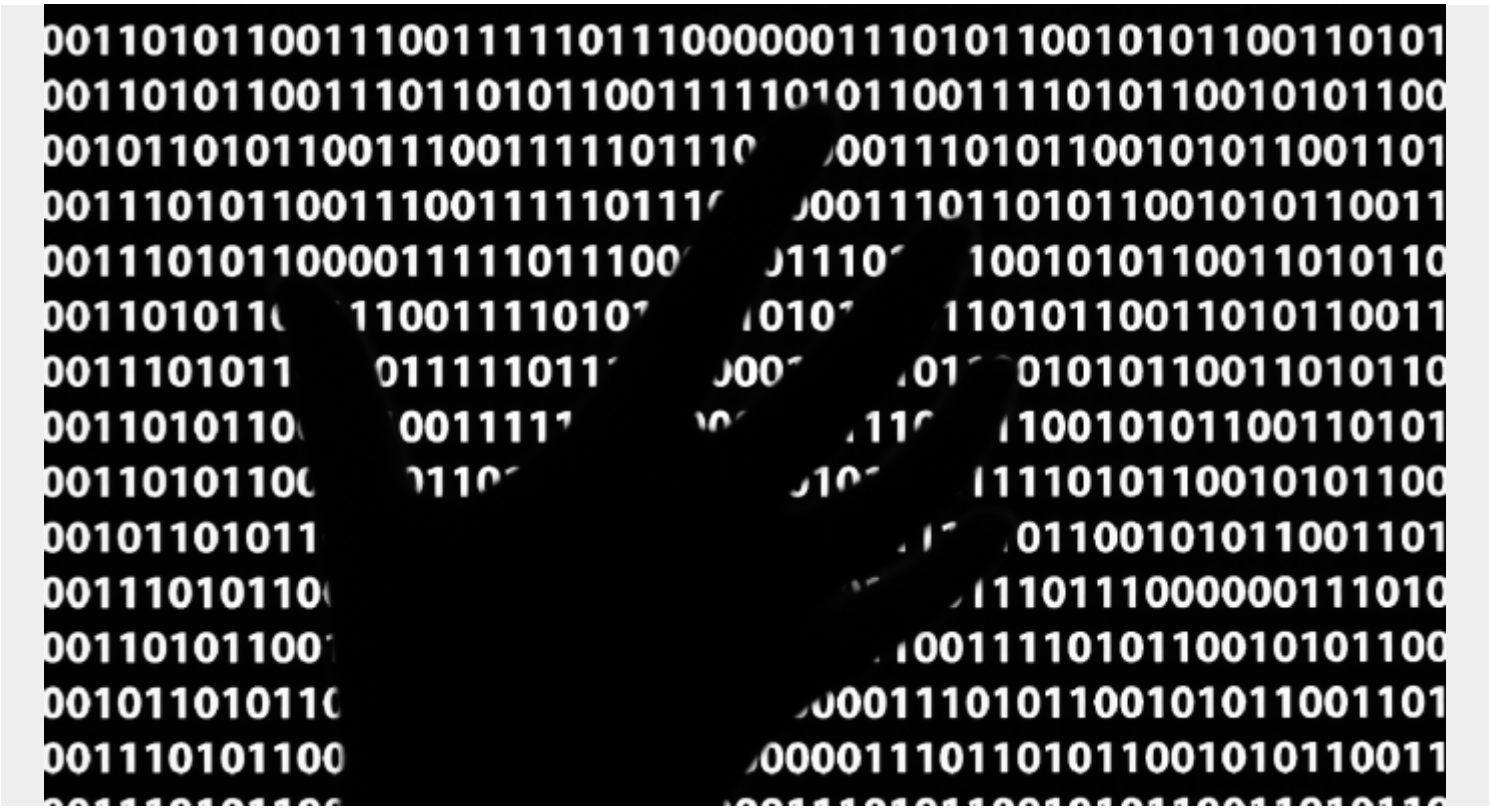
AI CYBERATTACKS & HOW THEY WORK, EXPLAINED



Artificial intelligence (AI) has created new possibilities for business organizations. The ability to automate and augment human intelligence has allowed organizations to:

- Transform operations
- Adopt new business models
- Understand customer behavior
- Predict cyberattacks

With these capabilities, organizations are able to adapt their operations and prepare for the challenges and opportunities—before they occur.

Unfortunately, artificial intelligence has also empowered cybercriminals. Taking advantage of sophisticated and intelligent technology solutions, they can:

- Find loopholes in corporate IT networks
- Launch large-scale Denial of Service (DoS) attacks
- Counter the limited security capabilities of an average organization.

Cyberattacks that harness AI might be the biggest threat facing organizations today—so let's take a look at how this changes the enterprise cybersecurity landscape.

Modern cyberattacks attack data

Until recently, many cyberattacks were intended to compromise networks and access sensitive information. These stats are beginning to tell a newer, bigger story:

- In the first six months of 2020, 36 billion data records were compromised.
- Only 5% of an organization's files are protected, on average.
- The average data breach costs $3.86 million in damages. That cost is bound to increase thanks to prevailing stringent regulations, like GDPR, that impose financial penalties on organizations failing to adopt the necessary security measures.
- It takes 200 days+ on average to first identify a security breach and up to 280 days to contain the damages.

Now consider the size of our digital universe, the big data that represents the opportunity for business organizations to leverage AI solutions in making well-informed decisions. We have already produced 44 zettabytes (44×1021) of data. By the year 2025, we expect to have generated 175 zettabytes of information.

So, what does this big data mean for attackers? The threat surface just got exponentially bigger. Consider these AI adoption trends:

- 58% of the respondents believe they have used AI in at least one of the business functions (McKinsey).
- The number of organizations adopting AI technologies has increased by more than 270% since 2015 (Gartner, 2019).
- More than 50% of the organizations reported improvement in productivity due to AI technology investments (PWC, 2018).
- The global AI market is expected to reach $327.5 billion by 2021, at a year on year growth of 16.4% (IDC).

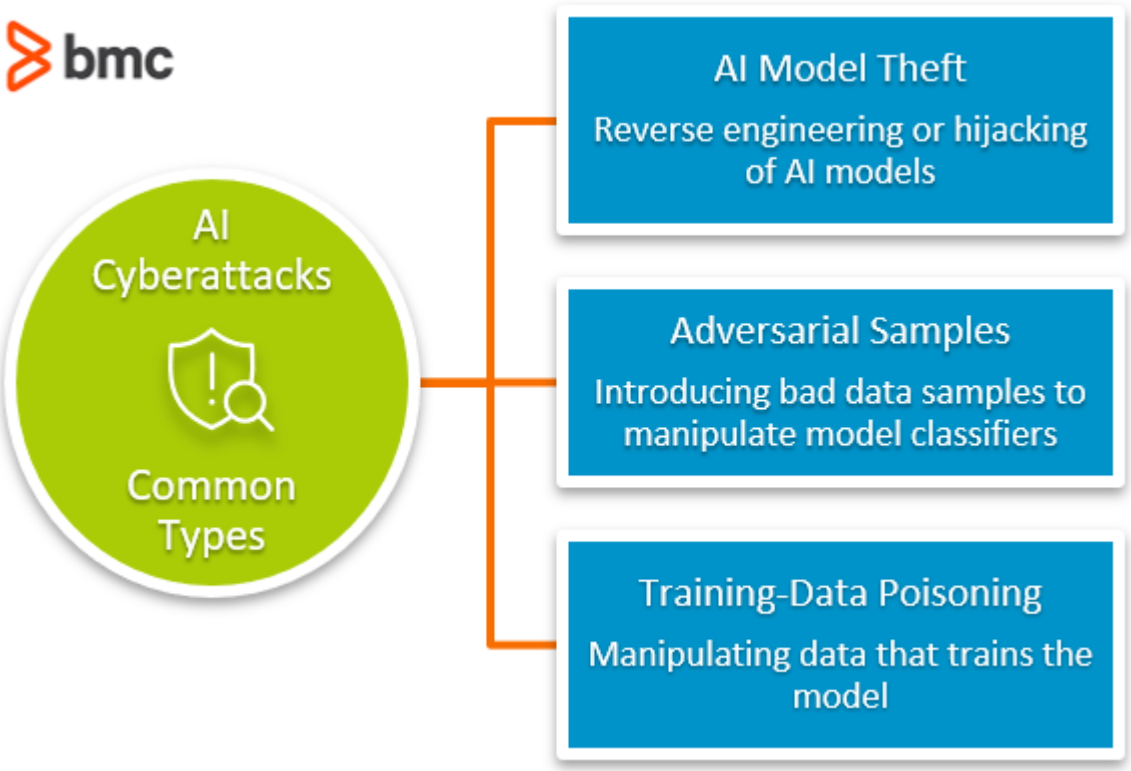
These trends have given rise to a new form of cyberattack threat vertical: AI cyberattacks.

What are AI cyberattacks?

AI Cyberattacks is the term for any offensive maneuvers launched on:

- AI systems
- Data
- The data processing pipeline

Since most AI practitioners excel at making sense of the available information, they are rarely the security experts who can protect their systems and data. Cybercriminals have found ways to compromise these systems, which has led to the concept of adversarial AI. This type of cyberattack jeopardizes the potential of data and AI systems to deliver the promised value to the business.



Common types of AI cyberattack

According to Gartner:

> *Through 2022, 30% of all AI cyberattacks will leverage training-data poisoning, AI model theft or adversarial samples to attack AI-powered systems.*

In an effort to understand and mitigate potential threats, let's look at these three AI-focused attack models.

AI model theft

AI model theft is the reverse engineering or hijacking of AI models. Once a model is trained and embedded on a vulnerable hardware chip or a cloud network, cybercriminals can:

- Access the AI systems
- Reverse engineer the machine learning (ML)/AI models

Confidential AI models are also being deployed on public networks accessible through API queries. Algorithms can also be reconstructed based on the data being ingested and delivered as an output from deployed models.

Adversarial samples

Adversarial samples are small sample instances that introduce feature perturbations, causing AI models to learn from the manipulated data and therefore learn to classify incorrectly. The samples are counterfactual, and the ML models fail to interpret them. As a result, the model becomes a source of incorrect classification decisions.

For example, consider a self-driving Tesla that is designed to slow down ahead of a stop sign. If the

stop sign is manipulated or painted in another color, the car may fail to recognize the sign.

Training-data poisoning

Training-data poisoning is the manipulation of training data that AI practitioners use to train the model.

Once cybercriminals gain unauthorized access to the storage network, they can alter the data or introduce significantly different data sets that are fed to the learning model, together with the original data.

Unlike the classical adversarial AI cyberattack that relies on stealing a pre-trained model, machine learning data poisoning targets only the training information fed to the model. Fewer highly skewed samples of input data are needed to manipulate the learning of the model itself—a benefit for the cybercriminals. Once an ML model is trained on the poisoned data, especially in the case of unsupervised learning models, you'll likely require deep AI expertise in order to identify possible issues with training data.

For instance, the accuracy and loss minimization of your model can change readily when a compromised model is tested against the uncorrupted datasets.

Planning for AI cyberattacks

Being aware of AI cyberattacks is the first step in preventing them. Security best practices encourage categorizing every potential threat into one function of the CIA triad, the most essential IT security concept. The MITRE ATT&CK Framework is a free resource that can also help inform your risk management practices.

Related reading

- BMC Security & Compliance Blog
- BMC Machine Learning & Big Data Blog
- Cybercrime Rising: 6 Steps To Prepare Your Business
- Introduction To Data Security
- IT Security Policy: Key Components & Best Practices for Every Business
- What's Artificial Artificial Intelligence? The Reality of AAI