From Supervised Learning into Generative AI: new perspectives in Cybersecurity

Webinar: “Innovative Cyber Investigations with AI”

Mário Marques da Silva

Professor at Universidade Autónoma de Lisboa

Director of the Department of Engineering and Computer Sciences

Researcher at Instituto de Telecomunicações

[mmsilva@autonoma.pt]

Fundação para a Ciência e a Tecnologia

Agenda

1. Introduction

2. Uses of AI in Cybersecurity

3. Conclusions

Agenda

1. Introduction

2. Uses of AI in Cybersecurity

3. Conclusions

1. Introduction

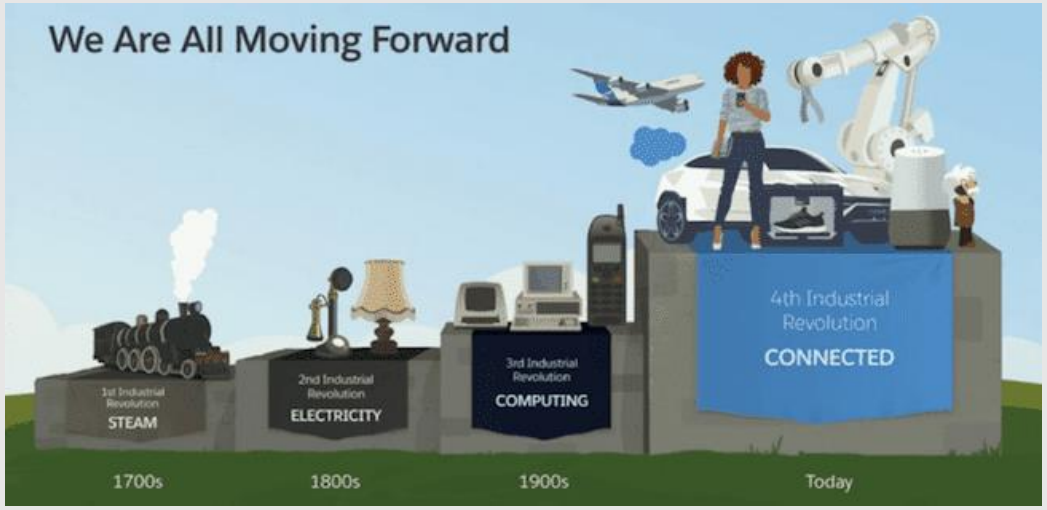
What is “Digital Transformation”?

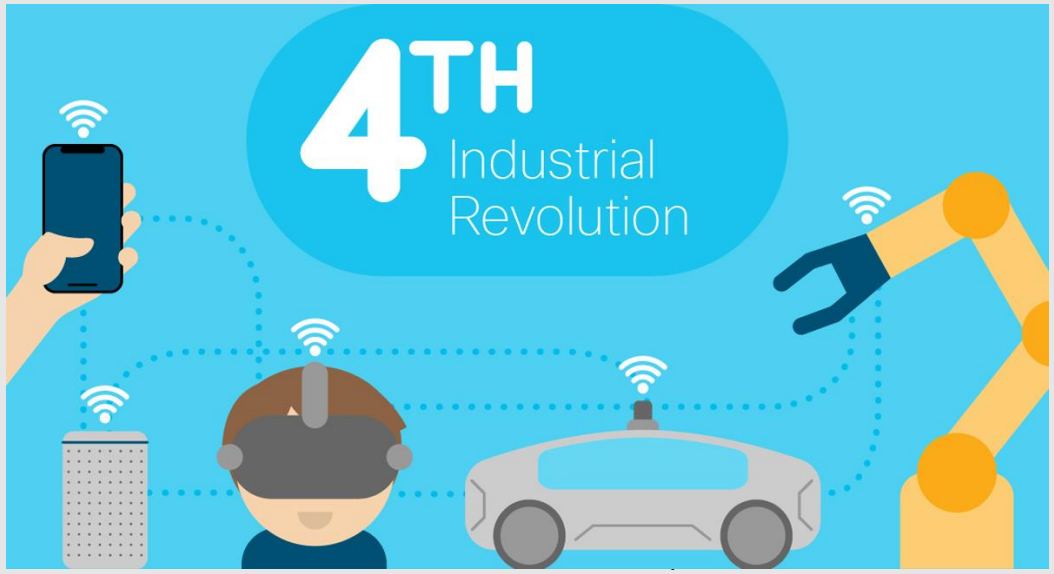
- Analog Electronics (Fourties)
- Phase 1: Digital Electronics (Eighties)
- Phase 2: All-over-IP

3rd Industrial Revolution

Information Age

- **Phase 3: Automation** || 4th Industrial Revolution - *Knowledge Age*
  - *ChatGPT versus static information*





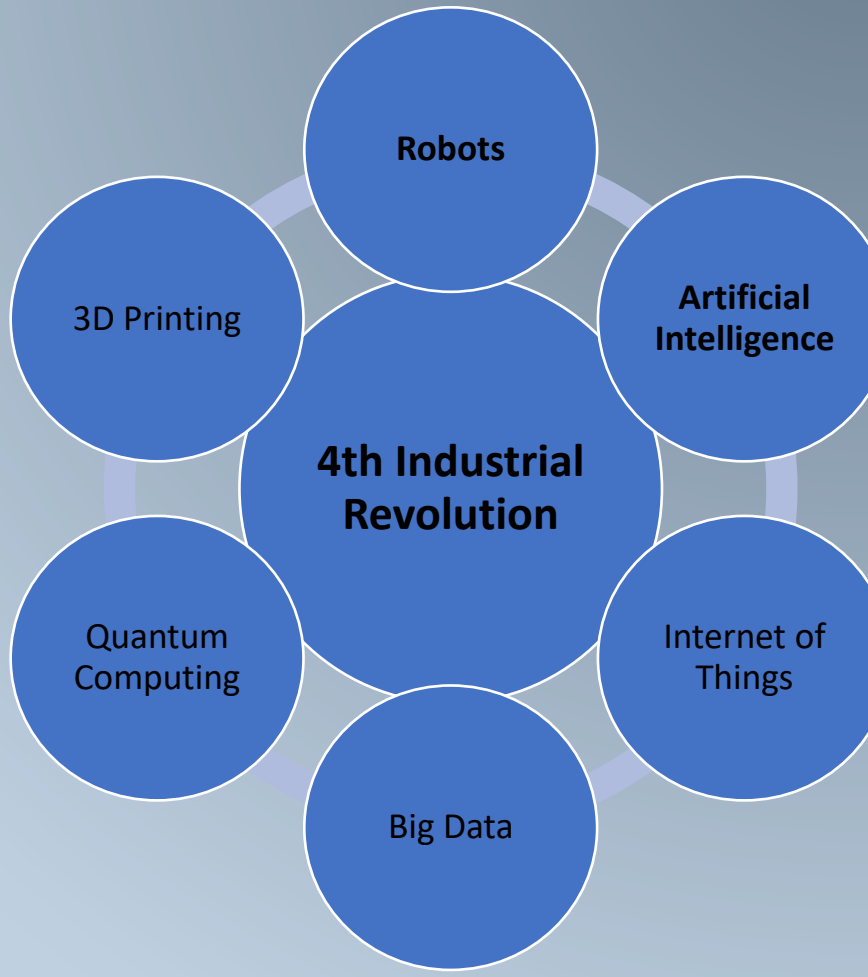
1. Introduction

The **4th Industrial Revolution** is characterized by the **massive use of Robots**, as well as **Artificial Intelligence, Big Data, Internet of Things, 3D Printing, Quantum Computing**.

More **efficient use of the Resources**.

These technologies potentiate the **replacement of Humans by Robots** in diferent areas.

- **Some Jobs disappear, others are created**, requiring a great human **adaptation** to this new **Paradigm**.
- **Knowledge Age – routine decisions are implemented by machines.**
  - Recalibration of skills / educational frameworks: **Scientific, technical and human skills**, critical thinking, analytic thinking, social and adaptability skills, emotional intelligence, and abstract thinking, became more important assets than tangible goods.
  - **Corporates are becoming capital-intensive**.



1. Introduction

Artificial Intelligence

What is Machine Learning?

Field of Study that gives computers the ability to learn without being explicitly programmed [Arthur Samuel, 1959]

Voice-to-text conversion

Image recognition

Forecasting of prices

Medical Applications (e.g., diagnose diseases from X-rays)

Autonomous Driving (Reinforcement Learning)

Recommender Systems (Netflix, Spotify, etc.).

Marketing

Robots

1. Introduction

- **Cybersecurity** is the practice of **protecting systems, networks, and programs** from digital attacks.

- With the **increase of the digital** world (**4th Industrial Revolution, IoT**), the digital attacks are **increasing** rapidly and becoming **more sophisticated**, requiring new cybersecurity approaches.

- **Rapid Detection** is not compatible with **Human Incident Response** → Need **Automation and Intelligence**.

- **Machine Learning: Mostly used in Cybersecurity** because it is **suitable for processing large datasets in real-time, to detect deviations quickly**.
- **Generative AI**: Can **simulate various attack scenarios** to help test and improve security defenses but is **less commonly used for real-time detection**.

Agenda

1. Introduction

2. Uses of AI in Cybersecurity

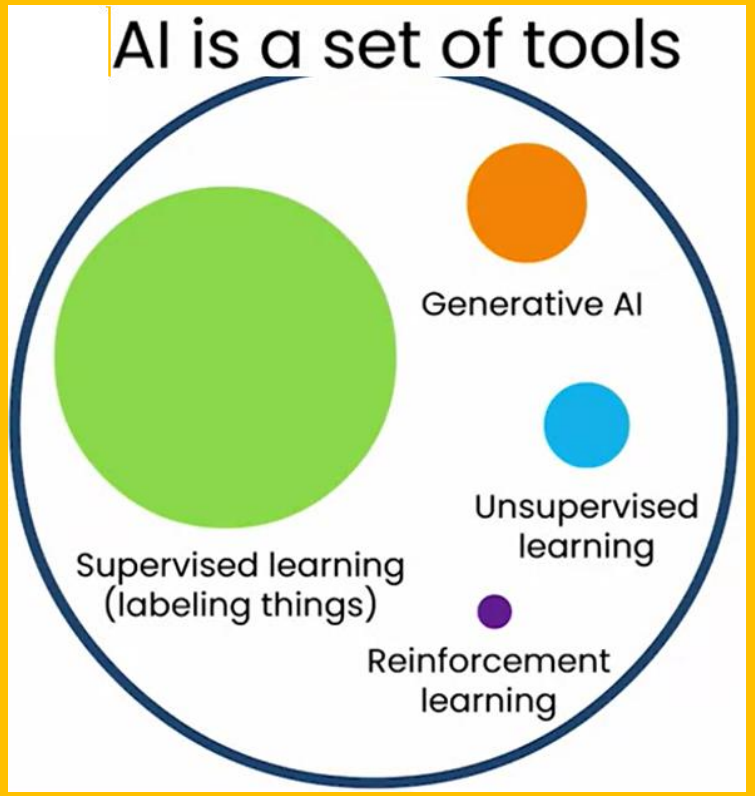
3. Conclusions

Use Machine Learning in Cyberdefense

- Suitable for processing **large datasets in real-time analytics, to detect deviations quickly** (e.g. unusual login times or locations, detecting unauthorized access attempts in real-time).
- Automates tasks like **log analysis and threat classification**.
- Learn from past incidents to classify data, **suggesting optimal responses (reduces human errors and response time)**.
- Provides **speed** and **precision** in analyzing millions of events per second, **reducing false positives.**
  - **Supervised Learning**: effective in **recognizing** known **attack patterns** (ex: block 99% of ransomware attacks by learning typical user behaviors).
  - **Unsupervised Learning**: effective in **classifying data in clusters** (e.g., K-means algorithm) or **detecting anomalies** (ex: to find a security breach before it caused damage).

2. Uses of AI in Cybersecurity

Supervised Learning with <u>one Neuron</u>



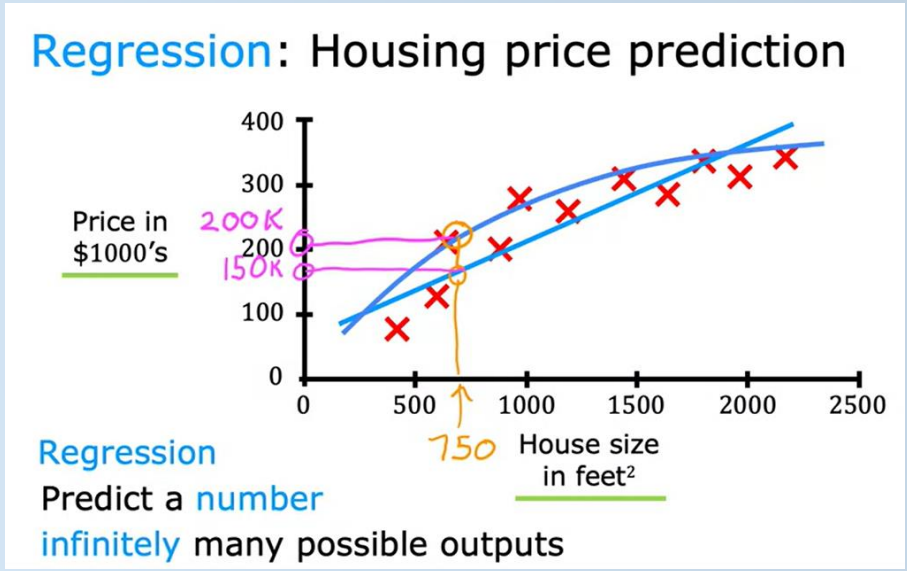
Supervised Learning: after training with examples (data set), such as images of dogs and cats or previous prices of houses, it learns **classifying data** or to **predict values** [regression] (ex: value of a house).

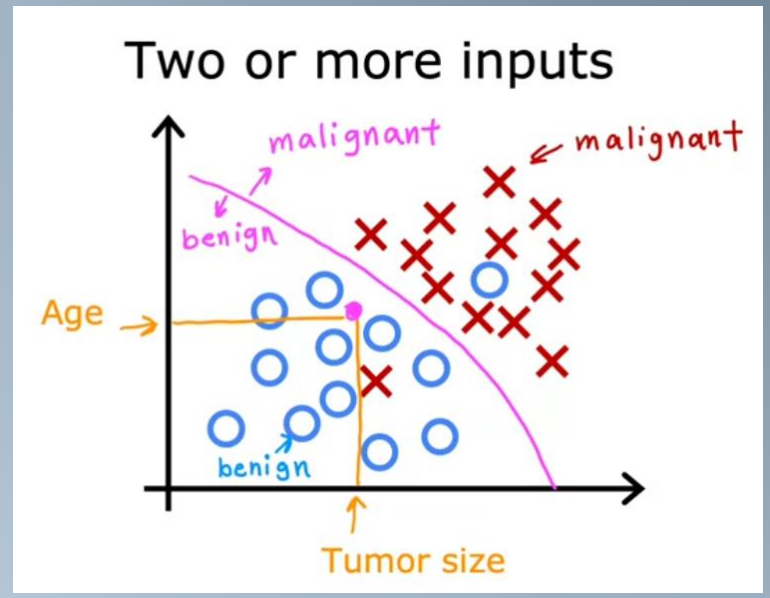
2. Uses of AI in Cybersecurity

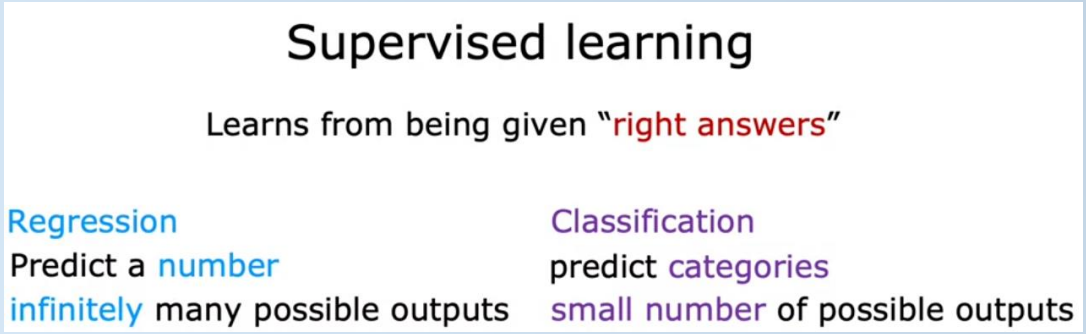
Main Types of **Supervised** Learning: (1) Regressions and (2) Classification (Logistic Regression)

Data Set: Learn from previous samples.

After learning, forecast new results.







Typical Features in Cybersecurity:

- Bandwidth
- Time of the day
- Login location
- User profile

2. Uses of AI in Cybersecurity

Supervised Learning

Training set: Data used to train the model

size in feet²	price in $1000's
2104	400
1416	232
1534	315
852	178
...	...
3210	870

2. Uses of AI in Cybersecurity

Supervised Learning

Terminology

Training set: Data used to train the model

	x size in feet²	y price in $1000's
(1)	2104	400
(2)	1416	232
(3)	1534	315
(4)	852	178
...	...	...
(47)	3210	870

$m = 47$

$x^{(1)} = 2104 \quad y^{(1)} = 400$

$(x^{(1)}, y^{(1)}) = (2104, 400)$

Notation:

x = "input" variable
feature

y = "output" variable
"target" variable

m = number of training examples

(x, y) = single training example

$(x^{(i)}, y^{(i)})$

$(x^{(i)}, y^{(i)})$ = i^{th} training example

(1^{st}, 2^{nd}, 3^{rd} ...)

CLASSIFICATION (Logistic Regression)

Question	Answer "y"	
Is this email spam?	no	yes
Is the transaction fraudulent?	no	yes
Is the tumor malignant?	no	yes

y can only be one of two values

"binary classification"

class = category

false true

0 1

useful for classification

"negative class"

≠ "bad"

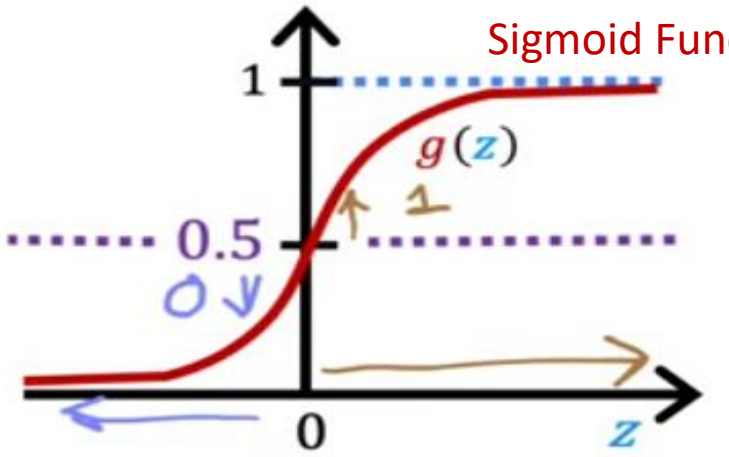
absence

"positive class"

≠ "good"

presence

Decision Boundary



$f_{\vec{w},b}(\vec{x})$

$z = \vec{w} \cdot \vec{x} + b$

z

$g(z) = \frac{1}{1+e^{-z}}$

$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w} \cdot \vec{x} + b}_{z}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$$

Probability

$= P(y = 1 | x; \vec{w}, b)$ 0.7 0.3

0 or 1?

Is $f_{\vec{w},b}(\vec{x}) \geq \underbrace{0.5}_{\text{threshold}}$?

Yes: $\hat{y} = 1$ No: $\hat{y} = 0$

When is $f_{\vec{w},b}(\vec{x}) \geq 0.5$?

$g(z) \geq 0.5$

$z \geq 0$

$\vec{w} \cdot \vec{x} + b \geq 0$ $\vec{w} \cdot \vec{x} + b < 0$

$\hat{y} = 1$ $\hat{y} = 0$

Logistic Regression

Gradient descent

cost

$$J(\vec{w}, b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) + (1-y^{(i)})\log\left(1-f_{\vec{w},b}(\vec{x}^{(i)})\right)\right]$$

repeat {

Repeat for each different feature

$j = 1 \ldots n$

$$w_j = w_j - \alpha\frac{\partial}{\partial w_j}J(\vec{w}, b)$$

$$\frac{\partial}{\partial w_j}J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m}\left[(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})x_j^{(i)}\right]$$

$$b = b - \alpha\frac{\partial}{\partial b}J(\vec{w}, b)$$

$$\frac{\partial}{\partial b}J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m}(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})$$

} simultaneous updates

2. Uses of AI in Cybersecurity

Logistic Regression - Gradient Descent

How to adjust weights 'w' of features such that the Cost [Error] is minimized?

Gradient descent for logistic regression

repeat {

looks like linear regression!

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) \right]$$

} simultaneous updates

Same concepts:

- Monitor gradient descent (learning curve)
- Vectorized implementation
- Feature scaling

Linear regression $f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

Logistic regression $f_{\vec{w},b}(\vec{x}) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$

2. Uses of AI in Cybersecurity

Supervised Learning with Multiple Neurons

Neural Networks (Deep Learning)

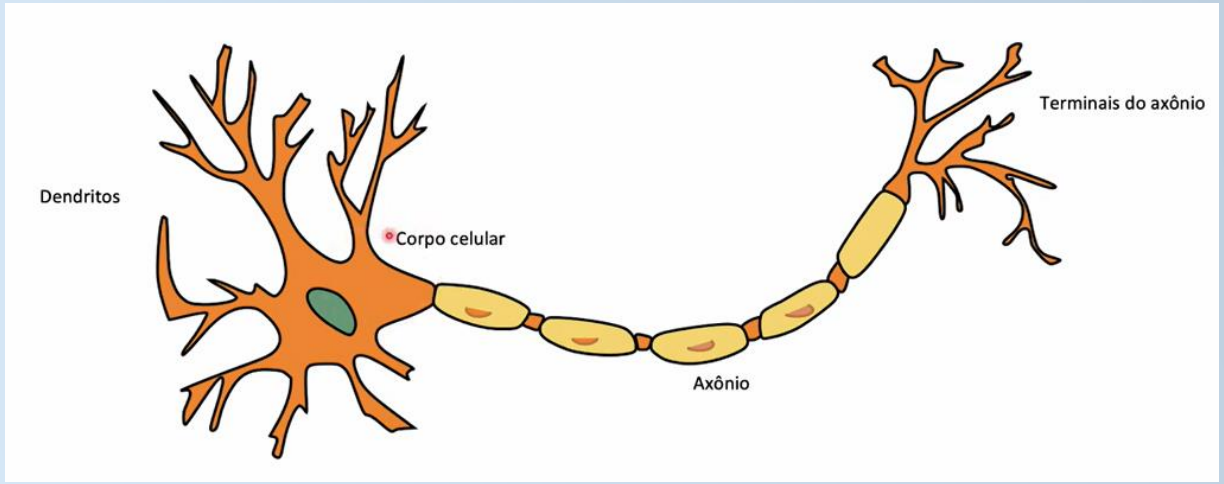
- **Deep Learning:** uses **several layers of artificial neurons**, to learn patterns in data (make predictions).
- Artificial neurons are connected through synapses, which are the electrical equivalents of biological synapses.

Use Deep Learning in Cyberdefense

- Effective to **identify unusual behavior** in **complex data** (e.g., **biometrics**, or **image recognition** in security recordings or **deeper analysis of network traffic**).

2. Uses of AI in Cybersecurity

Network with 1 Neuron

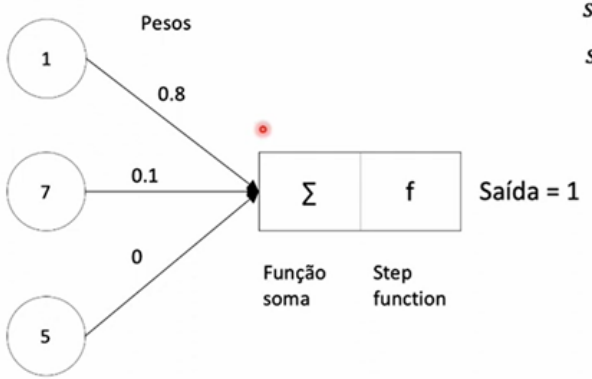


$$soma = \sum_{i=1}^{n} xi * wi$$

$soma = (1 * 0.8) + (7 * 0.1) + (5 * 0)$

$soma = 0.8 + 0.7 + 0$

$soma = 1.5$



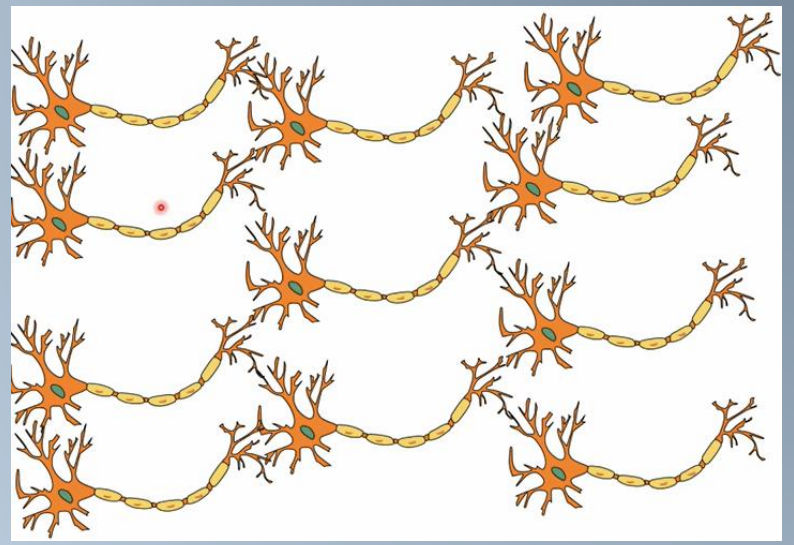
Activation Function

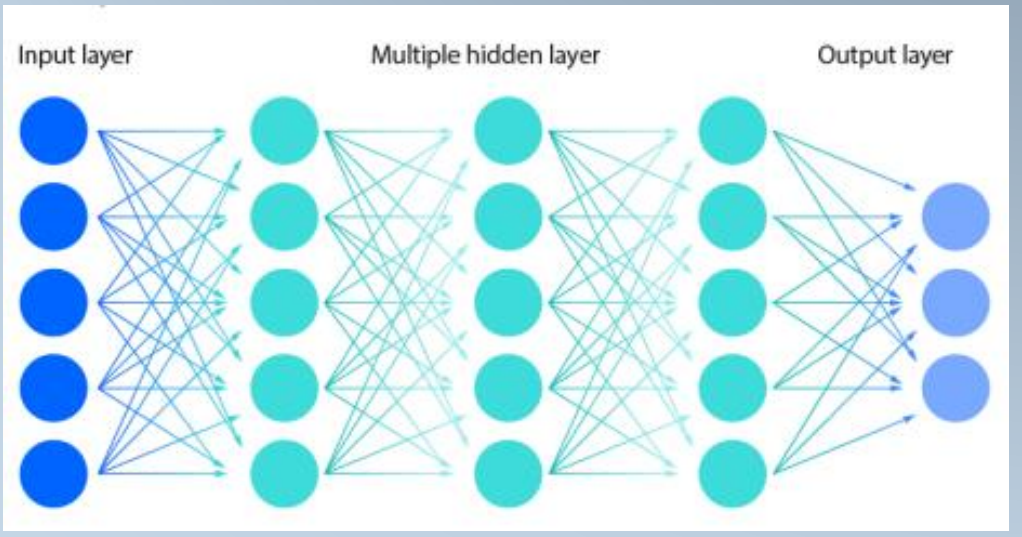
Step function (função Degrau)

Maior ou igual a 1 = 1
Caso contrário = 0

Multi-Neuron Network

(Neural network - Deep Learning Network)

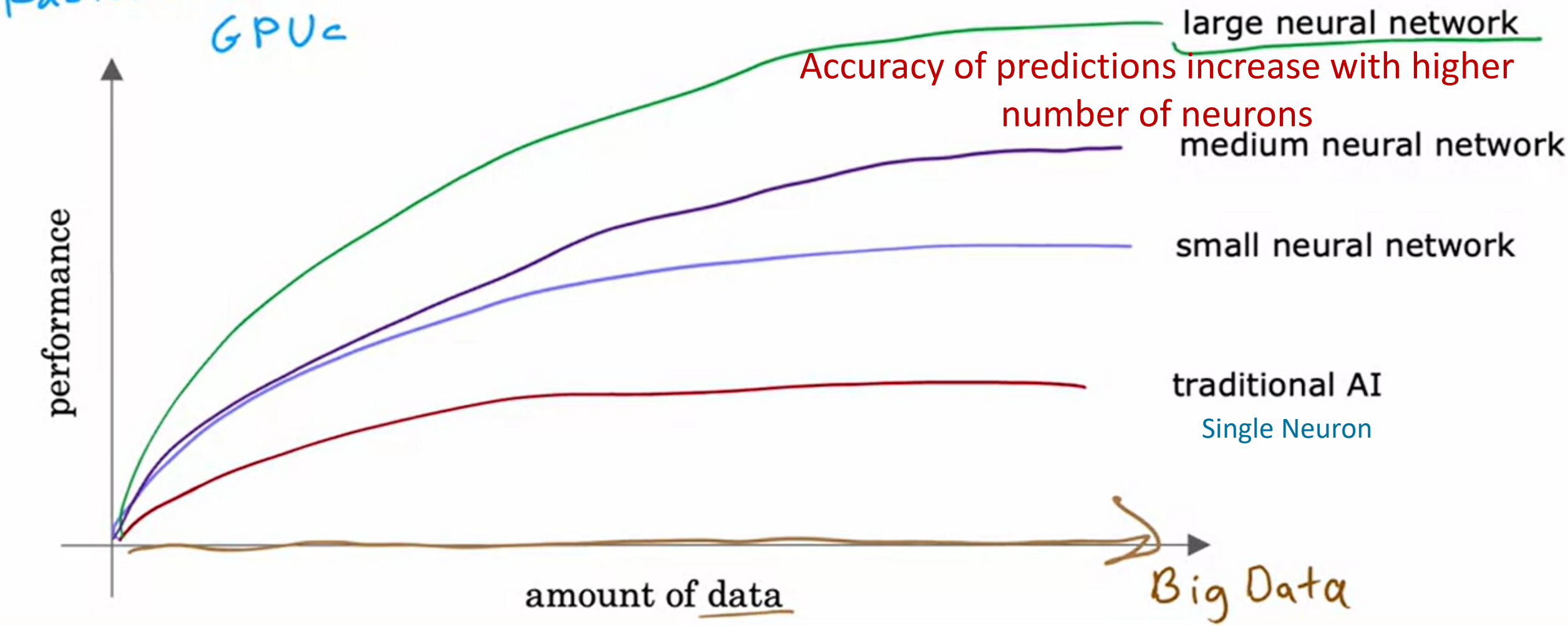




2. Uses of AI in Cybersecurity

Why Now?

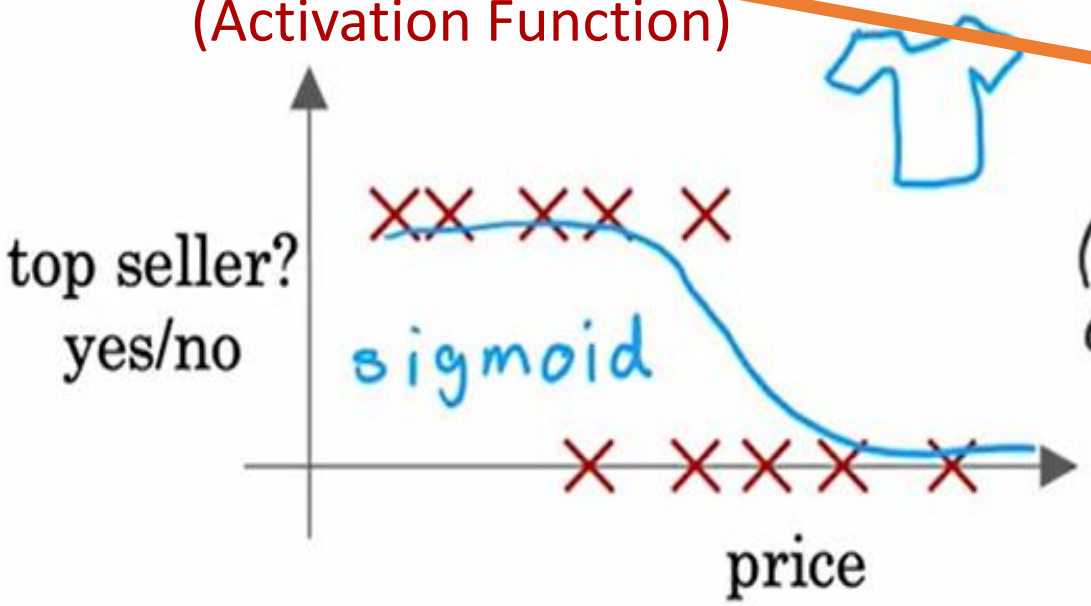
Faster computer processors
GPUs



2. Uses of AI in Cybersecurity

Demand Prediction

Sigmoid Function
(Activation Function)



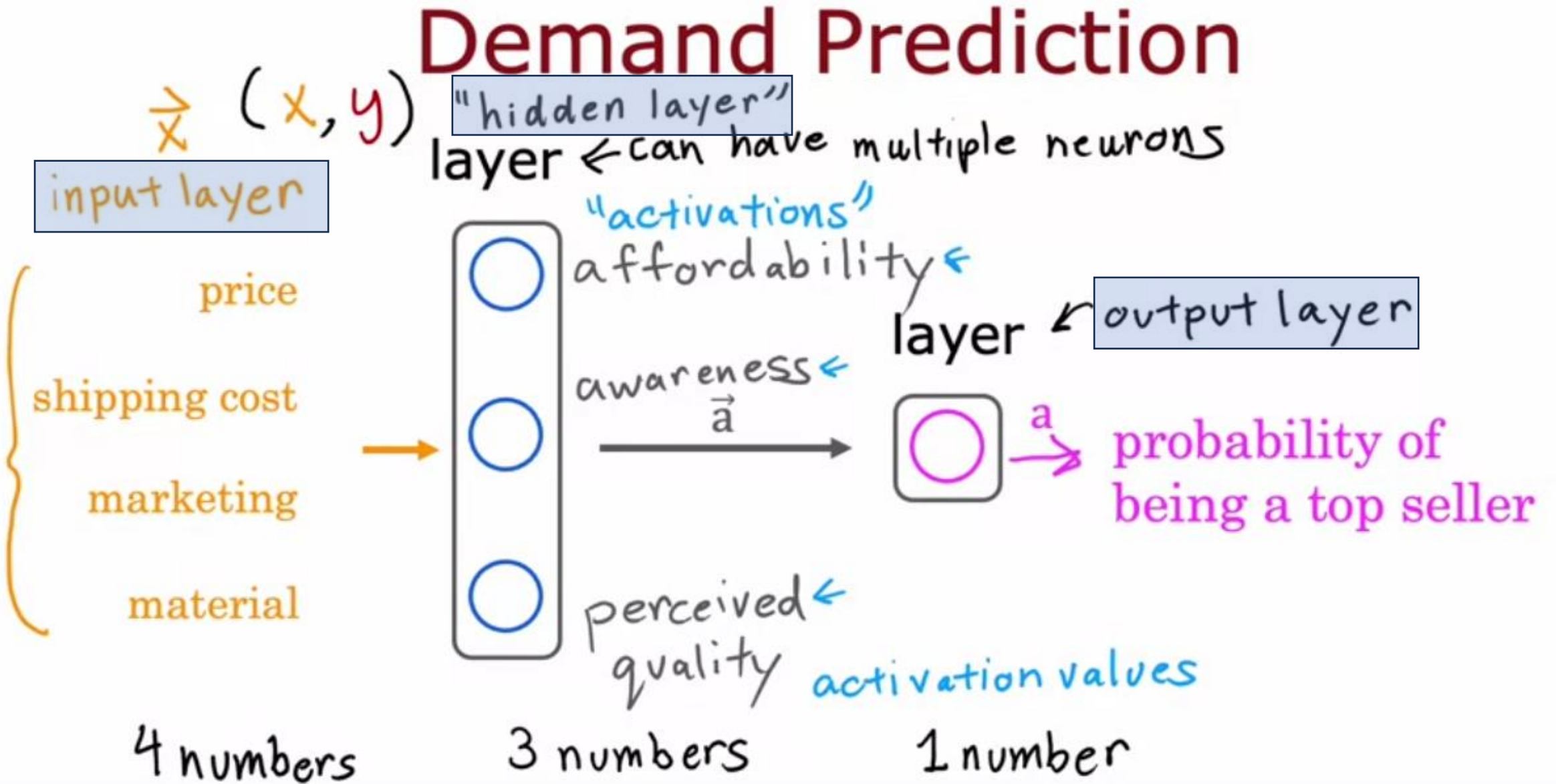
$x = price$ input

$a = f(x) = \frac{1}{1 + e^{-(wx+b)}}$ output

activation

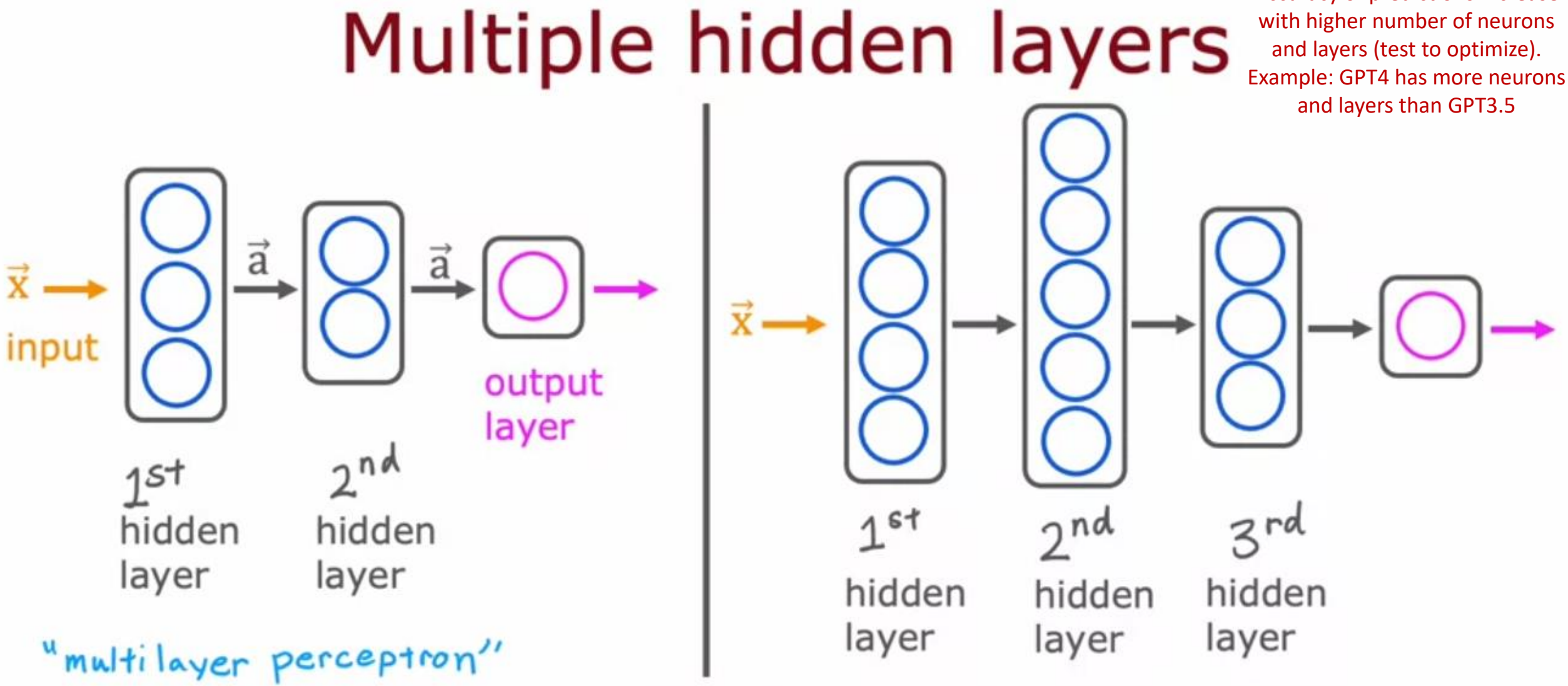
SINGLE NEURON CONFIGURATION

2. Uses of AI in Cybersecurity



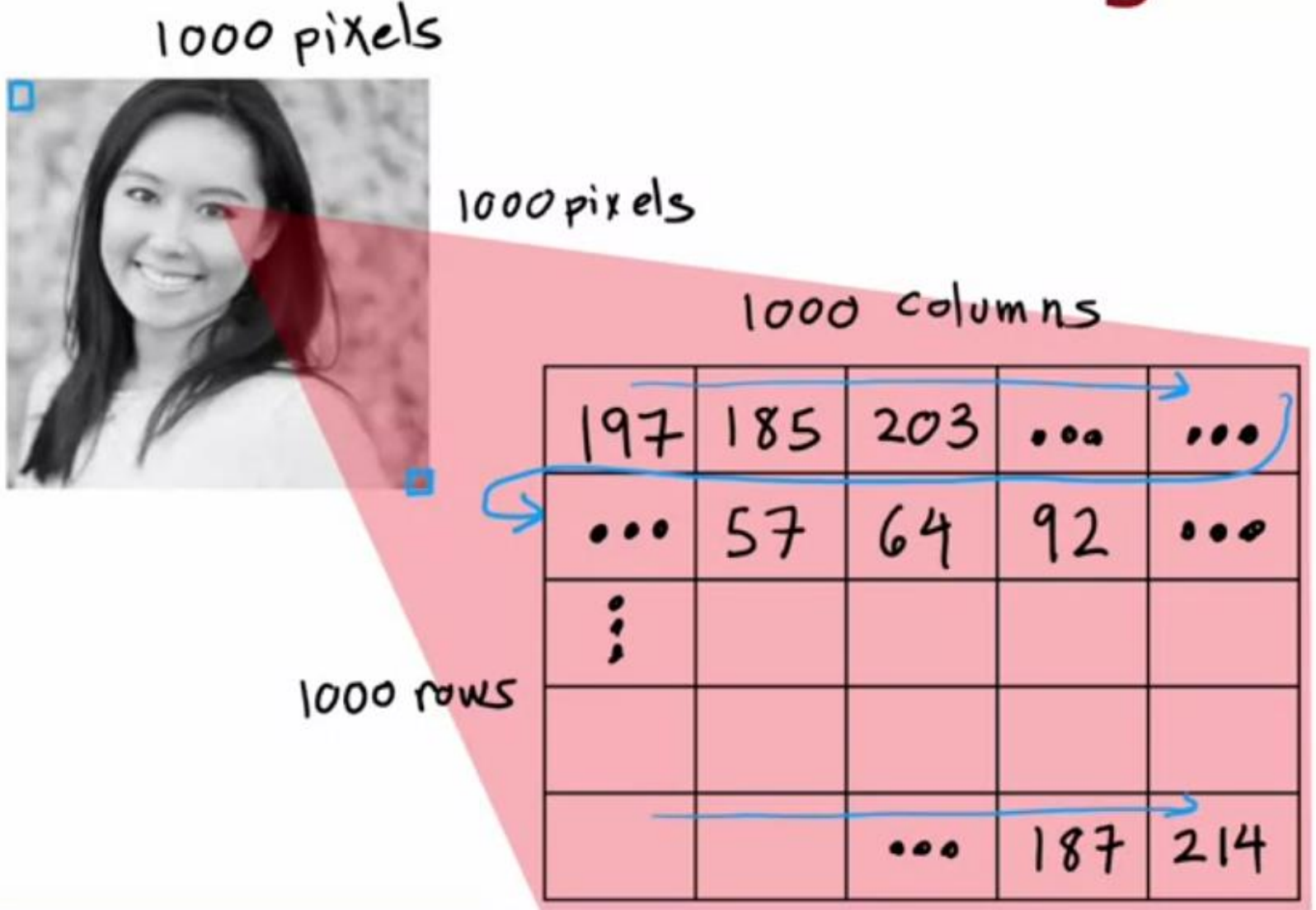
2. Uses of AI in Cybersecurity

Accuracy of predictions increase with higher number of neurons and layers (test to optimize). Example: GPT4 has more neurons and layers than GPT3.5



2. Uses of AI in Cybersecurity

Face recognition

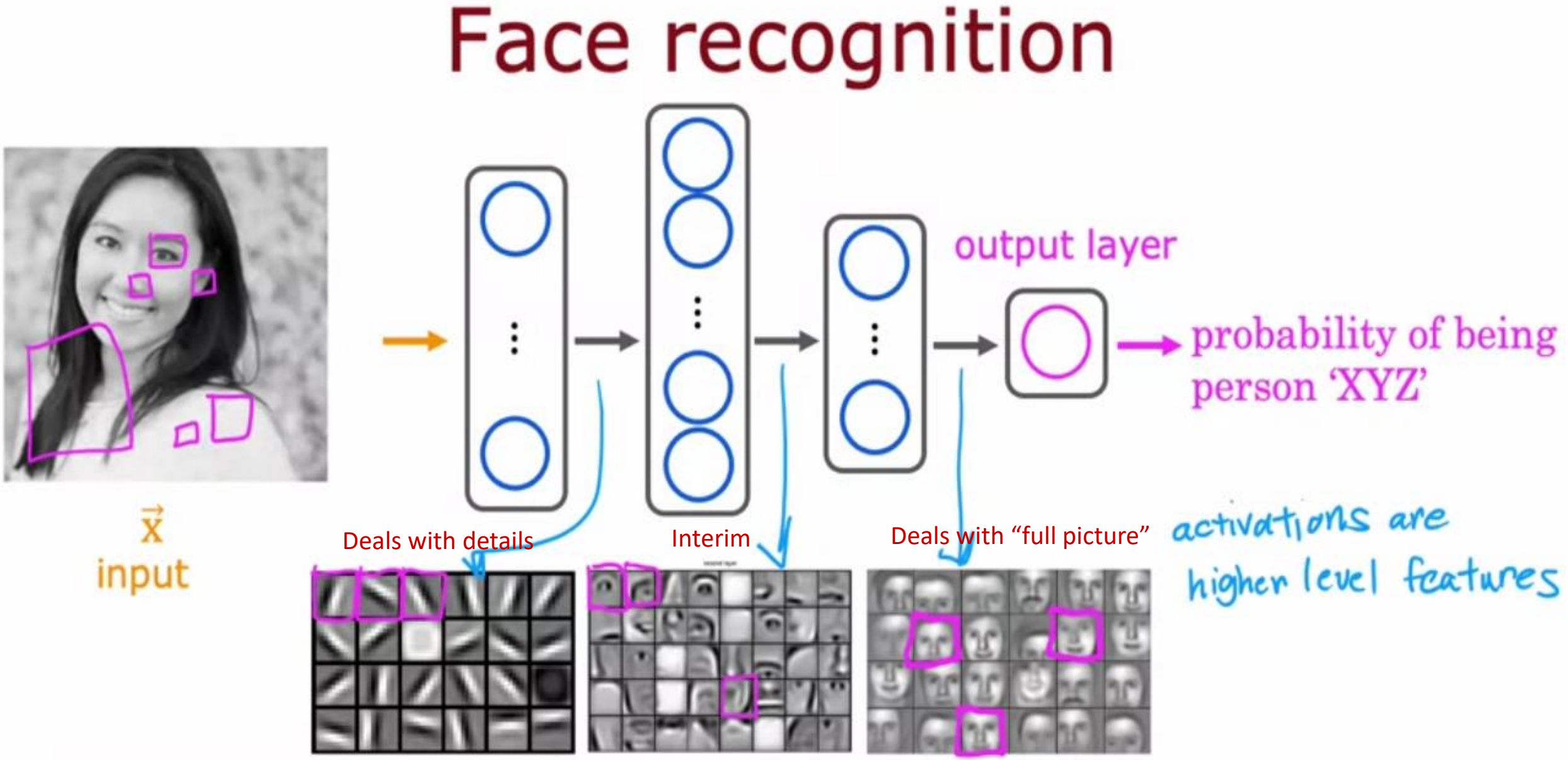


$\vec{x} =$

197
185
203
⋮
57
64
92
⋮
187
214

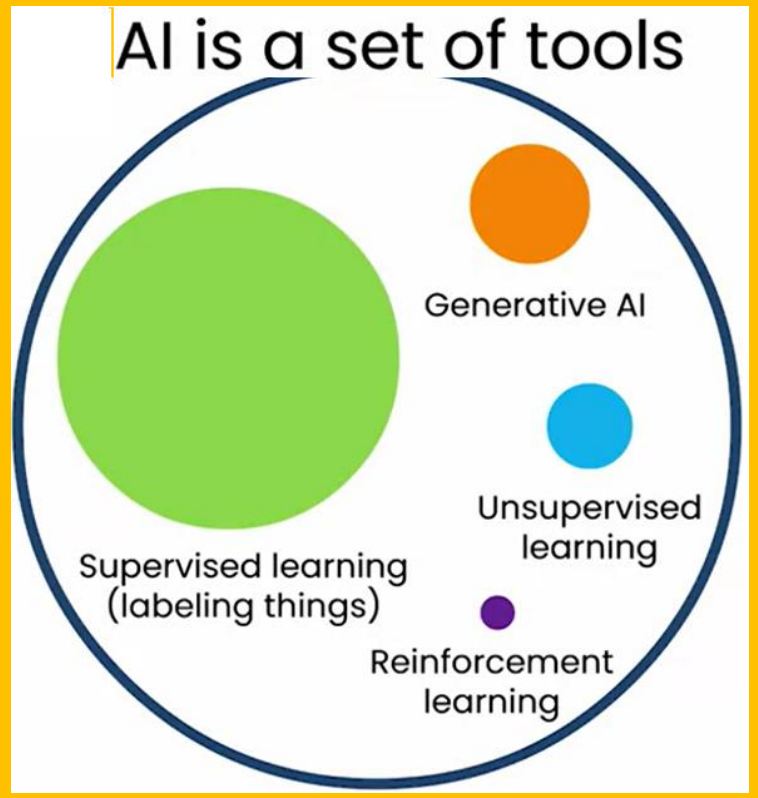
1 million

2. Uses of AI in Cybersecurity



2. Uses of AI in Cybersecurity

Unsupervised Learning



Unsupervised Learning: the model learns to identify patterns in the data on its own (autonomous training).

Beyond Supervised Learning

Unsupervised Learning

- Unsupervised Learning
  - Clustering
  - Anomaly detection
- Recommender Systems
- Reinforcement Learning

2. Uses of AI in Cybersecurity

Unsupervised Learning

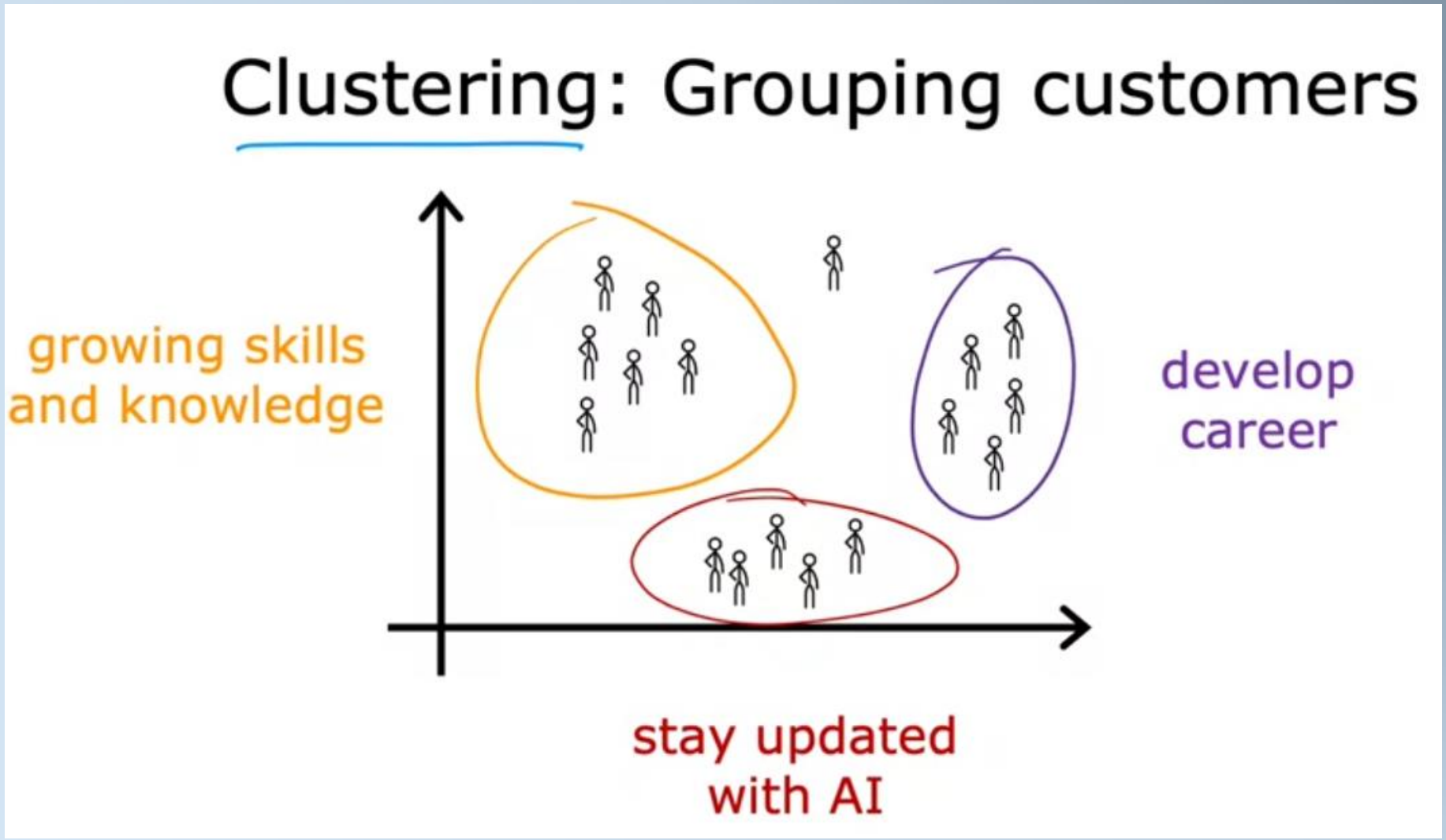
Clustering

Clustering algorithms, like K-means, can group network behaviors and **create profiles of "normal" activities**.

- Any activity that does not fit into these groups can be **flagged as potentially suspicious or anomalous**, indicating a possible attack, malicious activity, or unusual traffic.
- Can be used to:
  - Group types of malware, etc.
  - Identify unusual patterns over different time periods (in a more flexible way than supervised learning).
  - Detect events like intrusion attempts, unauthorized access, or unusual user behaviors.

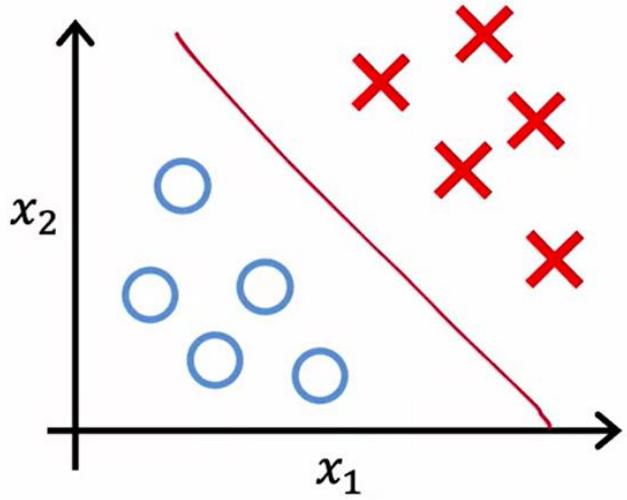
2. Uses of AI in Cybersecurity

Unsupervised Learning – Clustering (find something in data that allows grouping) – The algorithm does it by it self.



2. Uses of AI in Cybersecurity

Supervised learning

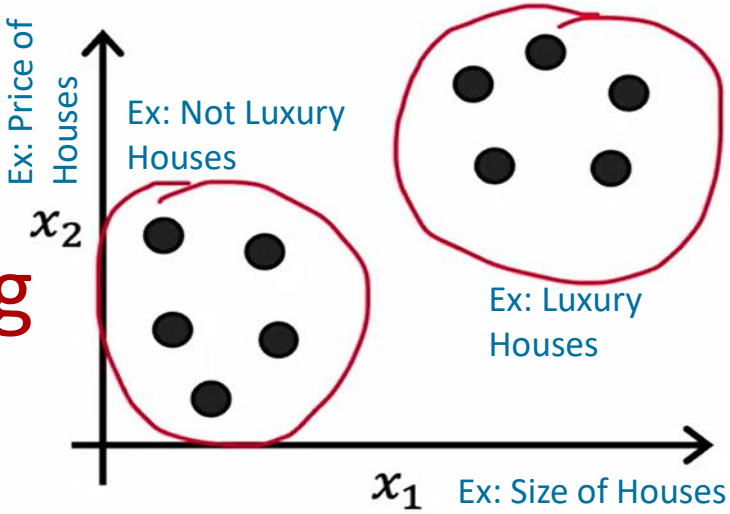


Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \ldots, (x^{(m)}, y^{(m)})\}$

There is “right answer”, i.e., there is y

Unsupervised learning

Used for Clustering



Training set: $\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots, x^{(m)}\}$

There is no “right answer”, i.e., there is no y

2. Uses of AI in Cybersecurity

Unsupervised Learning

Anomaly Detection

Since it doesn't require labeled data, it is particularly useful in **detecting new or evolving threats**.

- Can monitor network traffic **to identify unusual patterns that might indicate intrusions or attacks**.
- Detect **insider threats**, where a **legitimate user may engage in activities that suggest malicious intent** (e.g., accessing a large volume of sensitive files, logging in at unusual times, or using devices or abnormal IP addresses).
- Perform additional checks to confirm real fraud vs false alarm.

2. Uses of AI in Cybersecurity

Anomaly detection example

how often log in?
how many web pages visited?
transactions?
posts? typing speed?

Fraud detection:

- $x^{(i)}$ = features of user i's activities
- Model $p(x)$ from data.
- Identify unusual users by checking which have $p(x) < \varepsilon$

perform additional checks to identify real fraud vs. false alarms

Manufacturing:

$x^{(i)}$ = features of product i

airplane engine
circuit board
smartphone

ratios

Monitoring computers in a data center:

$x^{(i)}$ = features of machine i

- x_1 = memory use,
- x_2 = number of disk accesses/sec,
- x_3 = CPU load,
- x_4 = CPU load/network traffic.

2. Uses of AI in Cybersecurity

Density estimation

Dataset: $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

probability of x being seen in dataset

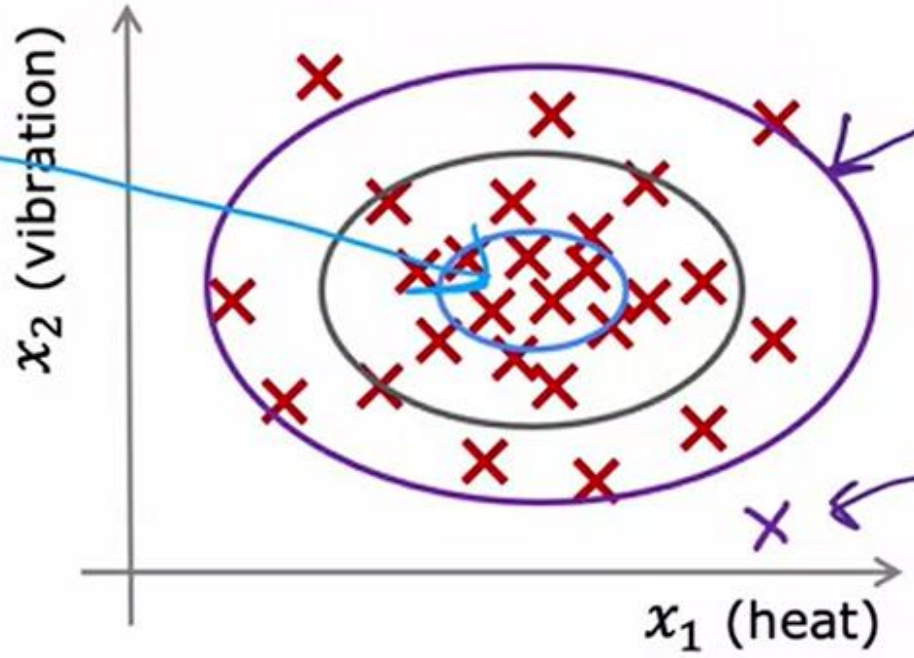
Model $p(x)$

Is x_{test} anomalous?

high probability

$p(x_{test}) \geq \varepsilon$

OK (normal)



low probability

$p(x_{test}) < \varepsilon$

epsilon: small number

unlikely anomaly

2. Uses of AI in Cybersecurity

Gaussian (Normal) distribution

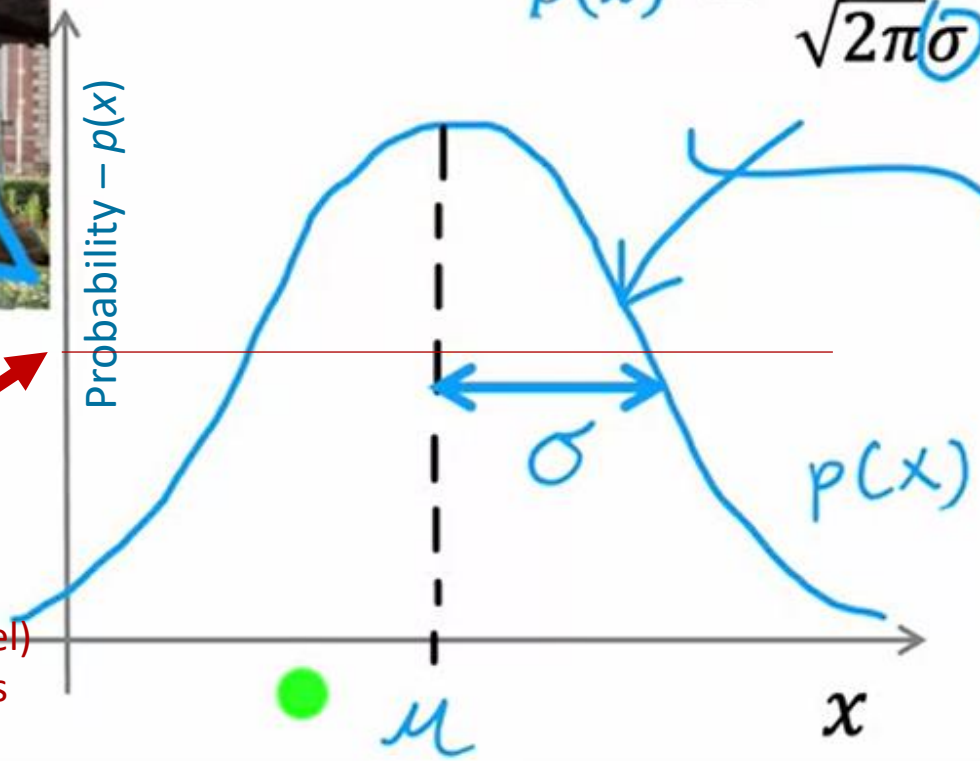
σ standard deviation

σ^2 variance

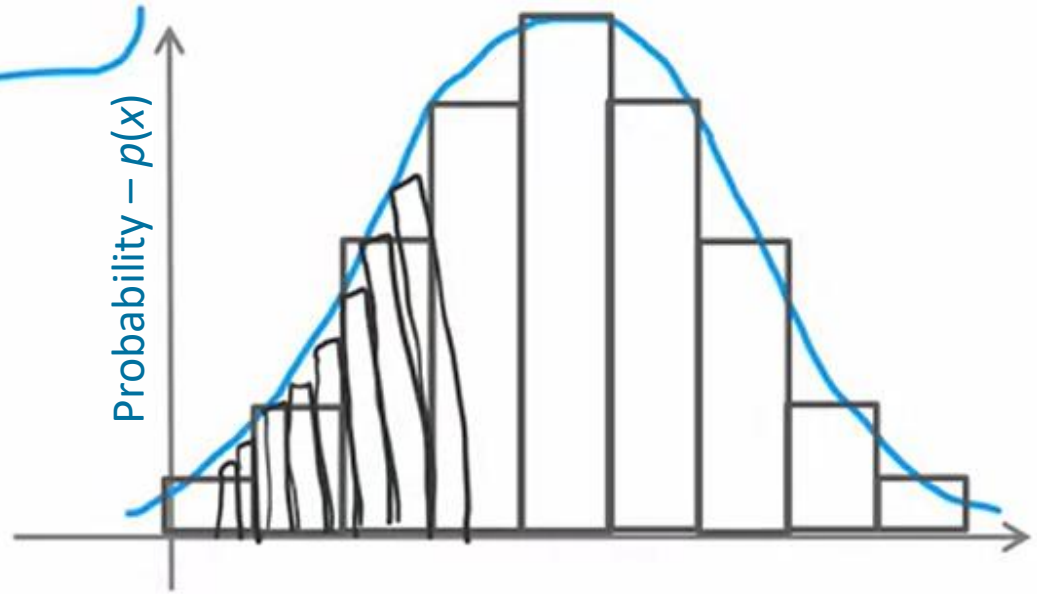
Say x is a number.
Probability of x is determined by a Gaussian with mean μ, variance σ^2.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

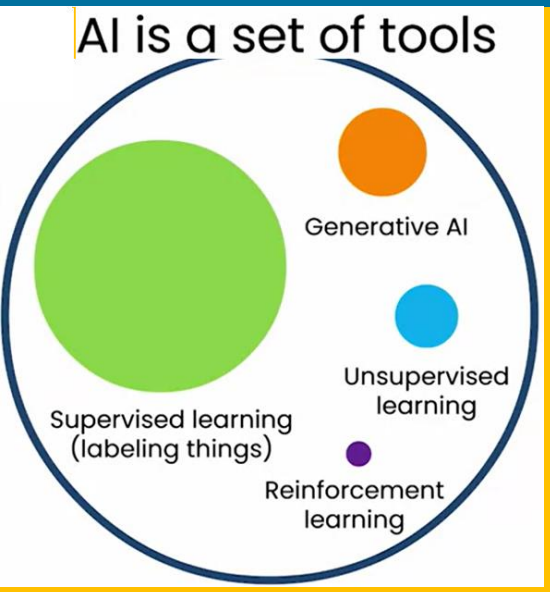
$\pi = 3.14$



What is the threshold (probability level) below which is anomaly?



2. Uses of AI in Cybersecurity



Generative AI

Large Language Models (LLMs)

→ Large Multimedia Models (LMMs)

Generative Learning: gives response with words (or images, videos) that are **statistically correlated** with the question (may experience hallucinations). **Uses Supervised Learning** to give responses.

In Cybersecurity used in:

- **Programming**: Create Malware
- **Natural Language Interaction**: Create fake e-mails, simulate traffic, etc.
- **Video and Image Creation**: Deepfake for social engineering (ex: fake videos), etc.

2. Uses of AI in Cybersecurity

Use Generative AI in Cyberdefense

- Typically, not used for real-time detection itself, but **can simulate potential attack strategies to:**
  1. **Train ML models, using synthetic datasets** (generated with Gen AI).
  2. **Train Corporate Employees** (simulating attacks)**.**

Use Generative AI in Cyberdefense

Examples of Use Cases to **improve security team awareness** and **to improve their security systems**:

1. **Simulating Phishing Attacks**: can **create realistic phishing emails** that mimic common tactics used by cybercriminals.
2. **Deepfake for Social Engineering**: convince an employee of a company to transfer funds to an account they control. To do this, they **use generative AI to create a deepfake video** of the company's CEO.
3. **Creating Malware Variants**: can develop different **variations of malware that evade traditional detection systems** to **enhance their antivirus** (new threats) → Help Hackers.
4. **Intrusion Testing Automation**: can simulate attacks on a network or application, such as Distributed Denial of Service (DDoS, ex: Syn Attack or Botnet) or SQL injection attacks.

2. Uses of AI in Cybersecurity

How Large Language Models (LLMs) work

LLMs are built by using supervised learning (A→B) to repeatedly predict the next word.

Use Tokens, instead of words…

My favorite food is a bagel with cream cheese

Input (A)	Output (B)
My favorite food is a	bagel
My favorite food is a bagel	with
My favorite food is a bagel with	cream
My favorite food is a bagel with cream	cheese

Same for programming

When we train a very large AI system on a lot of data (hundreds of billions of words), we get a Large Language Model like ChatGPT. (as a supervised Learning system, requires training data)

2. Uses of AI in Cybersecurity

Image, Audio and Video generation

A beautiful, pastoral mountain scene.
Landscape painting style (Midjourney)

Two cute kittens playing (DALL-E)

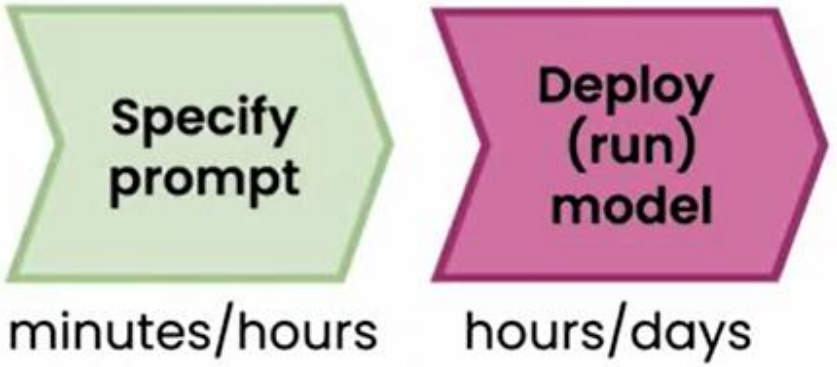
2. Uses of AI in Cybersecurity

Workflow using Generative AI

Supervised learning

Prompt-based AI

(Generative AI)



Much **quicker** than Supervised Learning and **supports non-structured data** (text, audio, images…)

Considerations on the use of AI in Cyberdefense

- Data Requirements: Both traditional ML and generative AI require **large datasets**, but **ML is more dependent on labeled data** (Gen AI uses unstructured data).

- Cost Considerations: **Implementing ML models can be expensive** due to data storage and processing power needs.

- Generative AI Risks: Can also be **used by hackers to create** sophisticated phishing attacks or malware.

Agenda

1. Artificial Intelligence

2. Uses of AI in Cybersecurity

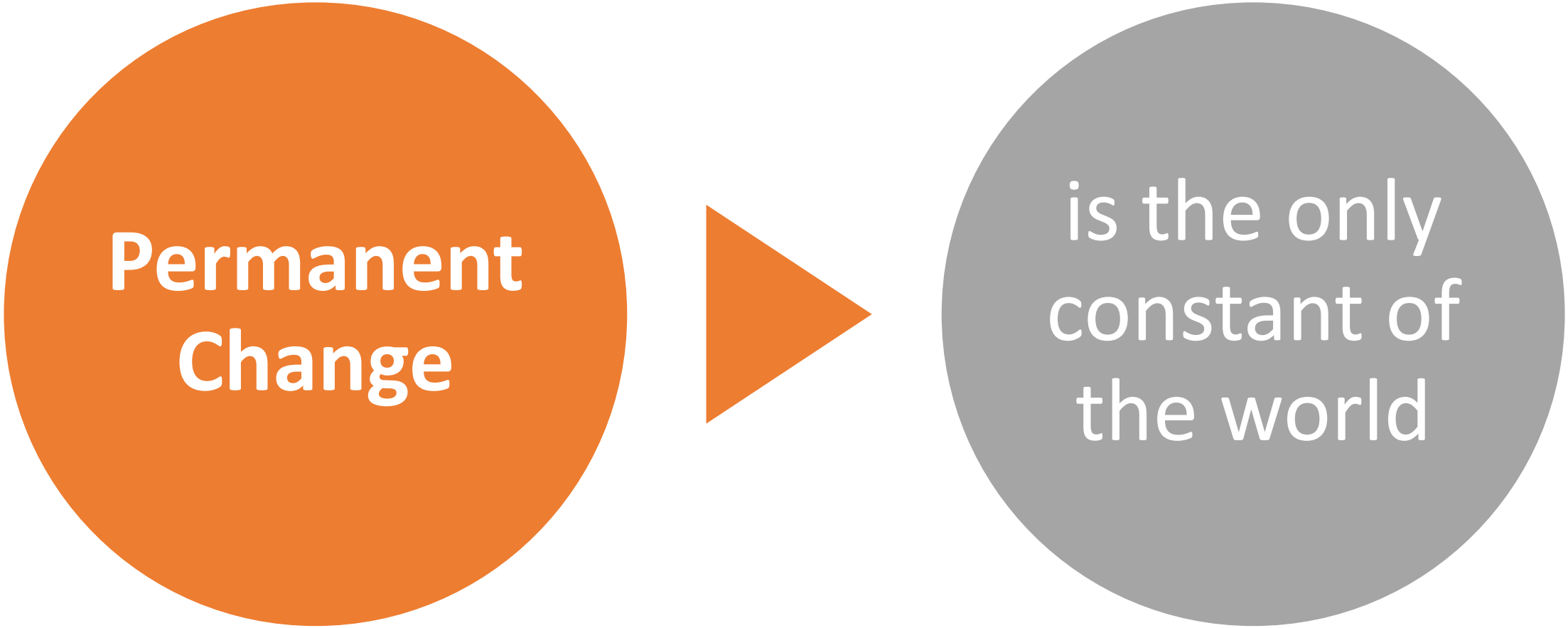
3. Conclusions

Conclusions

- Why Both? : **Combining** ML for detection and generative AI for simulation offers a **robust defense strategy -> Institutions should invest in both!**

- Predictions: **AI integration in cybersecurity** is expected to solve **95% of incidents by 2030**.

- AI is not just a **trend**, but a **necessity** to **protect the ever-evolving digital world**.


Permanent Change
is the only constant of the world