William Adoukonou

University College Birmingham  [Company address]

Artificial Intelligence in cyber security

# Task 1 - Introduction

*(IBM, 2023)* describes AI as technology that “enables computers and machines to simulate human learning, comprehension, problem solving.” This writing will explain the main techniques of AI such as machine learning, deep learning, and anomaly detection and show examples of each. To show these technique’s effectiveness there will also be a summarized, real-world example where AI has been used successfully. This will be based on the Helios by UVeye. This report will also go over the challenges that come with adopting this kind of technology such as data false positives, adversarial attacks, and privacy concerns. It will also go over issues such as AI over-reliance. For a look at the future of AI, there will be a section on upcoming trends such as AI in IoT security alongside a summary at the end.

Contents

[Task 1 - Introduction 1](#_Toc184724519)

[Task 2 - Define key AI techniques 3](#_Toc184724520)

[Machine Learning 4](#_Toc184724521)

[Deeper Learning 4](#_Toc184724522)

[AI for Phishing 4](#_Toc184724523)

[AI for DDoS 5](#_Toc184724524)

[AI for Malware 5](#_Toc184724525)

[Anomaly Detection 5](#_Toc184724526)

[Task 3- AI Real World Case Study 6](#_Toc184724527)

[How does it work 6](#_Toc184724528)

[Impact on Cyber Security Industry 6](#_Toc184724529)

[Task 4 – Challenges and Limitations in AI 7](#_Toc184724530)

[False Positive 7](#_Toc184724531)

[Adversarial Attacks 7](#_Toc184724532)

[Privacy Concerns 8](#_Toc184724533)

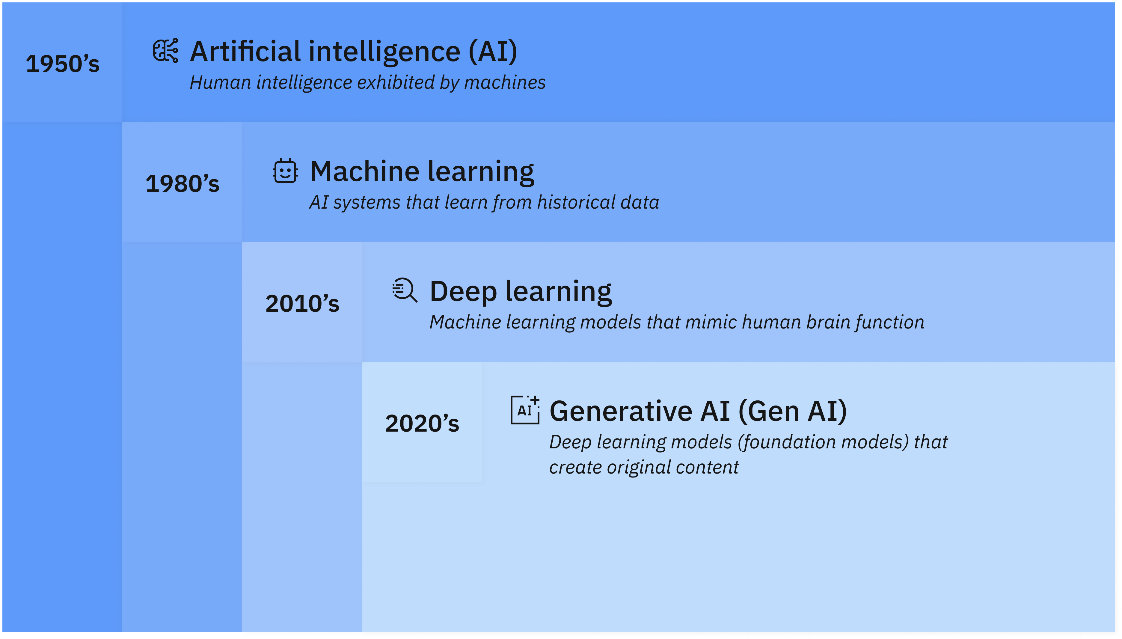
[Task 5 – Future trends in AI Cyber Security (AI for IoT) 8](#_Toc184724534)

[Pattern Recognition 8](#_Toc184724535)

[Environmentally Friendly 9](#_Toc184724536)

[Conclusion 9](#_Toc184724537)

# Task 2 - Define key AI techniques

AI or Artificial Intelligence is a way to “simulate human learning” and capabilities within computers *(IBM, 2024)*. This enables computers to understand creativity and problem solving. Throughout time, the concept of AI has developed into a learning model that goes deeper and deeper, allowing machines to “think” with less human assistance. This image from IBM shows the progression of machine learning across 70 years.

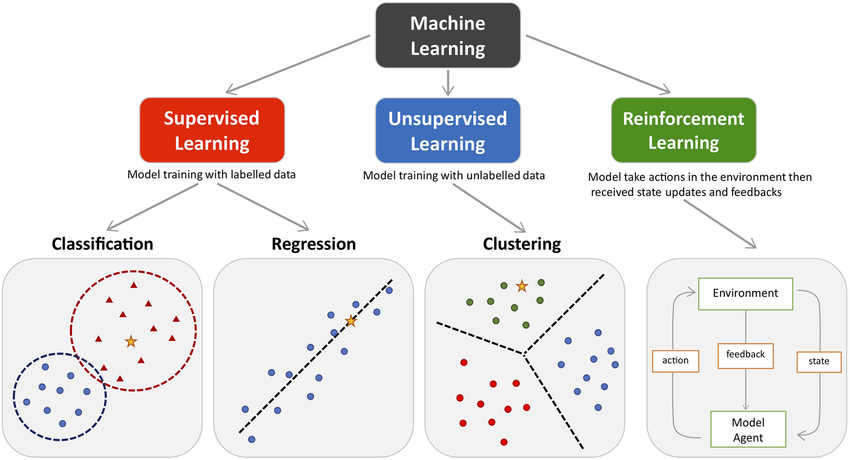
*(Figure 1, AI over time. IBM, 2024)*

## Machine Learning

This is simply just machines learning from information that was already available. Technology such as the internet is incredibly beneficial as there is an almost infinite amount of data that we can use to teach machines. One of the most common cases of machine learning is in image detection. Let's say you want a machine to be able to tell you what is in an image, the program will be taught what objects are and then will have the ability to distinguish what is on the screen based on what it has learned in its training phase. Some of its real-world applications are:

Facial Recognition – machines can learn to recognize faces as a security measure

Email Automation – machine learning can be used to learn which spam mail to filter out



*Example of how data can be sorted to train a machine. ResearchGate, 2024*

## Deeper Learning

AI goes deeper than machines mimicking human learning, it also includes deeper learning. This is a subset of machine learning that goes into more detail. It's a Later developed AI model that more precisely copies human thinking. Just like the brain, it uses “artificial neural networks to mimic learning processes” *(Coursera, 2024).* An example is PyTorch developed by Facebook AI Research. It's known for its dynamic computational graph. Deeper learning is better than machine learning due to its increased precision in machine learning tasks such as “advanced speech recognition, language processing, and decision-making.” *(Grammarly, 2024)*

Generative AI – A newer AI model that can create its own content based on its own intelligence. Adds a response to AI as well as comprehension. One example is Midjourney, a text-to-image model that can create artistic images from text descriptions. *(datacamp, 2023)*

## AI for Phishing

How does this work for cyber security then? Take virus and threat protection for example. Social engineering attacks such as Phishing emails are not easy for a computer to 100% spot and filter so it would be beneficial if these antiviruses were trained using artificial intelligence. The software could “detect anomalies and spot new strategies” to keep businesses secure. *(Barracuda, 2023)*. It would do this by reading the email, scanning for human errors and inconsistencies, then comparing its results to previous phishing attacks.

## AI for DDoS

When any data traffic goes through a network, it is sent in packets. The aim of a DDOS attack is to flood it with traffic by using “fake” packets. This drastically slows down performance on the site and if it is more than the network can handle, it will crash. This doesn’t seem very harmless for the average person as it can be solved with a reboot after a certain amount of time but imagine a big company with customers who rely on their services for day-to-day activities. The goal for minimizing the impact of a DDoS attack is to respond early, this can be achieved with the help of AI.

## AI for Malware

AI can be used for malware similarly to its uses in DDoS prevention. This is done by embedding it into an antivirus. Antiviruses work by performing system scans on every directory, searching for any files that seem out of place. If a file is found, it alerts the user and gives them the option to “quarantine” and then delete the file. It is then placed in a folder where it cannot harm the system. With machine learning, it can learn the patterns that malware takes when infecting systems and spot an intrusion earlier. It can also adapt to where they like to hide and which naming schemes they like to use and search those places 1st.

## Anomaly Detection

When machine learning is used as a preventive measure, it's imperative that the machine knows how to detect serious outliers. This is called anomaly detection. These inconsistencies come in the form of “unusual patterns, random spikes, or any data point that doesn't fit with most of the data collected prior.” *(IBM, 2023).* Anomaly detection’s main use cases are virus protection and fraud prevention. Imagine a bank wants to be able to spot fraudulent activity on someone’s account as fast as possible. The machine will look for these certain features:

Velocity – is there an increase in transaction amounts per day?

Volume – are there large, unusual transactions outside of normal spending habits?

Login Anomalies – are there unusual login times, locations, or devices? *(Fraud.com, 2023)*

Password Anomalies – have there been frequent password resets or changes to security questions? *(Proofpoint, 2023)*

When any of these are triggered, it will create an anomaly in the stored data, leading the machine to confirm fraud has occurred. *(Fingerprint, 2024)*

# Task 3- AI Real World Case Study

The UVeye system is a prime example of AI-powered threat detection working in the real world. UVeye is a company from Israel that created a system that can detect suspicious items implemented into automobiles. It can spot car bombs or other items by using artificial intelligence.

## How does it work

They have a new product, Helios. By using high-resolution cameras, it creates 3D images of the vehicle's undercarriage. It then uses AI algorithms to analyze this image, looking for items like Improvised explosive devices (IEDs) and other suspicious items. This would also be beneficial in more secure places and vehicles such as aircraft in the airport. The Helios can be used as an anti-terror attack device which would increase customer safety at airports and train stations. Its benefits include speed and accuracy. AI algorithms can detect subtle anomalies that might be missed by human eyes, also “at a much faster rate.” *(IIOT World, 2024).* Having the inspections be done by a computer also reduces the need for manual labor, speeding up security and increasing efficiency.

<https://www.youtube.com/watch?v=BDE8oHQHsCs>

## Impact on Cyber Security Industry

This piece of technology acts as a precursor to AI cyber defenses. “The blueprint for AI scanning hardware can be used in different professions” *(Acronyms, 2024).* This is a similar method used in AI systems used to monitor network traffic and manage databases.

Some of the challenges this might face come from the false positives the AI can be fed. Like other models, the Helios device has the chance to be manipulated and provide inaccurate results. Hackers have the potential to feed it false information which throws off its detection rate. This type of device would need robust defenses in place to prevent this. Another problem this type of device might encounter is the privacy concerns. AI-powered systems may “collect and process large amounts of sensitive data, raising concerns” since users may not know how much of their vehicle is visible to the device and its operators. *(TaylorWessing, 2024)*

# Task 4 – Challenges and Limitations in AI

The problem with training a machine to perform like a human is that there will always be errors. As mentioned earlier, false positives can create challenges as a machine in its earlier stages of learning will be inaccurate.

## False Positive

An example of a common false positive is in AI plagiarism checkers. For example, Turnitin’s software boasts a “less than 1% false positive rate” *(Turnitin, 2023)*. Because of this 1%, machines must lean towards either “trust or not to trust”. In this specific case a plagiarism checker would assume positive intent. This means it would side with the student if there were definitive proof of detection. In other areas like antivirus, the machine would decide between safe or suspicious and should choose to quarantine the file in most cases. The place false positives have in machine learning can slow down the progress of AI. This is because people are less likely to trust implementations of AI in society if they can generate incorrect results. Imagine AI had replaced a service such as facial recognition. If false positive rates were higher, it could lead to the wrong people receiving fines or punishments, damaging its reputation.

## Adversarial Attacks

A stop sign with a couple of signs

Description automatically generatedIf someone had more malicious intentions, they could use an Adversarial attack. This is when a user intentionally “tricks a model into making predictions that are wrong” *(ITM, 2024)*. This is achieved by making small edits in the input data. These changes are barely recognizable by humans however they can cause an AI model to predict completely wrong.

*Figure 2, AI model tricked by adversarial attack. ITM, 2024*

This is what we call an Evasion Attack. The image shows how small alterations caused the machine accuracy to drop to less than 1%. If this software was in a vehicle, it would be dangerous. It wouldn’t stop at the sign since it hasn’t recognized it.

## Privacy Concerns

There are many benefits of AI but like all emerging technologies, there are significant privacy concerns. The main issue stems from sensitive data collection. AI models like anomaly detection “require the collection and analysis of sensitive personal information” *(Experian. 2022).* This means user data such as financial information may end up in a database and its up to the company to keep this data secure. People don’t like when big tech companies have large amounts of their data.

To further add to this point, another issue is data retention. Big AI companies are taking massive amounts of user data and storing it forever. Therefore, companies now must disclose what they plan on doing with your data in line with the GDPR (2018).

Another problem in artificial intelligence is bias and discrimination. Imagine there is an AI model that is being trained to determine if you should hire someone. If the hiring algorithm is trained on data that disproportionately favors certain groups, it could carry on this trend when it is released. *(WTWCO, 2023).* This is unethical and will cause people to lose trust in AI since it’s been pre-programmed unfairly.

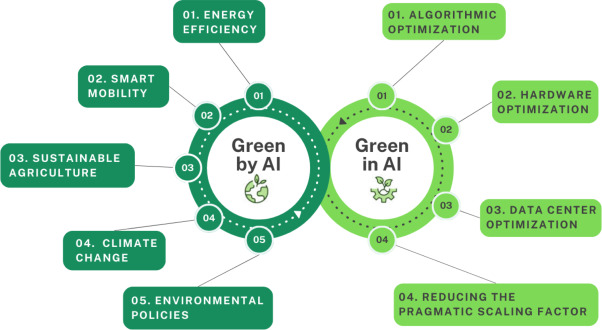
# Task 5 – Future trends in AI Cyber Security (AI for IoT)

IoT or internet of things is what we call a network of devices that have smaller scaled technology that communicates with each other as well as exchanges data over the internet *(WMM, 2023*). This includes sensors, software, and other devices that can be found in household objects or even major tools used in the cyber industry. *(Oracle, 2020).* This technology is growing due to its massive potential with artificial intelligence.

## Pattern Recognition

First of all, it can take advantage of AI pattern recognition capabilities. In a house, a smart thermostat will have a sensor that is part of an IoT, AI can learn the owner’s patterns and automatically adjust the temperature even if they aren’t there. Adding a backup layer of safety in homes. In the industry, AI can then use its pattern recognition tools to prevent attacks on IoT devices like CCTV cameras. Even technology as small as a door sensor can be victim to an IoT attack. *(PTC, 2024).* This can be even used in the health care industry. Patients with biometric organs need 24/7 remote monitoring to make sure the devices don’t fail. These devices can be manufactured with pattern recognition tools to “monitor vital signs, track patient data, and improve remote healthcare.” *(ByteBeam, 2023)*

## Environmentally Friendly

IoT devices are used in the eco industry to help monitor plants and the overall environment. With the power of AI, these devices can consume less energy and be more efficient. One example is in the agricultural industry. The IoT devices that monitor weather conditions and crop health can use AI predictive technologies to accurately trigger other devices such as sprinklers to keep vegetation healthy while using the least number of resources *(PeerBits, 2023)*. When these items are shipped and transported to different countries, IoT sensors in the vehicle’s GPS can determine the shortest routes to the destination based on traffic habits allowing for quicker travel and less emissions, keeping the environment safer.

*Figure 3, AI solutions for the environment. ScienceDirect, 2024*

# Conclusion

In summary, Artificial Intelligence is defined by its main techniques: machine learning, deeper learning, and generative intelligence. Machine learning allows us to “feed” computers information in order to train to recognize things. Deeper Learning is a subset of machine learning that has a more advanced ability to operate more human-like. Lastly, we have generative AI. It’s a newer, more self-sufficient subset that can create its own content based on what it’s learned from machine/deeper learning.

These AI techniques can be used to strengthen cyber security. It can combat attacks such as Phishing DDoS and Malware attacks by learning the patterns that standard viruses, network attacks and scam emails use to prevent them from infecting the system. These inconsistencies it finds in normal system behavior is called Anomaly Detection

These technologies aren’t a thing of the future. We are already seeing them be implemented into the real world. The UVeye Helios is an example of how machines can learn by being fed examples of automobile tampering as well as image recognition of car bombs to protect vehicles from these attacks.

This doesn’t mean all AI technologies are 100% safe and accurate though. There are some impacts and challenges that come with the adoption of AI technology in the real world. The main one in machine learning is false positives. This is when a machine predicts an incorrect answer as correct. This can have worse consequences in other scenarios such as a plagiarism checker where a used may be penalized for the machine’s mistake. Another issue is adversarial attacks. This is when users intentionally feed a machine incorrect information, causing errors in the predictions. This can cause privacy concerns such as AI bias/discrimination and data retention policies which would make the public not trust in AI technology as much.

One newer adaptation of AI in cyber security is AI for IoT (internet of things). Cyber security IoT devices such as a CCTV network can take advantage of AI pattern recognition to learn and prevent IoT attacks on them. A generalized application of this technology is used in GPS devices for smart vehicles. To reduced emissions, GPS’s can use AI to predict faster routes to burn fuel for the least amount of time, helping the environment.

*References*

andrewwoods. “Fraud Detection Using Machine Learning and AI.” *Experian UK*, 15 May 2022, www.experian.co.uk/blogs/latest-thinking/guide/machine-learning-ai-fraud-detection/.

“Automatic Vehicle Inspection during COVID-19 - UVeye.” *UVeye*, 23 Mar. 2020, www.uveye.com/automatic-vehicle-inspection-during-covid-19/. Accessed 30 Oct. 2024.

Barnard, Joel, and Cole Stryker. “What Is Anomaly Detection? | IBM.” *Www.ibm.com*, 12 Dec. 2023, www.ibm.com/topics/anomaly-detection.

Barracuda INC. “Barracuda.” *Barracuda.com*, 2023, www.barracuda.com/.

Bhattacharyya , Atreyee. “Beware Data Bias in AI Models.” *WTW*, 23 Oct. 2023, www.wtwco.com/en-nl/insights/2023/10/beware-data-bias-in-ai-models.

Bolón-Canedo, Verónica , et al. “A Review of Green Artificial Intelligence: Towards a More Sustainable Future.” *Neurocomputing*, vol. 599, no. 128096, 1 June 2024, pp. 1872–8286, https://doi.org/10.1016/j.neucom.2024.128096.

---. “A Review of Green Artificial Intelligence: Towards a More Sustainable Future.” *Neurocomputing*, vol. 599, no. 31, 1 June 2024, pp. 128096–128096, https://doi.org/10.1016/j.neucom.2024.128096.

Chalishazar, Tej. “Exploring the Applications of IoT in Different Industries.” *Peerbits*, 25 Apr. 2023, www.peerbits.com/blog/iot-applications-in-different-industries.html.

Chea, Evelyn. “Fingerprint.” *Fingerprint*, 2024, fingerprint.com/blog/account-takeover-solutions/. Accessed 24 Nov. 2024.

Chechitelli, Annie. “Understanding False Positives within Our AI Writing Detection Capabilities.” *Www.turnitin.com*, 16 Mar. 2023, www.turnitin.com/blog/understanding-false-positives-within-our-ai-writing-detection-capabilities.

Grammarly. “Deep Learning: Everything You Should Know.” *Deep Learning: Everything You Should Know | Grammarly*, 13 June 2024, www.grammarly.com/blog/ai/what-is-deep-learning/. Accessed 24 Nov. 2024.

International, Fraudcom. “Transaction Monitoring - Everything You Need to Know.” *Fraud.com*, 21 Apr. 2023, www.fraud.com/post/transaction-monitoring.

ITM. “Protect IoT-Applications from Adversarial Evasion Attacks – ITM-Predictive.” *Itm-P.com*, 2024, www.itm-p.com/protect-iot-applications-from-adversarial-evasion-attacks. Accessed 12 Nov. 2024.

Lloyd-Davies, Frazer. “AI in Cybersecurity: Revolutionising Threat Detection and Prevention.” *Acronyms Ltd*, 2 Apr. 2024, www.acronyms.co.uk/blog/ai-in-cybersecurity/.

Moffa, Anthony. “The Key to Using AI to Enhance IoT Security.” *Ptc.com*, 2024, www.ptc.com/en/blogs/iiot/using-ai-to-enhance-iot-security. Accessed 7 Dec. 2024.

Oracle. “Accelerate Your Operations with IOT.” *Oracle.com*, 2020, www.oracle.com/uk/internet-of-things/.

Proofpoint. “What Is Account Takeover Fraud (ATO)? Prevention & More | Proofpoint UK.” *Proofpoint*, 10 July 2023, www.proofpoint.com/uk/threat-reference/account-takeover-fraud.

Pykes, Kurtis. “How to Use Midjourney: A Comprehensive Guide to AI-Generated Artwork Creation.” *Datacamp.com*, DataCamp, 27 July 2023, www.datacamp.com/tutorial/how-to-use-midjourney-a-comprehensive-guide-to-ai-generated-artwork-creation. Accessed 30 Oct. 2024.

Staff, Coursera. “Deep Learning vs Machine Learning: A Beginner’s Guide.” *Coursera*, 2024, www.coursera.org/gb/articles/ai-vs-deep-learning-vs-machine-learning-beginners-guide.

Stryker, Cole , and Eda Kavlakoglu. “What Is Artificial Intelligence (AI)?,” *Https://Www.ibm.com/Topics/Artificial-Intelligence*, 16 Aug. 2024.

Stryker, Cole, and Eda Kavlakoglu. “What Is Artificial Intelligence (AI)?” *IBM*, IBM, 16 Aug. 2024, www.ibm.com/topics/artificial-intelligence.

---. “What Is Artificial Intelligence (AI)?” *IBM*, IBM, 16 Aug. 2024, www.ibm.com/topics/artificial-intelligence.

Team Bytebeam. “IoT Innovations | Top 33 Industrial IoT Device Examples.” *Bytebeam*, Bytebeam, 28 Aug. 2023, bytebeam.io/blog/top-industrial-iot-device-examples/.

V, Andrew, and er. “The Rise of Internet of Things (IoT): Balancing Benefits and Responsibilities in a Connected World.” *International Journal of Swarm Intelligence and Evolutionary Computation*, vol. 12, no. 4, 7 Aug. 2023, pp. 1–1, www.walshmedicalmedia.com/open-access/the-rise-of-internet-of-things-iot-balancing-benefits-and-responsibilities-in-a-connected-world-122275.html, https://doi.org/10.35248/2090-4908.23.12.322. Accessed 7 Dec. 2024.

Viorel, Paul. “An Introduction to AI Anomaly Detection.” *IIoT World*, 2 Feb. 2024, www.iiot-world.com/artificial-intelligence-ml/artificial-intelligence/an-introduction-to-ai-anomaly-detection/.

Yates, Michael, and Andi Terziu. “AI – the Threats It Poses to Reputation, Privacy and Cyber Security, and Some Practical Solutions to Combating Those Threats.” *Www.taylorwessing.com*, 22 Apr. 2024, www.taylorwessing.com/en/global-data-hub/2024/cyber-security---weathering-the-cyber-storms/ai---the-threats-it-poses-to-reputation.