

أساسيات الأمن السيبراني المدعوم بالذكاء الاصطناعي (MSCS640) الذكاء الاصطناعي والأمن السيبراني

د. المعتر بالله قرباب



أكاديمية جوعان بن جاسم للدراسات الدفاعية
Joaan Bin Jassim Academy for Defence Studies

7 سبتمبر 2026



أكاديمية جوعان بن جاسم للدراسات الدفاعية
Joaan Bin Jassim Academy for Defence Studies



المحتويات

- 1 من القواعد القديمة إلى الذكاء الجديد
- 2 مفاهيم الأمن الأساسية
- 3 فهم الخصم
- 4 البيانات: وقود الذكاء الاصطناعي
- 5 مركز العمليات الأمنية الحديث



أهداف التعلم

بنهاية هذا الفصل، ستكون قادراً على:

- شرح قيود الأمن القائم على التوقعات باستخدام "هرم الألم" (Pyramid of Pain).
- مقارنة التعلم الخاضع للإشراف (Supervised) وغير الخاضع للإشراف (Unsupervised) في سياق أمني.
- تطبيق إطار سلسلة القتل السيبراني (Cyber Kill Chain) لتحليل الاختراقات في العالم الحقيقي.
- تقييم تحديات جودة البيانات وضرورة استخراج الميزات (feature extraction).
- تقييم أداء نماذج الذكاء الاصطناعي باستخدام مصفوفة الارتباك (Confusion Matrix) وأنواع الأخطاء.



التحول في الأمن السيبراني

النموذج القديم

كانت الأنظمة الأمنية صارمة، مثل حارس يتحقق من قائمة. إذا لم يكن التهديد على القائمة (توقيع)، يُسمح له بالدخول.

الواقع الحديث

المهاجمون آليون وسريعون ومتطورون. يسرقون بيانات الاعتماد ويستغلون "النوافذ المفتوحة". لم تعد الدفاعات التقليدية القائمة على القواعد كافية.

- يجب أن تنتقل إلى الدفاع المدعوم بالذكاء الاصطناعي (AI-powered defense).
- المطلوب: جودة نموذج الذكاء الاصطناعي تعتمد على جودة بياناته. فهم "كيفية" الهجمات السيبرانية هو شرط أساسي لبناء "ماذا" من أدوات الذكاء الاصطناعي.



المحتويات

- 1 من القواعد القديمة إلى الذكاء الجديد
- 2 مفاهيم الأمن الأساسية
- 3 فهم الخصم
- 4 البيانات: وقود الذكاء الاصطناعي
- 5 مركز العمليات الأمنية الحديث



حدود التوقعات

- الكشف القائم على التوقعات (Signature-Based Detection): يعتمد على تحديد الأنماط الفريدة (التجزئات مثل MD5/SHA-256).
- العملية: تحديد الفيروس ← حساب التجزئة ← إضافتها إلى قاعدة البيانات ← المسح والحظر.

المشكلة: البرمجيات الخبيثة متعددة الأشكال (Polymorphic Malware)

يستخدم المهاجمون أدوات لتغيير الكود قليلاً في كل مرة ينتشر فيها. هذا يغير التجزئة. بالنسبة لنظام قائم على التوقعات، يبدو كملف جديد وآمن تماماً.



ميزة الذكاء الاصطناعي: التحليل السلوكي

الذكاء الاصطناعي وتعلم الآلة [2] يغيّران التركيز من "التوقعات السيئة" (Bad Signatures) إلى "السلوك السيئ" (Bad Behavior).

المفهوم: خط الأساس (The Baseline)

يراقب الذكاء الاصطناعي الشبكة لحساب النشاط "الطبيعي".

- طبيعي: المستخدم أيسجل الدخول في الساعة 9:00 صباحاً من لندن.
- شاذ (Anomaly): المستخدم أيسجل الدخول في الساعة 3:00 صباحاً من بلد مختلف.

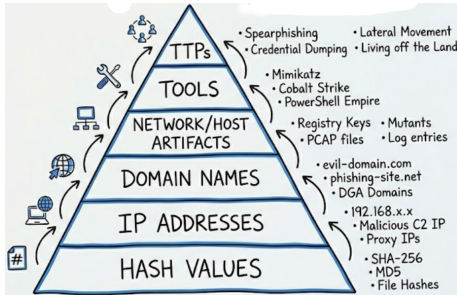
- النتيجة: الكشف عن هجمات اليوم صفر (Zero-Day Attacks) (استغلالات لم تُشاهد من قبل) بناءً على الانحراف عن خط الأساس [3].



I (The Pyramid of Pain) هرم الألم

ابتكره ديفيد بيانكو، هذا المفهوم يصنف المؤشرات حسب مقدار "الألم" الذي تسببه للمهاجم إذا حضرناها.

- القاع (قيم التجزئة وعناوين IP): من السهل على المهاجمين تغييرها. حظرها يسبب لا ألم.
- القمة (TTPs - السلوك): التكتيكات والتقنيات والإجراءات. هذا هو كيف يعمل المهاجم.



موقع الذكاء الاصطناعي

يعمل الذكاء الاصطناعي في قمة الهرم. من الصعب جداً على المهاجمين التخلي عن أساليبهم أو مهاراتهم. الكشف عن السلوك يسبب أقصى درجات الألم.



التعلم الخاضع للإشراف مقابل غير الخاضع للإشراف

التعلم غير الخاضع للإشراف (بالملاحظة) (Unsupervised Learning)

- المفهوم: تغذية البيانات الخام بدون تسميات. يقوم الذكاء الاصطناعي بتجميع التشابهات وإبراز القيم الشاذة.
- الأفضل لـ: التهديدات الداخلية والبحث عن هجمات جديدة وغير معروفة.

التعلم الخاضع للإشراف (بالأمثلة) (Supervised Learning)

- المفهوم: التدريس باستخدام التسميات ("هذه برمجية خبيثة"، "هذا آمن").
- الأفضل لـ: كشف البرمجيات الخبيثة وتصنيف الهجمات المعروفة.



القسم 1: تقييم (1/3)

السؤال 1

لماذا يكون الكشف القائم على التوقعات غير فعال غالباً ضد التهديدات الحديثة؟

- أ. لأنه يستهلك الكثير من قوة المعالجة.
- ب. لأنه لا يمكنه كشف البرمجيات الخبيثة متعددة الأشكال (Polymorphic Malware) التي تغير تجزئتها.
- ج. لأنه يعمل فقط على أنظمة تشغيل ويندوز.
- د. لأنه يتطلب اتصالاً نشطاً بالإنترنت ليعمل.



القسم 1: تقييم (1/3) - الإجابة

السؤال 1

لماذا يكون الكشف القائم على التوقعات غير فعال غالباً ضد التهديدات الحديثة؟

- أ. لأنه يستهلك الكثير من قوة المعالجة.
- ب. لأنه لا يمكنه كشف البرمجيات الخبيثة متعددة الأشكال (Polymorphic Malware) التي تغير تجزئتها. (✓)
- ج. لأنه يعمل فقط على أنظمة تشغيل ويندوز.
- د. لأنه يتطلب اتصالاً نشطاً بالإنترنت ليعمل.

الشرح:

يعتمد الكشف القائم على التوقعات على بصمات ثابتة (تجزئات). البرمجيات الخبيثة متعددة الأشكال تغير بنية كودها (وبالتالي تجزئتها) مع كل تكرار، مما يجعل التوقيع الثابت عديم الفائدة.



القسم 1: تقييم (2/3)

السؤال 2

وفقًا لـ "هرم الألم" (Pyramid of Pain)، أي مؤشر يسبب أكبر قدر من الألم للمهاجم عند اكتشافه وحظره؟

- أ. عناوين IP.
- ب. قيم التجزئة.
- ج. أسماء النطاقات.
- د. التكتيكات والتقنيات والإجراءات (TTPs).



القسم 1: تقييم (2/3) - الإجابة

السؤال 2

وفقاً لـ "هرم الألم" (Pyramid of Pain)، أي مؤشر يسبب أكبر قدر من الألم للمهاجم عند اكتشافه وحظره؟

- أ. عناوين IP.
- ب. قيم التجزئة.
- ج. أسماء النطاقات.
- د. التكتيكات والتقنيات والإجراءات (TTPs). (✓)

الشرح:

تمثل TTPs سلوكيات ومهارات المهاجم. بينما يمكن للمهاجم تغيير عنوان IP أو تجزئة ملف بسهولة، فإن تغيير أساليبه وأدواته الأساسية يتطلب وقتاً كبيراً وإعادة تدريب، مما يسبب "أقصى درجات الألم".



القسم 1: تقييم (3/3)

السؤال 3

أي نوع من تعلم الذكاء الاصطناعي هو الأنسب للكشف عن التهديدات الداخلية حيث يكون غلط الهجوم المحدد غير معروف؟

- أ. التعلم الخاضع للإشراف (Supervised Learning).
- ب. التعلم غير الخاضع للإشراف (Unsupervised Learning).
- ج. التعلم المعزز (Reinforcement Learning).
- د. مطابقة التوقعات (Signature Matching).



القسم 1: تقييم (3/3) - الإجابة

السؤال 3

أي نوع من تعلم الذكاء الاصطناعي هو الأنسب للكشف عن التهديدات الداخلية حيث يكون نمط الهجوم المحدد غير معروف؟

أ. التعلم الخاضع للإشراف (Supervised Learning).

ب. التعلم غير الخاضع للإشراف (Unsupervised Learning). (✓)

ج. التعلم المعزز (Reinforcement Learning).

د. مطابقة التوقيعات (Signature Matching).

الشرح:

يتعلم التعلم غير الخاضع للإشراف عن طريق الملاحظة بدون تسميات. إنه ينشئ خط أساس للسلوك الطبيعي ويحدد القيم الإحصائية الشاذة (الانحرافات)، مما يجعله مثالياً للعثور على التهديدات الخفية أو غير المعروفة مثل الموظف الماروق.



المحتويات

- 1 من القواعد القديمة إلى الذكاء الجديد
- 2 مفاهيم الأمن الأساسية
- 3 فهم الخصم
- 4 البيانات: وقود الذكاء الاصطناعي
- 5 مركز العمليات الأمنية الحديث



ثالث CIA

كل ضابط أممي ونموذج ذكاء اصطناعي يحمي أحد هذه الأهداف الثلاثة:

- 1 السرية (Confidentiality): الحفاظ على الأسرار.
 - دفاع الذكاء الاصطناعي: كشف تسريب البيانات.
- 2 النزاهة (Integrity): الدقة والموثوقية.
 - دفاع الذكاء الاصطناعي: كشف التعديلات غير المصرح بها على الملفات.
- 3 التوافر (Availability): عمل الأنظمة عند الحاجة.
 - دفاع الذكاء الاصطناعي: التنبؤ بالأعطال أو تصفية حركة مرور هجمات الحرمان من الخدمة الموزعة (DDoS).



سطح الهجوم (The Attack Surface)

مجموع كل النقاط التي يمكن للهجوم محاولة الدخول منها.

- سطح الشبكة: المنافذ المفتوحة، شبكات Wi-Fi، تكوينات السحابة.
- السطح البشري: غالباً ما يكون الحلقة الأضعف. عرضة لهندسة الاجتماعية (Social Engineering) (التصيد الاحتيالي (Phishing)).
- سطح البرمجيات: الثغرات (الأخطاء) في التطبيقات المثبتة.



القسم 2: تقييم (1/3) - الإجابة

السؤال 1

في ثالث CIA، أي مفهوم يضمن عدم تغيير البيانات أو حذفها من قبل أفراد غير مصرح لهم؟

- أ. السرية (Confidentiality).
- ب. النزاهة (Integrity). (✓)
- ج. التوافر (Availability).
- د. التفويض (Authorization).

الشرح:

النزاهة هي ضمان أن المعلومات دقيقة وموثوقة. إذا قام مهاجم بتعديل معاملة بنكية أو حذف ملف سجل، فقد انتهك نزاهة النظام.



القسم 2: تقييم (2/3) - الإجابة

السؤال 2

ما هو المبدأ الأساسي لـ "الدفاع في العمق" (Defense in Depth)؟

أ. الاعتماد فقط على الذكاء الاصطناعي لالتقاط جميع التهديدات.

ب. استخدام جدار حماية واحد قوي جداً.

ج. استخدام طبقات متعددة من الضوابط الأمنية. (✓)

د. فصل الشبكة عن الإنترنت.

الشرح:

يستخدم الدفاع في العمق نهجاً متعدد الطبقات (مثل قلعة بها خندق وجدران وحراس). إذا تجاوز المهاجم المحيط (جدار الحماية)، فإن حماية نقطة النهاية أو تشفير البيانات تعمل كدفاع احتياطي.



القسم 2: تقييم (3/3)

السؤال 3

رسائل التصيد الاحتيالي التي تخدع الموظفين للكشف عن كلمات المرور تستهدف أي جزء من سطح الهجوم؟

- أ. سطح الشبكة.
ب. سطح البرمجيات.
ج. السطح البشري.
د. السطح المادي.



القسم 2: تقييم (3/3) - الإجابة

السؤال 3

رسائل التصيد الاحتيالي التي تخدع الموظفين للكشف عن كلمات المرور تستهدف أي جزء من سطح الهجوم؟

- أ. سطح الشبكة.
ب. سطح البرمجيات.
ج. **السطح البشري. (✓)**
د. السطح المادي.

الشرح:

يشير السطح البشري إلى الموظفين والمستخدمين داخل المنظمة. هجمات الهندسة الاجتماعية مثل التصيد الاحتيالي تستغل علم النفس البشري (الثقة، الخوف، الإلحاح) بدلاً من الأخطاء الفنية في البرمجيات.



المحتويات

- 1 من القواعد القديمة إلى الذكاء الجديد
- 2 مفاهيم الأمن الأساسية
- 3 فهم الخصم
- 4 البيانات: وقود الذكاء الاصطناعي
- 5 مركز العمليات الأمنية الحديث



سلسلة القتل السيبراني (لوکھید مارتن) (The Cyber Kill Chain)

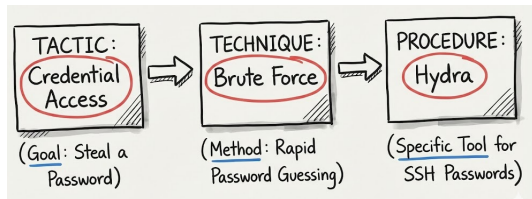
نموذج خطي متدرج لهجوم سيبراني. كسر أي حلقة يوقف الهجوم.

- 1 الاستطلاع (Reconnaissance): البحث عن الهدف.
- 2 التسليح (Weaponization): إنشاء الحمولة الخبيثة.
- 3 التسليم (Delivery): إرسال السلاح (مثل البريد الإلكتروني).
- 4 الاستغلال (Exploitation): تشغيل الثغرة.
- 5 التثبيت (Installation): تثبيت باب خلفي.
- 6 القيادة والتحكم ((Command and Control (C2)): التحكم عن بعد في الضحية.
- 7 الإجراءات على الأهداف (Actions on Objectives): سرقة البيانات أو تدميرها.



إطار II MITRE ATT&CK

- التكتيكات (الهدف): لماذا يفعلون هذا؟ (على سبيل المثال، الوصول إلى بيانات الاعتماد ((Credential Access)).
- التقنيات (الطريقة): كيف يفعلون ذلك؟ (على سبيل المثال، القوة الغاشمة ((Brute Force)).
- الإجراءات (التفاصيل): ما الأداة التي استخدموها؟ (على سبيل المثال، هيدرا ((Hydra)).



الأهمية بالنسبة للذكاء الاصطناعي

يسمح لنا بتصنيف الأحداث بدقة. بدلاً من مجرد "تنبيه"، يمكن للذكاء الاصطناعي أن يقول "تم كشف تقنية: القوة الغاشمة ((Brute Force))."



القسم 3: تقييم (1/3)

السؤال 1

في سلسلة القتل السيبراني (Cyber Kill Chain)، ما هي مرحلة "التسليم" (Delivery)؟

- أ. إنشاء الفيروس.
- ب. نقل الحمولة المسلحة إلى الضحية (على سبيل المثال، عبر البريد الإلكتروني).
- ج. البحث عن موظفي الهدف.
- د. سرقة البيانات من قاعدة البيانات.



القسم 3: تقييم (1/3) - الإجابة

السؤال 1

في سلسلة القتل السيبراني (Cyber Kill Chain)، ما هي مرحلة "التسليم" (Delivery)؟

- أ. إنشاء الفيروس.
- ب. نقل الحمولة المسلحة إلى الضحية (على سبيل المثال، عبر البريد الإلكتروني). (✓)
- ج. البحث عن موظفي الهدف.
- د. سرقة البيانات من قاعدة البيانات.

الشرح:

التسليم هو مرحلة النقل. إنها الطريقة التي يوصل بها المهاجم الكائن الخبيث (الذي تم إنشاؤه في مرحلة التسليح) إلى نظام الهدف، وعادة ما يكون ذلك من خلال رسائل البريد الإلكتروني التصيدية أو محركات أقراص USB أو مواقع الويب الخبيثة.



القسم 3: تقييم (2/3)

السؤال 2

ماذا يمثل "التكتيك" (Tactic) في إطار MITRE ATT&CK؟

- أ. أداة البرمجيات المحددة المستخدمة.
- ب. الهدف الفني للخصم ("لماذا").
- ج. الطريقة المستخدمة لتحقيق الهدف ("كيف").
- د. عنوان IP للمهاجم.



القسم 3: تقييم (2/3) - الإجابة

السؤال 2

ماذا يمثل "التكتيك" (Tactic) في إطار MITRE ATT&CK؟

- أ. أداة البرمجيات المحددة المستخدمة.
- ب. الهدف الفني للخصم ("لماذا"). (✓)
- ج. الطريقة المستخدمة لتحقيق الهدف ("كيف").
- د. عنوان IP للمهاجم.

الشرح:

في MITRE ATT&CK، تمثل التكتيكات الهدف عالي المستوى للمهاجم، مثل "الوصول الأولي" أو "التنفيذ" أو "التسلل". إنها تجيب على لماذا يقوم المهاجم بتنفيذ إجراء.



القسم 3: تقييم (3/3)

السؤال 3

- ما هي خطوة "الحركة الجانبية" (lateral movement) الرئيسية في اختراق شركة تارجت؟
- الانتقال من بوابة مورد أنظمة التدفئة والتهوية وتكييف الهواء إلى الشبكة الرئيسية.
 - اقتحام المتجر الفعلي ليلاً.
 - نقل البيانات من الولايات المتحدة إلى روسيا.
 - التبديل من خوادم ويندوز إلى لينكس.



القسم 3: تقييم (3/3) - الإجابة

السؤال 3

ما هي خطوة "الحركة الجانبية" (lateral movement) الرئيسية في اختراق شركة تارجت؟

أ. الانتقال من بوابة مورد أنظمة التدفئة والتبريد وتهوية وتكييف الهواء إلى الشبكة الرئيسية. (✓)

ب. اقتحام المتجر الفعلي ليلاً.

ج. نقل البيانات من الولايات المتحدة إلى روسيا.

د. التبدیل من خوادم ویندوز إلى لینکس.

الشرح:

اخترق المهاجمون أولاً مورداً خارجياً أقل أماناً (HVAC). ثم استخدموا تلك البيانات الاعتمادية للتحرك بشكل جانبي عبر الشبكة للوصول إلى الهدف ذي القيمة العالية: شبكة نقاط البيع.



مصادر البيانات الرئيسية

- تدفقات الشبكة (Network Flows): "فاتورة الهاتف". ملخصات حركة المرور (من، متى، كم). سريعة وخفيفة.
- التقاطات الحزم ((Packet Captures (PCAP): "التسجيل الكامل". كل التفاصيل. أحجام ملفات ضخمة.
- سجلات نقطة النهاية (Endpoint Logs): "يوميات الجهاز". تسجيلات الدخول، إنشاء العمليات (على سبيل المثال، Sysmon).
- سجلات الويب (Web Logs): "سجل الزوار". طلبات إلى خوادم الويب.

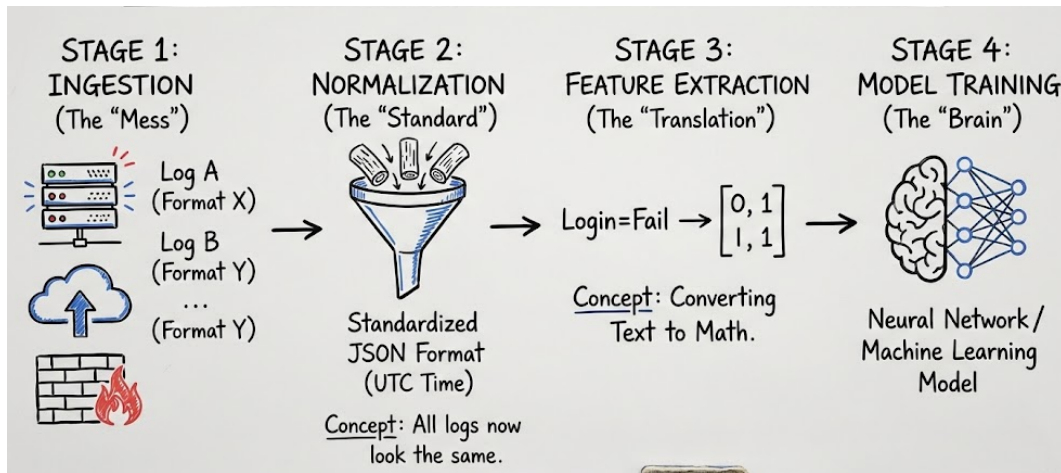


تحدي "البيانات الضخمة"

- الحجم (Volume): تيرابايت من البيانات يومياً.
- السرعة (Velocity): تحدث الهجمات في أجزاء من الثانية؛ المعالجة في الوقت الفعلي مطلوبة.
- التنوع (Variety): تنسيقات مختلفة (سجلات نصية، JSON، حزم خام).
- عدم التوازن (Imbalance): أكبر عقبة أمام الذكاء الاصطناعي.
- 999,990 حدث "آمن" مقابل 10 أحداث "هجوم".
- يواجه الذكاء الاصطناعي صعوبة في تعلم شكل الهجوم لأن الأمثلة نادرة.



خط أنابيب تحضير البيانات II



بايثون: محاكاة البيانات (1/2)

يحاكي هذا المقطع البرمجي تحميل بيانات السجل الخام. في سيناريو حقيقي، سيكون هذا ملف CSV كبير.

```
1 import pandas as pd
2 import io
3
4 # 1. Simulate raw log data (CSV format)
5 log_data = """timestamp,source_ip,status_code
6 2025-08-16 10:00:01,192.168.1.10,200
7 2025-08-16 10:00:05,192.168.1.10,404
8 2025-08-16 10:00:10,192.168.1.10,404
9 2025-08-16 10:00:15,10.0.0.5,200
10 """
11
12 # 2. Load the data into a DataFrame
13 df = pd.read_csv(io.StringIO(log_data))
```



بايثون: منطق استخراج الميزات (2/2)

هنا، نقوم بتحويل رموز "404" الختام إلى عدد رياضي. يمكن للذكاء الاصطناعي أن يتعلم أن العدد الكبير هو أمر مشبوه.

```

1 # 3. FEATURE EXTRACTION
2 # Logic: Group by IP address, count how many '404' errors appear.
3
4 features = df.groupby('source_ip')['status_code'].apply(
5     lambda x: (x == 404).sum()
6 ).reset_index(name='error_count_feature')
7
8 print("Data ready for AI Model:")
9 print(features)
10
11 # Output:
12 #   source_ip  error_count_feature
13 # 0      10.0.0.5                  0
14 # 1  192.168.1.10                  2

```



القسم 4: تقييم (1/3)

السؤال 1

- أي مصدر بيانات يوفر "التسجيل الكامل" لحركة مرور الشبكة، بما في ذلك محتوى المحادثة؟
- أ. تدفقات الشبكة (Network Flows).
 - ب. التقاطات الحزم (Packet Captures (PCAP)).
 - ج. سجلات DNS.
 - د. تنبيهات جدار الحماية.



القسم 4: تقييم (1/3) - الإجابة

السؤال 1

أي مصدر بيانات يوفر "التسجيل الكامل" لحركة مرور الشبكة، بما في ذلك محتوى المحادثة؟

- أ. تدفقات الشبكة (Network Flows).
- ب. التقاطات الحزم (Packet Captures (PCAP)). (✓)
- ج. سجلات DNS.
- د. تنبيهات جدار الحماية.

الشرح:

تلتقط بيانات PCAP المحملة الكاملة ورؤوس حزم الشبكة. بينما تعمل تدفقات الشبكة مثل فاتورة الهاتف (بيانات وصفية فقط)، تعمل PCAP مثل تسجيل التنصت على المحادثة الفعلية.



القسم 4: تقييم (2/3)

السؤال 2

ما هو "عدم توازن البيانات" (Data Imbalance) في سياق تدريب الذكاء الاصطناعي للأمن؟

- أ. وجود الكثير من البيانات لتخزينها.
- ب. وجود بيانات من مناطق زمنية مختلفة جداً.
- ج. وجود كمية هائلة من البيانات "الآمنة" وعدد قليل جداً من أمثلة "الهجمات".
- د. وجود بيانات غير مشفرة.



القسم 4: تقييم (2/3) - الإجابة

السؤال 2

ما هو "عدم توازن البيانات" (Data Imbalance) في سياق تدريب الذكاء الاصطناعي للأمن؟

أ. وجود الكثير من البيانات لتخزينها.

ب. وجود بيانات من مناطق زمنية مختلفة جداً.

ج. وجود كمية هائلة من البيانات "الآمنة" وعدد قليل جداً من أمثلة "الهجمات". (✓)

د. وجود بيانات غير مشفرة.

الشرح:

مجموعات بيانات الأمن منحرفة بشكل طبيعي. الهجمات الحقيقية نادرة مقارنة بحركة المرور الخلفية العادية. هذا يجعل من الصعب على نماذج الذكاء الاصطناعي تعلم خصائص الهجوم بشكل فعال دون معالجة خاصة.



القسم 4: تقييم (3/3)

السؤال 3

لماذا يعد "استخراج الميزات" (Feature Extraction) ضرورياً؟

- أ. لتشفير البيانات من أجل السلامة.
- ب. لتحويل السجلات النصية إلى قيم رقمية يمكن للشبكات العصبية معالجتها.
- ج. لضغط البيانات لتوفير مساحة على القرص.
- د. لحذف البيانات بعد قراءتها.



القسم 4: تقييم (3/3) - الإجابة

السؤال 3

لماذا يعد "استخراج الميزات" (Feature Extraction) ضرورياً؟

أ. لتشفير البيانات من أجل السلامة.

ب. لتحويل السجلات النصية إلى قيم رقمية يمكن للشبكات العصبية معالجتها. (✓)

ج. لضغط البيانات لتوفير مساحة على القرص.

د. لحذف البيانات بعد قراءتها.

الشرح:

معظم خوارزميات تعلم الآلة هي دوال رياضية تتطلب مدخلات رقمية. تحتوي السجلات الخام على نصوص (على سبيل المثال، "Login Failed"). يقوم استخراج الميزات بترجمة هذا النص إلى أرقام (على سبيل المثال، عدد مرات الفشل = 1) يمكن للذكاء الاصطناعي حسابها.





مصفوفة الارتباك (The Confusion Matrix)

نقيس نجاح الذكاء الاصطناعي من خلال أخطائه.

آمن فعلياً	خبيث فعلياً	
إيجابي كاذب (ضوضاء)	إيجابي حقيقي (نجاح)	توقع الذكاء الاصطناعي: تهديد
سلبي حقيقي (هدوء)	سلبي كاذب (اختراق)	توقع الذكاء الاصطناعي: آمن

- إيجابي كاذب (النوع الأول) (False Positive (Type I)): إنذار كاذب. يؤدي إلى "إرهاق التنبيهات" لدى المحللين.
- سلبي كاذب (النوع الثاني) (False Negative (Type II)): هجوم مفقود. النتيجة الأكثر خطورة.



الذكاء الاصطناعي كمضاعف للقوة والذكاء الاصطناعي القابل للتفسير (XAI)

مضاعف القوة (Force Multiplier)

يزيد الذكاء الاصطناعي من الإنتاج. يمكن للمحلل مراجعة 500 تنبيه بدلاً من 50 لأن الذكاء الاصطناعي يقوم بتصفية الضوضاء.

مشكلة "الصندوق الأسود" (The "Black Box")

غالباً ما تعطي النماذج المعقدة إجابات بدون أسباب. إذا لم يفهم المحلل لماذا حظر الذكاء الاصطناعي مستخدماً، فلن يثق به.

- **الذكاء الاصطناعي القابل للتفسير (Explainable AI (XAI):** يتضمن أن النماذج تقدم أسباباً (على سبيل المثال، "تم الحظر لأن الموقع تغير بسرعة").
- يقول الذكاء الاصطناعي: "احظر المستخدم أ".
- يسأل المحلل: "لماذا؟"
- الذكاء الاصطناعي: (صمت).



القسم 5: تقييم (1/3)

السؤال 1

في مصفوفة الارتباك (Confusion Matrix)، ما هو "السليبي الكاذب" (False Negative)؟

- أ. ينبه الذكاء الاصطناعي على ملف آمن.
ب. يحدد الذكاء الاصطناعي فيروساً بشكل صحيح.
ج. يفشل الذكاء الاصطناعي في كشف هجوم فعلي (صمت أثناء الاختراق).
د. يتجاهل الذكاء الاصطناعي ملفاً آمناً بشكل صحيح.



القسم 5: تقييم (1/3) - الإجابة

السؤال 1

في مصفوفة الارتباك (Confusion Matrix)، ما هو "السليبي الكاذب" (False Negative)؟

أ. ينبه الذكاء الاصطناعي على ملف آمن.

ب. يحدد الذكاء الاصطناعي فيروساً بشكل صحيح.

ج. يفشل الذكاء الاصطناعي في كشف هجوم فعلي (صمت أثناء الاختراق). (✓)

د. يتجاهل الذكاء الاصطناعي ملفاً آمناً بشكل صحيح.

الشرح:

السليبي الكاذب هو خطأ من النوع الثاني حيث يتنبأ النظام بـ "آمن" ولكن الواقع هو "خبيث". هذا هو السيناريو الأكثر خطورة في الأمن السيبراني لأن الهجوم يستمر دون أن يلاحظه أحد.



القسم 5: تقييم (2/3)

السؤال 2

ما هي الوظيفة الأساسية لمنصة SOAR في مركز العمليات الأمنية؟

- أ. جمع السجلات.
- ب. أتمتة الاستجابات وتنفيذ الإجراءات ("الأيدي").
- ج. تدريب نموذج الذكاء الاصطناعي.
- د. استبدال جميع المحللين البشريين.



القسم 5: تقييم (2/3) - الإجابة

السؤال 2

ما هي الوظيفة الأساسية لمنصة SOAR في مركز العمليات الأمنية؟

أ. جمع السجلات.

ب. أتمتة الاستجابات وتنفيذ الإجراءات ("الأيدي"). (✓)

ج. تدريب نموذج الذكاء الاصطناعي.

د. استبدال جميع المحللين البشريين.

الشرح:

بينما يعمل SIEM كـ "الدماغ" (التحليل)، يعمل SOAR كـ "الأيدي". إنه ينسق الأدوات لاتخاذ إجراءات، مثل حظر عنوان IP تلقائياً على جدار الحماية أو تعطيل حساب مستخدم مخترق.



القسم 5: تقييم (3/3)

السؤال 3

لماذا يعد الذكاء الاصطناعي القابل للتفسير (XAI) حاسماً في مركز العمليات الأمنية؟

- أ. يجعل الذكاء الاصطناعي يعمل بشكل أسرع.
- ب. يسمح للذكاء الاصطناعي بتشفير البيانات بشكل أفضل.
- ج. يبني الثقة عن طريق إخبار المحللين لماذا تم اتخاذ القرار.
- د. يزيل الحاجة إلى تنظيف البيانات.



القسم 5: تقييم (3/3) - الإجابة

السؤال 3

لماذا يعد الذكاء الاصطناعي القابل للتفسير (XAI) حاسماً في مركز العمليات الأمنية؟

أ. يجعل الذكاء الاصطناعي يعمل بشكل أسرع.

ب. يسمح للذكاء الاصطناعي بتشفير البيانات بشكل أفضل.

ج. يبنى الثقة عن طريق إخبار المحللين لماذا تم اتخاذ القرار. (✓)

د. يزيل الحاجة إلى تنظيف البيانات.

الشرح:

يعاني المحللون من مشكلة "الصندوق الأسود" حيث لا يمكنهم التحقق من قرارات الذكاء الاصطناعي. يوفر XAI الأساس المنطقي (على سبيل المثال، "تم الحظر بسبب السفر المستحيل")، مما يسمح للبشر بالثقة والتحقق من الإجراءات الآلية.



- التطور: انتقلنا من التوقعات الثابتة إلى التحليل السلوكي الديناميكي.
- الاستراتيجية: نستخدم نماذج مثل هرم الألم (Pyramid of Pain) والدفاع في العمق (Defense in Depth) لهيكله دفاعاتنا.
- الخصم: فهم سلسلة القتل السيبراني (Cyber Kill Chain) وMITRE ATT&CK يسمح لنا بتسمية البيانات بفعالية.
- البيانات: يتطلب الذكاء الاصطناعي بيانات نظيفة وموحدة واستخراج الميزات (Feature Extraction) ليعمل.
- العمليات: في مركز العمليات الأمنية (SOC)، يوازن الذكاء الاصطناعي بين الإيجابيات الكاذبة والسلبيات الكاذبة، ويعمل كمضاعف للقوة للمحللين البشريين.





II المختصرات والمصطلحات

التزييف العميق Deepfakes هي مقاطع صوتية أو مرئية اصطناعية مولدة بواسطة الذكاء الاصطناعي AI تُستخدم لانتحال شخصية أفراد موثوق بهم، غالباً في مخططات الهندسة الاجتماعية social engineering المتطورة.

الاختبار الضبابي Fuzzing هي طريقة آلية يستخدمها المهاجمون لاكتشاف ثغرات برمجية جديدة عن طريق إدخال بيانات غير صالحة أو غير متوقعة أو عشوائية إلى مدخلات البرنامج.

الشبكات التوليدية التنافسية GANs (Generative Adversarial Networks) هي نوع من نماذج الذكاء الاصطناعي AI يمكن للمهاجمين تدريبها لإنشاء سلاسل جديدة وفعالة من البرمجيات الخبيثة malware تلقائياً تكون غير مرئية للدفاعات القائمة على التوقعات.

أنظمة كشف التسلل IDS (Intrusion Detection Systems) هي أدوات أمنية تراقب أنشطة الشبكة أو النظام بحثاً عن إجراءات ضارة أو انتهاكات للسياسات. غالباً ما تعتمد الإصدارات التقليدية على التوقعات signatures.

مؤشرات الاختراق IOCs (Indicators of Compromise) هي أدلة دقيقة أو أنماط أو بيانات جنائية رقمية forensic data توجد في السجلات أو الملفات وتشير إلى احتمال وقوع خرق أمني.



III المختصرات والمصطلحات

تعلم الآلة (Machine Learning) ML هو مجال فرعي من الذكاء الاصطناعي AI يمكن أنظمة الكمبيوتر من التعلم من البيانات وتحديد الأنماط دون برمجة صريحة. وهو التكنولوجيا الأساسية وراء اكتشاف التهديدات القائم على السلوك.

إطار عمل MITRE ATT&CK هو قاعدة معرفية وإطار عمل متاح عالمياً لتكتيكات وتقنيات الخسوف في العالم الحقيقي، يستخدمه متخصصو الأمن لفهم ونمذجة سلوك المهاجمين.

البرمجيات الخبيثة متعددة الأشكال Polymorphic malware هي برمجيات خبيثة تغير شفرتها أو ميزات باسمرار لإنشاء توقيعات signatures جديدة وفريدة في كل مرة تتكاثر فيها، مما يجعل من الصعب على برامج مكافحة الفيروسات التقليدية اكتشافها.

إدارة معلومات وأحداث الأمن SIEM (Security Information and Event Management) هي منصة مركزية تقوم بجمع وتجميع وتحليل بيانات السجلات والتنبيهات المتعلقة بالأمن من جميع أنحاء البنية التحتية لتكنولوجيا المعلومات في المؤسسة.

تنسيق وأتمتة واستجابة الأمن SOAR (Security Orchestration, Automation, and Response) هي منصة تساعد المؤسسات على الاستجابة للتنبيهات الأمنية بطريقة آلية باستخدام مسارات عمل محددة مسبقاً أو ما يعرف بـ "كتيبات اللعب" playbooks لتنفيذ إجراءات الاستجابة.



المختصرات والمصطلحات IV

التوقيع Signature هو نمط فريد، مثل بصمة رقمية، يرتبط بقطعة معروفة من التعليمات البرمجية الخبيثة أو طريقة هجوم محددة، وتستخدمه الأدوات الأمنية التقليدية للكشف.

هجوم سلسلة التوريد Supply chain attack هي استراتيجية هجوم غير مباشرة حيث يقوم الخصوم باختراق مؤسسة ما عن طريق استهداف العناصر الأقل أماناً في شبكة مورديها، مثل بائع برامج تابع لجهة خارجية.

صيد التهديدات Threat Hunting هي الممارسة الاستباقية للبحث في شبكات وبيانات المؤسسة للعثور على التهديدات المتقدمة التي أفلتت من الدفاعات الأمنية الآلية الحالية.

هجوم يوم الصفر Zero-day attack هو هجوم سيراني يستغل ثغرة برمجية غير معروفة سابقاً لا يتوفر لها تصحيح أو إصلاح، مما يجعله فعالاً للغاية ضد الدفاعات التقليدية.



شكراً لكم!

