CSET CENTER for SECURITY and EMERGING TECHNOLOGY

الأمن السيبراني والذكاء الاصطناعي

لحقوق الطبع والنشر محفوظة © 2021 لجامعة جورجتاون

12 آب 2021

زميل باحث | **Mr. Wyatt Hoffman**

wyatt.hoffman@georgetown.edu | cset.georgetown.edu

نظرة عامة

موجز

- أساسيات (الذكاء الاصطناعي مقابل التعلم الآلي)
- الذكاء الاصطناعي للهجوم السيبراني
- الذكاء الاصطناعي للدفاع السيبراني
- قرصنة الذكاء الاصطناعي
- الأبعاد الاستراتيجية
- توصيات للتعاون

الأسئلة الرئيسية:

- ما الذي يقدمه الذكاء الاصطناعي للأمن السيبراني؟ ما هي حدوده؟
- كيف يمكن للذكاء الاصطناعي إعادة تشكيل مشهد التهديد السيبراني؟
- كيف يمكن للذكاء الاصطناعي تغيير الديناميكيات الاستراتيجية للمنافسة السيبرانية؟

الأساسيات

1: الذكاء الاصطناعي مقابل التعلم الآلي

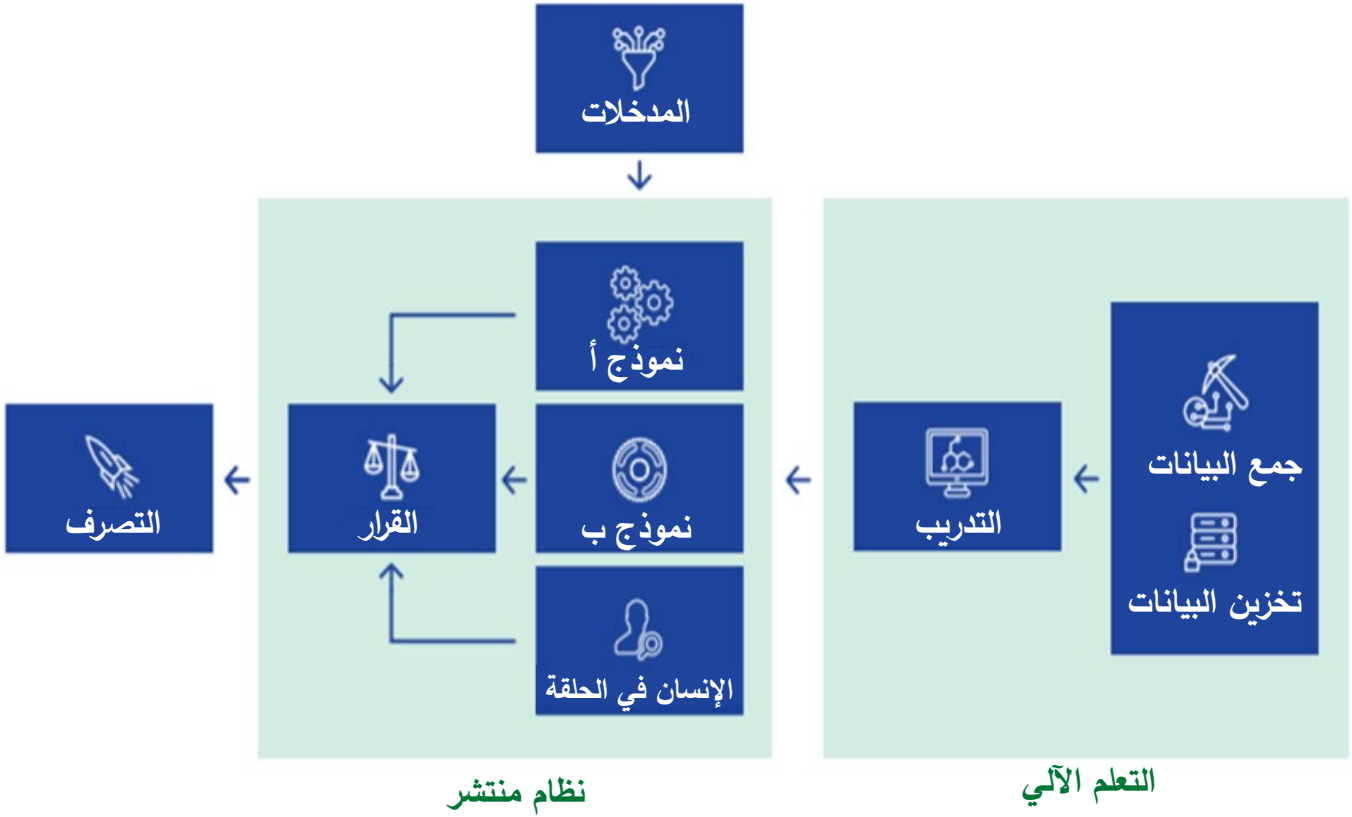
"تستخدم أنظمة التعلم الآلي قوة الحوسبة لتنفيذ الخوارزميات التي تتعلم من البيانات."

– بن بوكانان، "ثالوث الذكاء الاصطناعي وما يعنيه بالنسبة لاستراتيجية الأمن القومي"

المصدر: آرتيم أوبرمان، "الذكاء الاصطناعي مقابل التعلم الآلي مقابل التعلم العميق"، نحو علم البيانات، 29 تشرين الأول/ أكتوبر 2019

الأساسيات

2: كيف يعمل التعلم الآلي



المصدر: أندرو لوهن، "قرصنة الذكاء الاصطناعي: كتاب تمهيدي لصانعي السياسات في مجال التعلم الآلي للأمن السيبراني" CSET، كانون الأول/ ديسمبر 2020

الأساسيات

3: نقاط القوة والقيود في التعلم الآلي

نقاط القوة

- **الأداء الخارق:** يمكن للتعلم الآلي أن يكتشف أنماطًا لا يستطيع أن يلاحظها البشر، وتكون مفيدة لعمل التنبؤات
- **التكيّفية:** يمكن لأنظمة التعلم الآلي أن تستمر في التعلم أثناء نشرها
- **الأتمتة:** يمكن لأنظمة التعلم الآلي أداء المهام التي تتطلب خبرة بشرية

القيود

- **يعتمد على البيانات:** يعتمد النجاح بشكل حاسم على بيانات التدريب العالية الجودة والكمية
- **كثيف الموارد:** التدريب والتشغيل يتطلبان قوة حوسبة كبيرة
- **هَش:** لا يمكن لأنظمة التعلم الآلي أن تجاري بشكل جيد التغيرات البيئية أو المدخلات العدائية التي تنتهك الافتراضات المستفادة من التدريب
- **قابلية التفسير:** أنظمة التعلم الآلي هي "صناديق سوداء" يصعب فهم قراراتها

ك ئ ع ظ ئ ي ك ى ك ذ ر ع لا ج ـ ئ ـ س ئ غ ن ئ.

الذكاء الاصطناعي للجريمة السيبرانية

المصدر: لوكهيد مارتن، "سلسلة القتل السيبراني" https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

التطبيقات على المدى القريب:

- الصيد الآلي لنقاط الضعف
- التصيد الاحتيالي ذو الاستهداف المحدد والهندسة الاجتماعية

المزيد من التكهنات:

- انتشار أذكى
- قدرات أكثر خلسةً ومراوغةً وخبثًا
- عمليات هجومية أكثر قوة

الذكاء الاصطناعي للدفاع السيبراني

المصدر: DARPA، "التحدي السيبراني الكبير" https://www.darpa.mil/program/cyber-grand-challenge

التطبيقات على المدى القريب:

- الاصطياد الآلي لنقاط الضعف
- اكتشاف البرامج الضارة والتطفل من خلال تمكين التعلم الآلي

المزيد من من التكهنات:

- تدابير الدفاع النشطة (مثل مصائد المخترقين القادرة على التكيف)
- تحريك الدفاعات المستهدفة

قرصنة الذكاء الاصطناعي

1: التعلم الآلي الخصومي

نهجان رئيسيان:

- ***تجنب:*** المدخلات الحرفية التي تنتهك افتراضات النموذج
- ***تسمم:*** العبث ببيانات التدريب بهدف تدريب النظام بشكل خاطئ أو وضع باب خلفي

الشكل 3
لم يتأثر تصنيف مبنى هيلي هول التابع لجامعة جورج تاون في الصورة العلوية وتمت مهاجمته ليبدو في نظام التعلم الآلي على أنه ديناصور ثلاثي القرون في الصورة السفلى. بحسب عين الإنسان، تبدو الصورتان متطابقتين.

المصدر: Lohn, “Hacking AI”

قرصنة الذكاء الاصطناعي

2: قرصنة الدفاعات السيبرانية القائمة على التعلم الآلي

Malware	SHA256	Score Before	Score After
CoinMiner	1915126c27ba8566c624491bd2613215021cc2b28e5e6f3af69e9e994327f3ac	-826	884
Dridex	c94fe7b646b681ac85756b4ce7f85f4745a7b505f1a2215ba8b58375238bad10	-999	996
Emotet	b3be486490acd78ed37b0823d7b9b6361d76f64d26a089ed8fbd42d838f87440	-923	625
Gh0stRAT	eebff21def49af4e85c26523af2ad659125a07a09db50ac06bd3746483c89f9d	-975	998
Kovter	40050153dceec2c8fbb1912f8eeabe449d1e265f0c8198008be8b34e5403e731	-999	856
Nanobot	267912da0d6a7ad9c04c892020f1e5757edf9c4762d3de22866eb8a550bff81a	971	999

على سبيل المثال) تم اكتشاف "تحايل عالمي" في محرك مكافحة الفيروسات القائم على التعلم الآلي الذي أنتجته سيلانس.

سكاي لايت سايبر، "سيلانس، أنا أقتلك!
https://skylightcyber.com/2019/07/18/cylance-i-kill-you/

الآثار الاستراتيجية

1: كيف يمكن للذكاء الاصطناعي إعادة تشكيل مشهد التهديد السيبراني؟

يمكن للتعلم الآلي أن يمكِّن المهاجمين، أو يمنح تكافؤ الفرص للمدافعين:

- التطبيقات الهجومية والدفاعية التعويضية عبر "سلسلة القتل".
- يمكن للتعلم الإلكتروني أن يطلق العنان لمزايا دفاعية غير محققة: السيطرة على "الملعب"، والوصول إلى بيانات واسعة عن نشاط الشبكة.
- ولكن الدفاع يواجه تحديات فريدة من نوعها: مخاوف أكبر بشأن الموثوقية، وناقلات الهجوم التي تستهدف التعلم الآلي نفسه.
- وفي أسوأ الحالات، قد يؤجج التعلم الآلي هجمات أكثر خطورة وتدميرًا.

[illegible]

الآثار الاستراتيجية

2: كيف يمكن للذكاء الاصطناعي تشكيل الديناميكيات الاستراتيجية للمنافسة السيبرانية؟

الذكاء الاصطناعي يمكن أن يؤدي إلى زعزعة الاستقرار لعدة أسباب:

- إدخال مخاطر جديدة من الآثار غير المقصودة أو الأضرار الجانبية الناجمة عن القدرات الذاتية.
- تحفيز الحملات السيبرانية الأكثر عدوانية للإضرار بأنظمة التعلم الآلي أو تخريبها (مثل استهداف سلاسل التوريد) أو استهداف الثقة في التعلم الآلي نفسه.
- زيادة مخاطر التصعيد من الاشتباكات السيبرانية (على سبيل المثال، سوء تفسير عملية تجسس على أنها هجوم).
- توسيع نطاق التأثيرات المحتملة من العمليات السيبرانية التي تستهدف قدرات الذكاء الاصطناعي بشكل عام.

توصيات للتعاون

إيجاد حلول مشتركة لتعزيز بيئة تكنولوجيا وآمنة وسو تح

- مشاركة أفضل الممارسات من أجل سلامة وأمن الذكاء الاصطناعي
- التعاون على المتانة العدائية
- تأمين الأساس لتطوير الذكاء الاصطناعي (سلاسل التوريد ومصادر البيانات)

ث الكاك شدنطهد ثقف الكئ الإزئجن كه تملى

- تبادل المعلومات بشأن التهديدات المشتركة (مثل التهديدات الناشئة لنظم المراقبة الصناعية)
- مكافحة انتشار القدرات الهجومية
- الأعراف الدولية للعمليات السيبرانية الهجومية

لمزيد من القراءة

- **أتمتة الهجمات السيبرانية:** الضجيج والواقع من قبل بن بوكانان، جون بانسيمر، داكوتا كاري، جاك لوكاس وميخا موسر
- **العمليات السيبرانية المدمرة والتعلم الآلي** من تأليف داكوتا كاري ودانيال سيبول
- **التعلم الآلي والأمن السيبراني:** الضجيج والواقع من قبل ميكا موسر وأشتون غاريوت
- **قرصنة الذكاء الاصطناعي: كتاب تمهيدي لصانعي السياسات حول الأمن السيبراني الخاص بالتعلم الآلي** من تأليف أندرو لوهن
- **الذكاء الاصطناعي ومستقبل المنافسة السيبرانية** من قبل وايت هوفمان

متوفر على: cset.georgetown.edu

CSET CENTER for SECURITY and EMERGING TECHNOLOGY

- ابحث في https://cset.georgetown.edu/research/
- اشترك لتلقي الأبحاث في اليوم الذي تصدر فيه، واشترك في نشرتنا الإخبارية التي تصدر كل أسبوعين، واحصل على دعوة لحضور فعالياتناعلى https://cset.georgetown.edu/sign-up/
- شاهد ندوات مركز الأمن والتكنولوجيا الناشئة CSET على الويب واطلب إرشادات، إذا لزم الأمر. شارك الأسئلة والثغرات المعرفية الخاصة بك