باحثون يكتشفون أنّ أدواتِ الحفاظ على الخُصوصيّة تترك البيانات الخاصة بدونِ حماية



لم تعد أنظمة التعلم الآلي (ML) منتشرةً فقط في التقنيات التي تؤثر على حياتنا اليومية، بل أيضًا بتلك التي تراقبها، بما في ذلك أنظمة التّعرف على تعبيرات الوجوه.

تعتمد الشركات التي تَصنع وتستخدم مثل هذه الخدمات المنتشرة على نطاق واسع، على ما يُسمّى بأدواتِ الحفاظ على الخُصوصيّة التي غالبًا ما تَستخدم شبكات الخصومة التوليدية (**GANs**)، والتي يُنتِجها عادةً طرفٌ ثالثٌ لتنظيف صور هُويّة الأفراد.

لكن ما مدى جَودتها؟
وجد الباحثون في كليّة (تاندون) للهندسة بجامعة نيويورك **NNY** والذين اكتشفوا أُطرَ التعلّم الآلي وفقًا لهذه الأدوات، أنّ الإجابة "ليس كثيرًا" في الورقة البحثية "تخريب الخصوصية للحفاظ على شبكات **GANs**: إخفاء الأسرار في الصور النّظيفة"، والتي قُدّمت في شهر شباط/فبراير الماضي في مؤتمر **AAAI** الخامس والثلاثين للذكاء الاصطناعي.

اكتشف فريق بقيادة سيدهارث جارج **Siddharth Garg**، الأستاذ المساعد في معهد الهندسة الكهربائيّة وهندسة الكمبيوتر في جامعة نيويورك ـكليّة تاندونـ أنّه لا يزال من الممكن استعادة البيانات الخاصة من الصور التي "نُظِّفت" بواسطة مُميّزات التعلّم العميق، مثل حماية الخصوصية لشبكات **GANs** وحتى التي اجتازت الاختبارات التجريبية.

وجد الفريق بما في ذلك المؤلف الرئيسي كانغ ليو **Kang Liu** المرشح للدكتوراه، وبنجامين تان **Benjamin Tan**، الأستاذ المساعد ـ باحث في الهندسة الكهربائية وهندسة الكمبيوترـ أنّ تصميمات حماية الخصوصيّة لشبكات **GAN** يمكن تخريبها لاجتياز اختبارات الخصوصيّة، مع السماح باستخراج المعلومات السِّرية من الصّور النّظيفة.

تتمتّع أدوات الخصوصيّة المُستندة إلى التعلّم الآلي بإمكانية تطبيقٍ واسعة، من المُحتمل أن تكون في أيِّ مجالٍ حسّاسٍ للخصوصيّة، بما

في ذلك إزالة المعلومات ذات الصلة بالموقع من بيانات كاميرا المركبات، أو التعتيم على هُوية الشّخص الذي أنتج عيّنة بخط اليد، أو إزالة الرموز الشريطية من الصور.

لتصميم وتدريب الأدوات المستندة إلى **GAN** يتم الاستعانة بمصادرَ خارجيّة للبائعين بسبب التعقيد الذي ينطوي عليه الأمر .قال جارج: "تُستخدم العديد من أدوات الطرف الثالث لحماية خصوصية الأشخاص الذين قد يظهرون على كاميرا مراقبة أو كاميرا لجمع البيانات، حماية الخصوصية لشبكات **GAN** هذه للتلاعب بالصورة. صُمّمت إصدارات من هذه الأنظمة لتنظيف صور الوجوه وغيرها من البيانات الحسّاسة، إذ يتم الاحتفاظ بالمعلومات الهامة عن التطبيق فقط. وبينما اجتاز **PP-GAN** العدائي جميع اختبارات الخصوصية الحالية، اكتشفنا أنّه أخفى بالفعل بيانات سِرّية تتعلّق بمعلوماتٍ للسّمات الحسّاسة، حتى أنّها تسمح بإعادة بناء الصورة الأصلية الخاصة".

تُوفّر الدراسة معلوماتٍ أساسيةً عن شبكات حماية الخصوصيّة **PP-GAN** وما يرتبط بها من فحوصات الخصوصيّة التجريبية، كما تُصيغ سيناريو هجوم للسؤال عمّا إذا كان من المُمكن تخريب عمليّات التحقّق التجريبية للخصوصية، وتُحدّد نهجًا للتحايُل على فحوصات الخصوصيّة التجريبية.

يقدّم الفريق أوّلُ تحليلٍ أمني شامل لشبكات **GAN** التي تحافظ على الخصوصية، ويوضّح أنّ فُحوصات الخصوصيّة الحاليّة غير كافية لاكتشاف تسرُّب المعلومات الحسّاسة.

باستخدام نهج إخفاء جديد، قاموا بتعديل **PP-GAN** بشكل عكسي لإخفاء سر (معرف المستخدم)، من صور الوجه المُنظّفة المزعومة. لقد أظهروا أنّ **PP-GAN** العدائي المُقترح يُمكنه إخفاء السّمات الحسّاسة بنجاحٍ في صور الإخراج النّظيفة التي تجتاز فحوصات الخصوصية، بمُعدّل استرداد سِري بنسبة 100%.

لاحظَ جارج ومُعاونوه أنّ المقاييس التجريبية تعتمد على قدرات التعلّم لدى المميّزين وميزانيات التدريب، وقالوا إنّ فحوصات الخصوصية هذه تفتقر إلى الصرامة اللازمة لضمان الخصوصيّة. وأوضح جارج: "من وجهة نظرٍ عمليّة، تبدو نتائجنا بمثابة ملاحظة تحذيريّة ضد استخدام أدوات تنظيف البيانات، وتحديدًا شبكات حماية الخصوصية **PP-GAN**، المُصمَّمة من قبل أطراف ثالثة. أبرزت نتائجنا التجريبية عدم كفاية فُحوصات الخُصوصيّة القائمة على **DL** والمخاطر المحتملة لاستخدام أدوات **PP-GAN** من جهاتٍ خارجيّة غيرِ موثوق بها".

- التاريخ: 2021-04-08
- التصنيف: تكنولوجيا

#تكنولوجيا #الخصوصية

المصادر

- techxplore.com

المساهمون

- ترجمة
  - لبنى جمعة
- مُراجعة

- هبة العيوطي
- تحرير
  - عبد الفتاح أنور
- نشر
  - احمد صلاح