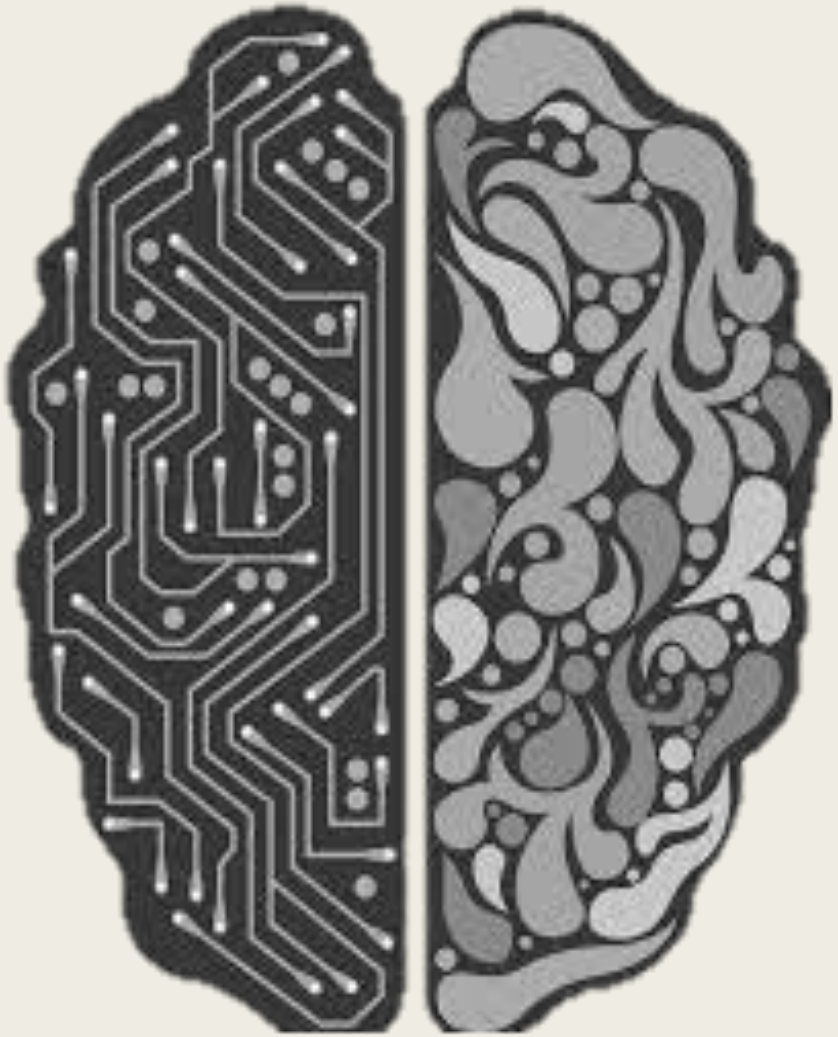
AI AND CYBERSECURITY: CHALLENGES IN EDUCATIONAL FIELD

S . Arnaud M. R. AHOUANDJINOU, Phd
Head of IT Security/ IFRI, UAC, Benin
4th Data Science School , IMSP, Benin 2023

Outline

- Introduction
- Motivation of attacks on Artificial Intelligence?
- Main Attacks on the AI
- Key points for a successful AI project
- Security measures at each stage of the AI Project: Think DevSecop
- Case of a new AI threat: DeepFake
- Conclusion

What is Artificial Intelligence?



- Artificial Intelligence can be defined as technology that demonstrates some form of basic intelligence, including the ability to learn and make decisions.
- Theoretically, advanced AI may even have the capacity to simulate and surpass the cognitive abilities of the human mind.
- This technology has been playing an increasingly large role in today's society.

How does Artificial Intelligence Work?

- The simple calculations and processes that occur in the human mind during every second of the day are actually much more complex than they initially seem, and developing a computer program to perform these same calculations takes time. Machine Learning is the term that is used to define how an AI “learns.”
- A technique known as genetic programming involves feeding an algorithm data and telling the algorithm what this data is.
- The algorithm interprets this data, and when it is asked questions about this data, developers determine if the results fall within a certain acceptable frame.
- Algorithms that succeed are built off and “reproduce,” while algorithms that fail are destroyed. This happens repeatedly until an algorithm emerges that can essentially understand and make decisions about the data that it is fed.[2]

Cybersecurity and Applications

- Cybersecurity is a field of study in the computer science industry that is dedicated to countering malicious attacks and exploits on networks. Having the proper countermeasures to these attacks are necessary, as the results of a successful attack can be catastrophic for the victim.

- AI technology has a promising future in the field of cybersecurity. Machine learning has been used to identify anomalous events on a network, which is often a signal of many types of attacks.[4] Because of its ability to detect these signs of an attack, AI can be incorporated into intrusion detection systems (IDSs), which monitor network traffic and detect malicious activity.[5]

Ethical Issues

- **Bias:** AI has the potential to be biased due to it needing to learn from information that is given to it. Therefore, it has the potential to inherit the biases of the human creators.3
- **More Advanced vs. Less Advanced:** As AI becomes more advanced, it becomes more integrated into our society. There are concerns on how AI would handle ethical dilemmas, and whether a more advanced AI would make more moral decisions over a less advanced AI..
- **Cybersecurity Issues:** AI may also pose some threats to the industries that they are implemented in. For example, in the cybersecurity field, an AI tasked with assigning users on a network certain privileges may automatically deny a user certain privileges based on biases that it may have.

IA in Cybersecurity

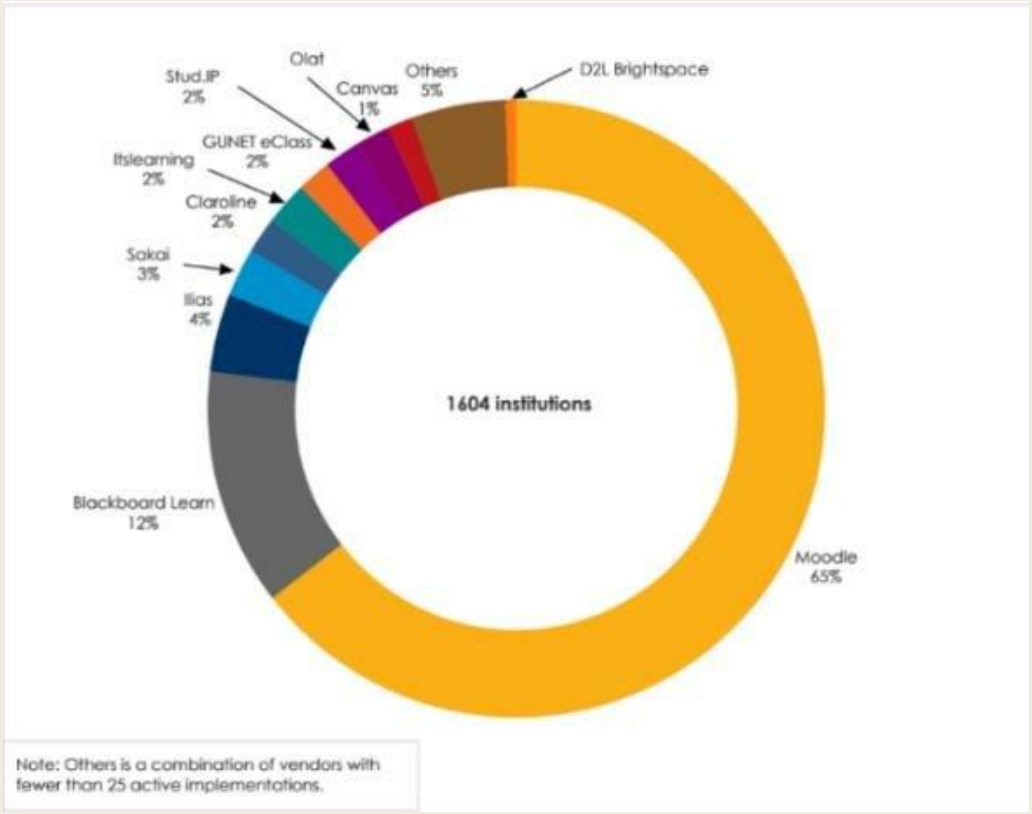
- Artificial Intelligence is undoubtedly becoming a large part of society today, and while this advanced technology introduces new issues, it also provides many benefits to a number of industries, including the cybersecurity field.
- It is important to note that this technology is still fairly new, and although there are ethical issues concerning the integration of AI into society, the technology continues to develop and improve.
- The benefits offered by AI in the cybersecurity field outweigh the controversies in, and since the industry of cybersecurity is ever-changing, AI is necessary to keep up with the changes.

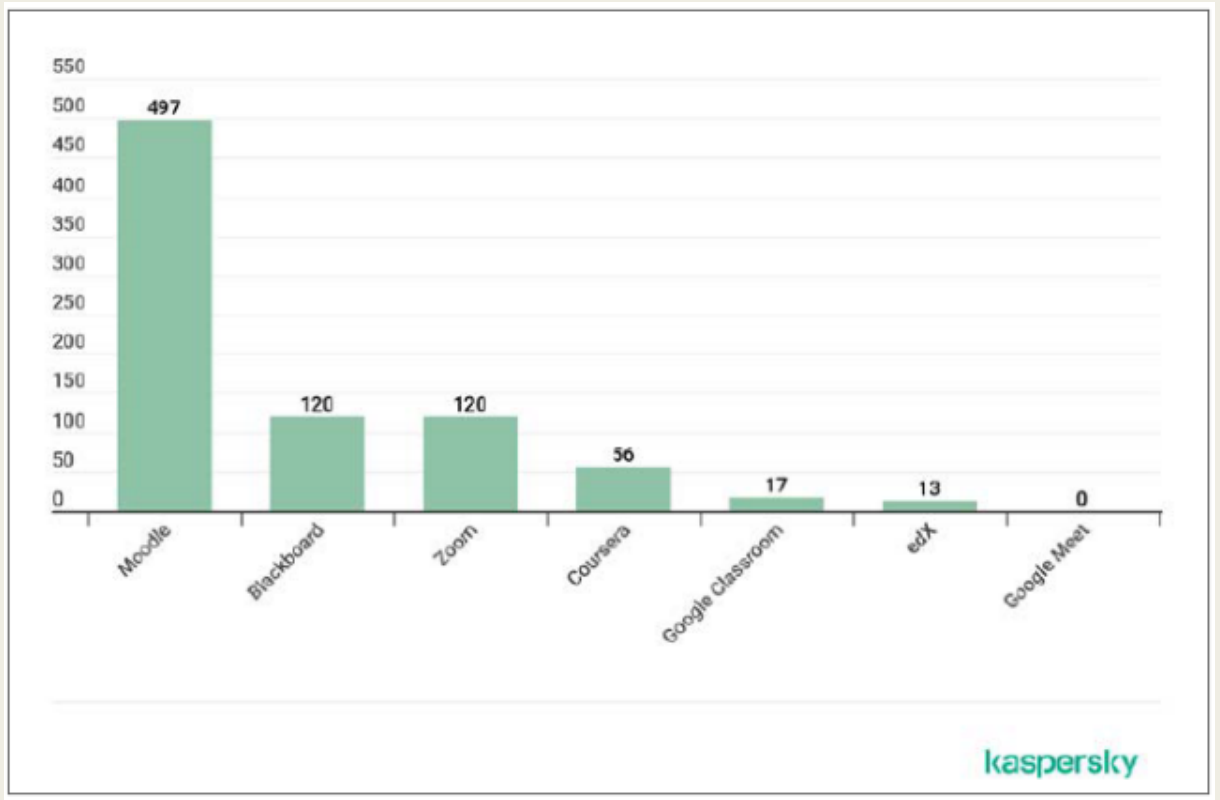
Cybersecurity in Education

- Cybersecurity refers to the protection of networks, devices, and data from unauthorized or unintended access or illegal use.
- The same bad actors that target enterprises also look for vulnerabilities in local school districts.
- Schools need enterprise-class security measures and hardware-enabled security to help protect their students, faculty, and data from cyberattacks.
- Improving cybersecurity is a top priority for organizations across all industries but is especially important in the education sector to help safeguard student and faculty privacy.

E-learning By LMS And Video conferencing Tools

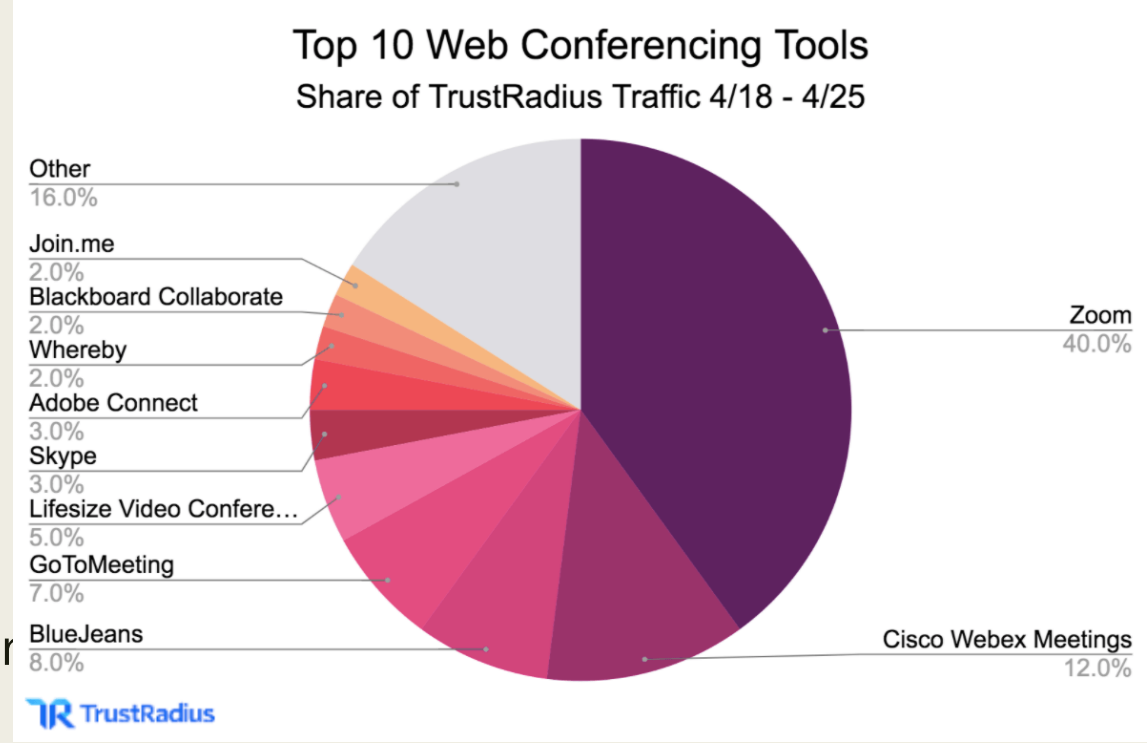
- LMSs were implemented until the pandemic, but during this period, were fully exploited.
- Complex courses were created, which had several types of activities, such as seminars, lessons, glossaries, practical tasks, assessment tests.
- According to the LMS Market report [8], in European HEIs, LMSs leaders are: Moodle (65%), Blackboard (12%), Ilias (4%) and Sakai (3%)





E-learning By LMS And Video conferencing Tools

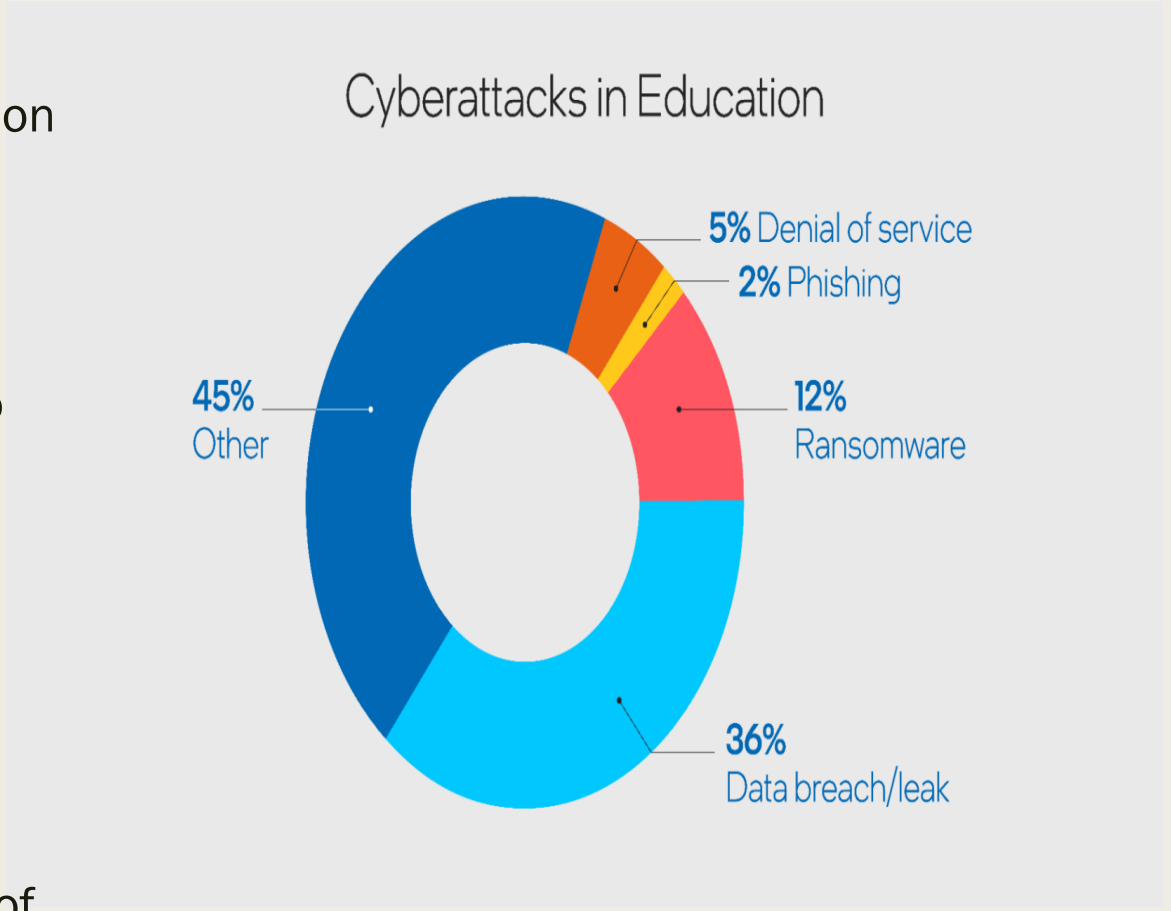
- VCT in these conditions, was the main source of communication.
- In this regard there are several applications, such as Zoom, GoToWebinar, Cisco WebEx, Livestorm, ON24, Adobe Connect, Microsoft Teams.
- Application concept is the same, different technologies used. According to the report submitted by the company Datanyze [9], the world leader in technography, top three VCA used in 2020, globally: Zoom, GoToWebinar ar Cisco Webex



Ranking	Technology	Domains	Market Share
1	Zoom	30583	36,15%
2	GoToWebinar	18486	21,85%
3	Cisco Webex	14628	17,29%

Mains Cyberattacks in Education : LMS case

- “Other” includes malware, meeting invasions, and website and social media defacement.
- These additional statistics only brush the surface on why cybersecurity is so important in education.
- One in three education devices contains sensitive data.[2]
- In a study of 5,400 IT decision-makers across 30 countries, education sectors are the most likely to admit security weaknesses.[3]
- 44% of IT managers in the education sector experienced a ransomware attack. This is the highest level of attack compared to a variety of other industries such as healthcare, IT, and local government.[3]
- 87% of educational establishments have experienced at least one attack.[4]
- Among all industries, the education sector is one of the least secure, and schools are the second most lucrative target for ransomware.[4]



E-learning Security Threats

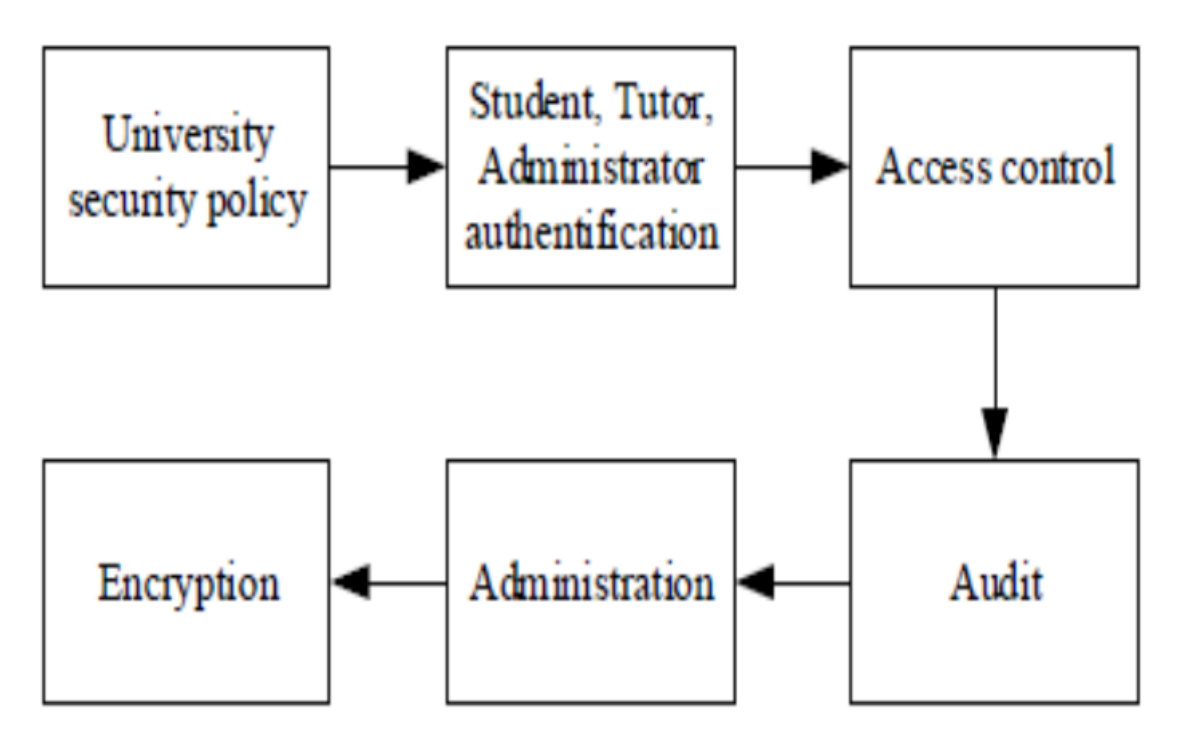
- **Confidentiality violation :** An unauthorized party gaining access of the assets present in E-Learning system.
- **Integrity Violation :** An unauthorized party accessing and tempering with an asset used in E-Learning system.
- **Denial of Service :** Prevention of legitimate access rights by disrupting traffic during the transaction among the users of E-Learning system.
- **Illegitimate use :** Exploitation of privileges by legitimate users.
- **Malicious program :** Lines of code to damage the other programs.
- **Repudiation :** Persons denial of participation in any transaction of documents.
- **Masquerade:** A way of behaving that hides the truth by the hackers.
- **Traffic analysis:** Leakage of information by abusing communication channel.
- **Brute-force attack:** An attempt with all possible combinations to uncover the correct one

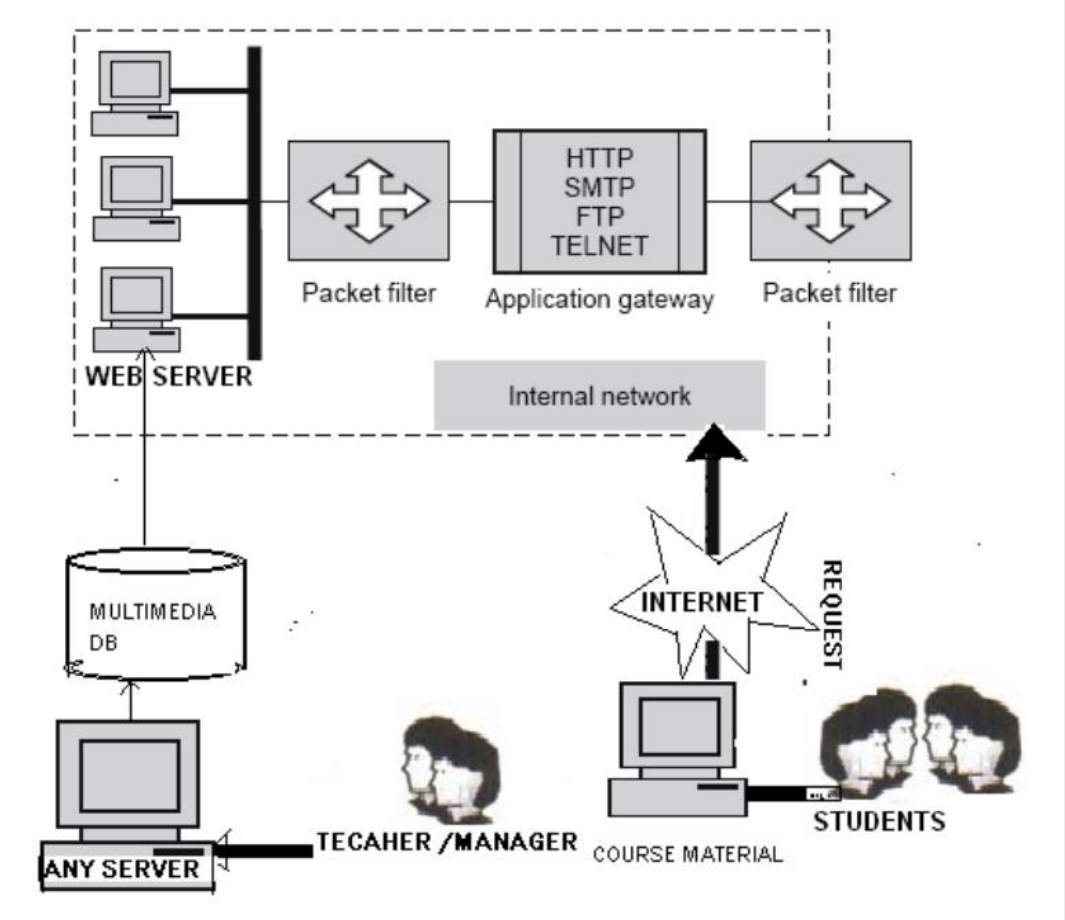
E-learning Security Risk Management

- Author's risk
- Teacher's risk
- Manager's risk.
- System Developer's risk
- Student's risk
- Others threats and risks in E-Learning :
  - ***Natural threats :*** *Natural threats may be caused by natural disasters like fire, storm, volcanic eruption, earthquake, floods*
  - ***Deliberative act :*** *Threats may come from fraud, blackmail, theft etc.*
  - ***Unintended :*** *There may be some unavoidable threats like Computer bug, power outage, handling error etc.*

- Identify the risk
- Analysis of the risk
- Provided a treatment of the risk
- Monitor the risk

REMEDIES OF RISKS : Access control using Firewall





Components of satisfactory security in e-learning

Access control using Firewall

Main Attacks on AI Apps in Education(1)

- **Hijacking the AI application:**
  - *by deliberately provoking an erroneous decision of the application with a chosen dataset. The circumvention of a facial recognition system to obtain logical or physical unauthorized access and carry out a theft is an example*
- **Sabotage the operation of AI:**
  - *preventing or disrupting the operation of the application. The attacker aims for a deterioration of the brand image or a slowdown in the activities of the targeted company. The attack on chatbotMicrosoft Tay in 2016, presented in the following section of the focus, is an emblematic example of sabotage.*
- **Understand and "reverse engineer" the model:**
  - *by studying his behavior. The work of data processing and modeling is often time-consuming and expensive for companies and the result with high added value. «Stealing» and then reselling a model can be very lucrative and buyers will be present to save time in the perpetual race for digital innovation..*

Main Attacks on AI Apps In Education(2)

- **Steal the data used by the application:**
  - *by querying them directly or cross-referencing the results provided by the application and trying to draw conclusions from them, or even steal the databases to which the AI has access.*
- **How to attack Artificial Intelligence:**
  - *Attacks specifically affecting machine learning-based applications can be grouped into three categories:*
    - Poisoning, or how to move the center of gravity of the AI
    - Inference, or how to make AI speak
    - Escape, or how to fool AI

Key points to Success Safety your project based AI in Education(1)

- **Protect data at every stage of the project and the BigData platform:**
  - *Data is the foundation of machine learning projects in companies. These projects require the manipulation of a large volume and a wide variety of data. Even before talking about protection against information leaks, it is essential to ensure that the desired use complies with the regulations in force, and in particular those related to data protection (GDPR, Health Data Host, PCI-DSS...).*
- **Securing the learning process:**
  - *Machine learning is both the key step in making the solution effective and relevant, and the real novelty compared to existing systems. We quickly understand that it can be a prime target for attackers. It is therefore appropriate to dedicate a special reflection to the protection of this stage. This protection must be at two levels: at the level of training data and at the level of the learning method.*

Key points to Success Safety your project based AI in Education(2)

- **Secure the application:**
  - *Many of the attack attempts can be contained by applying the secure development best practices already widely deployed in enterprises (for example, the rules of Open Web Application Security Project, or OWASP). However, these are not enough to protect against all cases of fraud related to the use of Machine Learning. The main safety measures specific to Machine Learning focus on three aspects: controlling its inputs, making processing more reliable and controlling its outputs*
- **Define its risk management and resilience strategy:**
  - *AI brings together a wide variety of applications; all do not have the sensitivity level of the autonomous car. Let's take the example of a chatbotintelligent: the level of risk in the case of a passive chatbotde, intended to provide personalized advice in response to questions such as "How much is my glasses reimbursed?" , will be less than that of a chatbottransactionnel, capable of performing operations such as the online creation of a third-party paying card. It is therefore also essential here to conduct a risk analysis in order to know where to prioritize efforts. These risk analyses must take into account the specific risks associated with the use of Machine Learning.*

New Threat for AI Apps in Education : DeepFake

- **Deepfake :**
  - *The Deepfakedis the modification of images, audios or videos via Artificial Intelligence (and in particular the « DeepLearning») to present a falsified vision («fake») of reality.*
- **Operation:**
  - *A Deepfakeest file created using two competing Artificial Intelligence systems: one is called the generator and the other the discriminater. The generator creates a fake file (image, audio, video, etc.) and then asks the discriminator to determine whether the file is real or fake. Together, the generator and the discriminator form what is called a Generative Adversarial Network (GAN). A learning dataset is provided to the generator to initialize the process. Then the method works in two-stage cycles:*

New Threat for AI Apps in Education : DeepFake

- ✓ *1. Discriminator Training: The discriminator is trained to differentiate between real files and fake files created by the generator.*
- ✓ *2. Generator Training: The generator creates files, which are evaluated by the discriminator. This allows the generator to improve by producing more and more credible files (judged as real) for the discriminater*

- **Results:**
  - *Once the generator has sufficiently progressed in the credibility of the created files, the discriminator is again trained to differentiate real files from fake files created by the generator (new step «1»).*
  - *The re-trained discriminater is now used to evolve the generator as part of a new step «2».*
  - *This cycle is repeated as many times as necessary to achieve the desired level of accuracy*

DeepFake Types

- Some types of DeepFake :
  - ***Deepfake audios ,*** *which imitate the voice of a targeted person from samples of his voice. They allow him to pronounce a given text as an input.*
  - ***Face-swapping,*** *which replaces in a video the face of a person with that of a targeted person from his photo.*
  - ***Deepfake lip-synching , which in a video adapts the movements of the face of a targeted person from an audio file of another person. It allows her to deliver the speech contained in the audio file in question without her having delivered it.***
  - ***Deepfake puppetry ,*** *which generates a video of a targeted person from an actor video provided as input. It is thus possible to create a video in which the target reproduces a speech played by the actor. A famous example of this technique is a video of Barack Obama stating a speech simulated by Jordan Peele, in which the gesture of the former president is reproduced in an extremely realistic way*

Conclusion

- ***The boom of Machine Learning leads to both increasingly powerful algorithms, for*** *applications, created for an entertainment purpose, represent a new powerful tool, accessible and easy to use for malicious people, There is no doubt that attacks using such applications will increase in the coming years.*
- *Fraud against the president, damage to the image, creation of false legal evidence, circumvention of biometric authentication: the cases of possible uses are numerous and frightening.*
- *Prevention solutions, such as the creation of anti-Deepfakeapplicated filters to the media before their publication. They introduce into the image a noise imperceptible to the naked eye but disturbing the learning of Deepfake algorithms (similar to the contradictory examples presented earlier).*
- *Solutions offer innovative mechanisms to guarantee the authenticity of files, like Amber Authenticate, a blockchain-based solution*