malware samples.
* **Network intrusion detection**: ML can be used to identify and prevent network intrusions.

In essence, ML security is about ensuring that ML systems are not only powerful and accurate but also safe and reliable, protecting them from various threats and vulnerabilities.

**Major Security Vulnerabilities and Risks in Machine Learning**

**1.Model poisoning:**

Model poisoning is a type of adversarial attack where an attacker manipulates a machine learning model's training data or parameters to cause it to behave in an undesirable way. This can be done by introducing biased or malicious data, altering model parameters, or exploiting vulnerabilities in the model's training process. The goal is to compromise the model's performance, security, or ethical behavior.

Attackers can directly modify the model's parameters or architecture to introduce vulnerabilities or backdoors. These backdoors can be triggered by specific inputs or prompts, causing the model to behave in an unintended way.

**Data Poisoning:**

Attackers can subtly alter the training data by mislabeling examples, adding false data points, or injecting malicious inputs. This can lead the model to learn incorrect patterns and make inaccurate predictions or generate harmful outputs.

**Model Theft**

Model theft, also known as model extraction, is a security threat in machine learning where an attacker gains unauthorized access to a trained model's parameters or architecture. This can involve querying the model, analyzing its output, and using this information to create a copy or exploit the model for various malicious purposes.

What is model theft? Model theft is a machine learning security threat that involves stealing a trained model's parameters or architecture. This can be done by querying the model and using the output to infer some of its parameters.

**Model Inversion**

In the context of machine learning, model inversion is a security threat where an attacker uses a model's output to infer details about its training data or architecture. Essentially, it's a type of attack where an attacker can "reverse engineer" a model to extract information from it, potentially revealing sensitive data used to train the model.

Suppose you train a machine learning model to predict whether a person has heart disease based on their medical history. Then, an attacker who doesn't have access to the person's medical history could use a model inversion attack to infer the person's medical history.

**Model Skewing**

Model skewing attacks occur when an attacker manipulates the distribution of the training data to cause the model to behave in an undesirable way.

Attackers modify the training data used by a machine learning algorithm, causing it to make a wrong decision.

**Different Attacks:**

**Input Manipulation attack:**

Input Manipulation Attacks is an umbrella term, which include Adversarial Attacks, a type of attack in which an attacker deliberately alters input data to mislead the model.

**Key Characteristics of Input Manipulation Attacks:**

* Subtlety: Changes are often too small for humans to detect.
* Targeted Impact: Designed to exploit specific vulnerabilities in a model.
* Broad Applicability: Can target various domains, from image recognition to cybersecurity systems.

**Data Privacy Attack**

Data privacy attacks, also known as data breaches, are cybersecurity incidents where sensitive information is accessed, stolen, or disclosed without authorization. These attacks can target individuals, organizations, or both, causing various harms like identity theft, financial losses, or reputational damage. Common types of attacks include phishing, malware, social engineering, and exploiting system vulnerabilities.

Types of Data Privacy Attacks:

* **Phishing:**

Cybercriminals use social engineering to trick individuals into revealing personal information, such as login credentials or credit card details, often through fake emails or messages.

* **Malware:**

Attackers inject malicious software (malware) into a victim's system to steal information or disrupt operations, often through malicious links, attachments, or websites.

* **Social Engineering:**

Attackers manipulate individuals or groups into compromising their security, often through deceptive tactics like impersonating trusted sources or creating a sense of urgency.

* **System Vulnerabilities:**

Cybercriminals exploit weaknesses in software, hardware, or network configurations to gain unauthorized access to systems and data.

* **SQL Injection:**

Attackers exploit vulnerabilities in databases to gain access to sensitive information or modify data.

* **Insider Threats:**

Unauthorized access or misuse of data by employees, contractors, or other individuals within an organization.

* **Ransomware:**

Attackers encrypt a victim's data and demand a ransom for its release, causing significant disruption and potential financial loss.

* **Denial-of-Service (DoS) Attacks:**

Attackers overwhelm a system with traffic to make it unavailable to legitimate users, often used to disrupt operations or extort payment.

* **Physical Security Compromises:**

Attackers may physically access data or devices to steal information, such as stealing laptops, hard drives, or paper documents.

* **Data Leakage:**

Unintentional or malicious disclosure of sensitive data, such as through email, cloud storage, or improperly secured devices.

**Data Poisoning Attack:**

Data poisoning, or AI poisoning, involves deliberately corrupting the training data of machine learning models to manipulate their behavior, resulting in biased or harmful outputs. Attackers inject malicious data to influence model decisions during the training phase, compromising its integrity and reliability.

Data poisoning is a type of cyberattack in which an adversary intentionally compromises a training dataset used by an AI or machine learning (ML) model to influence or manipulate the operation of that model.

**Data poisoning can be done in several ways:**

* Intentionally injecting false or misleading information within the training dataset
* Modifying the existing dataset
* Deleting a portion of the dataset

**Types of data poisoning**

Data poisoning attacks are typically classified based on the intended outcome of the attack. The two most common categories of data poisoning are:

1. Targeted data poisoning attacks: Targeted attacks occur when an adversary is attempting to manipulate the model’s behavior with respect to a specific situation. For example, a cybercriminal may train a cybersecurity tool to misidentify a specific file that they will use in a future attack or ignore suspicious activity from a certain user. Though targeted attacks can lead to serious and far-reaching consequences, they do not degrade the overall performance of an AI model.
2. Non-targeted data poisoning attacks: A non-targeted attack is when a cybercriminal manipulates the dataset to negatively impact the overall performance of the model. For example, the adversary may introduce false data, which in turn could reduce the accuracy of the model and negatively impact its predictive or decision-making capabilities**.**

**Adversarial Attack:**

An adversarial attack is a technique used to manipulate a machine learning model by introducing carefully crafted, often imperceptible changes to the input data, causing the model to make incorrect predictions or decisions. These "adversarial examples" are designed to exploit vulnerabilities in the model's decision-making process.

An adversarial attack is a method to generate adversarial examples. Hence, an adversarial example is an input to a machine learning model that is purposely designed to cause a model to make a mistake in its predictions despite resembling a valid input to a human.

How it Works:

* Input Perturbation:

Adversarial attacks involve adding small, carefully calculated "perturbations" or distortions to the input data.

* Deceptive Output:

These perturbations can be so subtle that they are not detectable by a human observer, but they can significantly alter the model's prediction.

* Exploiting Weaknesses:

The effectiveness of an adversarial attack depends on the model's sensitivity to changes in its input and its reliance on specific features for classification.

* Types of Attacks:
  + White-box attacks: Attackers have knowledge of the model's architecture and parameters, allowing them to craft more targeted attacks.
  + Black-box attacks: Attackers have limited or no knowledge of the model's internal workings and must rely on its output to generate adversarial examples.
* Goal of Adversarial Attacks:
  + Evasion: Deceiving the model into making incorrect predictions (misclassification).
  + Poisoning: Tampering with the training data to bias the model's behavior.
  + Information Extraction: Extracting information about the model's internal workings or decision boundaries.

**ML Supply Chain attack:**

In ML Supply Chain Attacks threat actors target the supply chain of ML models. This category is broad and important, as software supply chain in Machine Learning includes even more elements than in the case of classic software. It consists of specific elements such as MLOps platforms, data management platforms, model management software, model hubs and other specialized types of software that enable ML engineers to effectively test and deploy software.

**Transfer Learning**

Transfer learning attacks occur when an attacker trains a model on one task and then fine-tunes it on another task to cause it to behave in an undesirable way.

**Output integration attack**

In an Output Integrity Attack scenario, an attacker aims to modify or manipulate the output of a machine learning model in order to change its behavior or cause harm to the system it is used in.

**Deployment Vulnerabilities:**

Deployment vulnerabilities are weaknesses in the process of deploying applications or software that can be exploited by attackers. These vulnerabilities can arise from various sources, including misconfigurations, insecure dependencies, or flaws in the deployment process itself. They can lead to serious security risks, including data breaches, system downtime, and damage to reputation.

Types of Deployment Vulnerabilities:

* Configuration Errors:

Misconfigurations, such as default credentials or overly permissive access controls, can create vulnerabilities that attackers can exploit.

* Insecure Dependencies:

Using outdated or vulnerable third-party libraries or software components can introduce security flaws into the deployment.

* Code Tampering:

Unauthorized modifications to the code during the deployment process can lead to the introduction of malicious code or the loss of functionality.

* Vulnerable APIs:

API endpoints that are not properly secured can be exploited for unauthorized access or data breaches.

* Misconfigured CI/CD Pipelines:

Weaknesses in the continuous integration and continuous deployment (CI/CD) pipeline can allow attackers to compromise the deployment process.

* Shadow IT:

Using unvetted or unsupported tools and technologies can introduce vulnerabilities and security blind spots.

* Tool Sprawl:

The use of numerous tools and technologies in the deployment process can increase complexity and create security risks.

* Lack of Monitoring and Logging:

Insufficient monitoring and logging of deployment activities can make it difficult to detect and respond to security incidents.