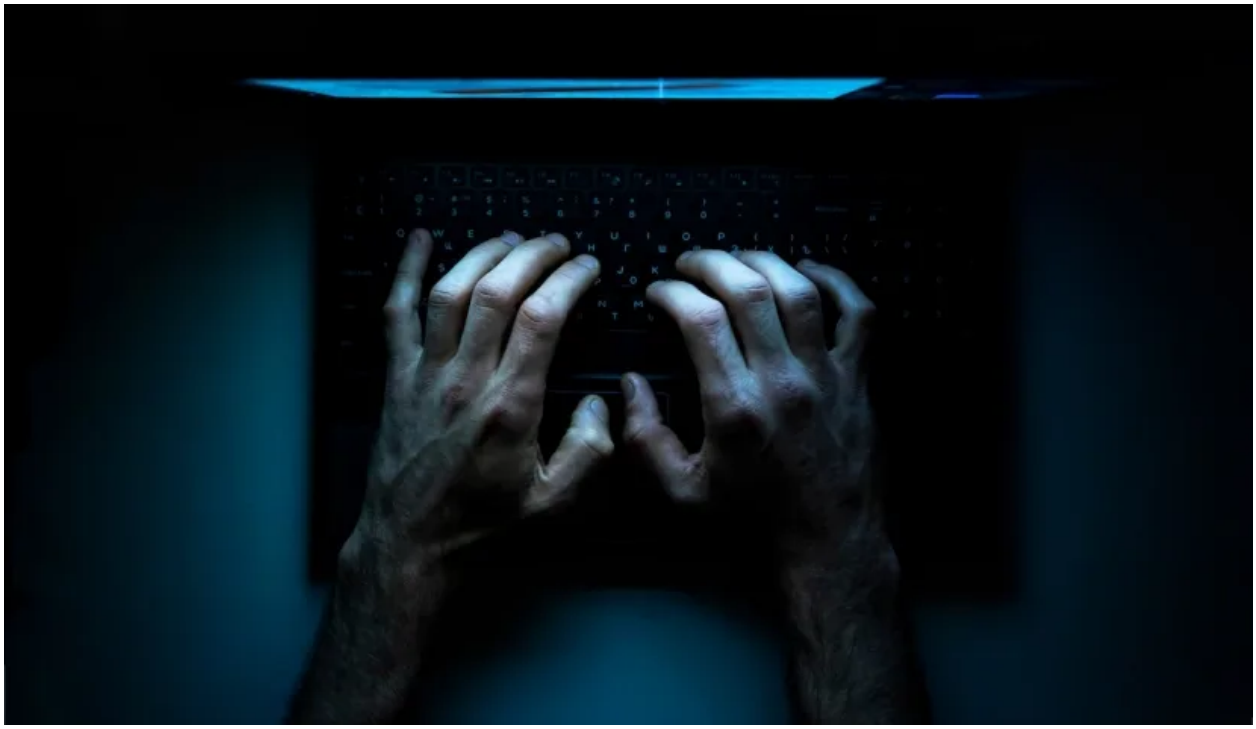
كابوس.. الذكاء الاصطناعي يولّد صورًا تنتهك براءة الأطفال

كتبه درو هارويل | 19 يوليو ,2023



ترجمة حفصة جودة

أثارت ثورة الذكاء الاصطناعي موجة من الصور النابضة بالحياة التي تُظهر الاستغلال الجنسي للأطفال، ما أثار قلق المحققين بشأن سلامة الأطفال خوفًا من أن يقوض ذلك جهود العثور على الضحايا وملاحقة المجرمين في العالم الحقيقي.

ساهمت أدوات الذكاء الاصطناعي التوليدي في إنتاج ما أطلق عليه أحد المحللين "سباق التسلح المفترس" في منتديات البيدوفيليا "الاستغلال الجنسي للأطفال" لأن باستطاعتهم خلق صورة حقيقة لأطفال يقومون بأفعال جنسية خلال ثوانٍ، تُعرف باسم "المواد الإباحية المتعلقة بالأطفال".

وُجدت الآلاف من الصور الجنسية للأطفال والمولدة عن طريق الذكاء الاصطناعي على منتديات في "الشبكة المظلمة" – وهي طبقة من الإنترنت تظهر فقط من خلال متصفحات خاصة – كما يقدم بعض المشاركين توجيهات تفصيلية لمساعدة المنجذبين جنسيًا للأطفال على صنع صورهم الخاصة.

تقول ريبيكا بورتنوف، مديرة علوم البيانات بمنظمة "Thorn" – وهي منظمة غير ربحية مهتمة بأمن الأطفال – "يُعاد استخدام صور الأطفال بما في ذلك محتوى ضحايا معروفين للحصول على هذه المخرجات الشريرة".

وتضيف "يعد التعرف على الضحايا مشكلة عويصة بالفعل، حيث تحاول قوات إنفاذ القانون العثور على الأطفال المتضررين، لكن تسهيل استخدام هذه الأدوات يمثل تحولًا بارزًا، فهو يجعل المشكلة أكثر تحديًا".

هذا السيل من الصور سيربك أنظمة التتبع المركزية المبنية بغرض حجب هذه المواد على الإنترنت، لأنها مصممة لاكتشاف صور الانتهاكات المعروفة وليس اكتشاف الصور الجديدة المولدة عن طريق الذكاء الاصطناعي.

كما يهدد ذلك أيضًا بإرهاق موظفي إنفاذ القانون الذين يعملون على التعرف على الضحايا الأطفال، حيث سيضطرون لقضاء مزيد من الوقت لتحديد إذا ما كانت الصورة حقيقية أم مزيفة.

أثارت الصور نقاشًا أيضًا إذا ما كانت تنتهك قوانين حماية الأطفال الفيدرالية لأنها تصور أطفالًا غير موجودين من الأساس، قال مسؤولون من وزارة العدل الذين يكافحون استغلال الأطفال إن هذه الصور غير قانونية أيضًا حتى لو كانت مولدة عن طريق الذكاء الاصطناعي، لكن حتى الآن لم يُتهم أي شخص بصنع مثل هذه الصور.

تسمح أداوت الذكاء الاصطناعي الجديدة المعروفة باسم "نماذج الانتشار" لأي شخص بإنشاء صورة فقط من خلال كتابة وصف قصير لما يود أن يراه، هذه النماذج مثل "DALL-E, Midjourney" و"Stable Diffusion" تمت تغذيتها بملايين الصور من الإنترنت التي تُظهر بعضها صورًا لأطفال حقيقيين ومن مواقع الصور والمدونات الشخصية، هذه البرامج تحاكي هذه النماذج البصرية لصنع صورتها الخاصة.

احتفى الجميع بهذه الأدوات لابتكاراتها البصرية واستُخدمت للفوز في مسابقات الفنون الجميلة وزخرفة كتب الأطفال وابتكار صور إخبارية مزيفة، وكذلك توليد صور إباحية لشخصيات تشبه البالغين لكنها غير حقيقية.

> وظف مختبر الأبحاث "OpenAI" المسؤول عن "Dall-E" و"ChatGPT" مراقبين بشريين لتطبيق القواعد بما في ذلك حظر أي مواد تنتهك الأطفال جنسيًا وحذفوا أي محتوى صريح من بيانات تدريب مولد الصور للحد من تعرضه لتلك المفاهيم

لكن ذلك ساهم أيضًا في تمكن البيدوفيليين من صنع المزيد من الصور التفصيلية لأن الأدوات الجديدة أقل تعقيدًا من الماضي، فالآن باستطاعتهم تركيب وجوه الأطفال على أجساد البالغين وتوليد العديد من الصور بأمر واحد فقط.

لكن لا يبدو واضحًا دائمًا على منتديات البيدوفيليا كيف تُصنع هذه الصور، لكن الخبراء المعنيين بأمن الأطفال يقولون إن بعضها يبدو معتمدًا على أداوت مفتوحة المصدر مثل "Stable

Diffusion" التي يمكن استخدامها دون أي قيود.

قالت شركة "Stability AI" التي تدير "Stable Diffusion" في بيان لها إنها حظرت صنع صور أطفال جنسية، وإنها تساعد تحقيقات إنفاذ القانون بشأن الاستخدامات غير القانونية والخبيثة، وقد حذفت بعض المواد من بيانات التدريب الخاصة بها للحد من القدرة على توليد هذا المحتوى المشين.

لكن بإمكان أي شخص تحميل البرنامج على جهازه الخاص واستخدامه بأي طريقة يريدها دون إشراف الشركة، فرخصة الأداة مفتوحة المصدر تطلب من المستخدمين عدم استغلال القاصرين وإيذائهم بأي شكل، لكن من السهل تجاوز ذلك ببعض سطور من البرمجة التي يستطيع المستخدم إضافتها للبرنامج.

ناقش مختبرو البرنامج لعدة أشهر مخاطر استخدام الذكاء الاصطناعي لمحاكاة وجوه وأجساد الأطفال، وقال أحد المعلقين إنه رأى أحدهم يستخدم البرنامج لمحاولة توليد صور بملابس السباحة لمثلة طفلة.

لكن الشركة دافعت عن منهجيتها مفتوحة المصدر وأهميتها لحرية إبداع المستخدمين، قال المدير التنفيذي للشركة عماد مستقي: "في النهاية إنها مسؤولية الناس أن يكونوا أخلاقيين وملتزمين بالقانون عند استخدامهم التكنولوجيا، وهذه الأشياء السيئة التي يصنعها الناس ستكون جزءًا صغيرًا فقط من إجمالي الاستخدام".

أما منافسا "Stable Diffusion" الرئيسيان: "Dall-E" و"Midjourney" فقد حظرا المحتوى الجنسي ولم يقدما مصدرًا مفتوحًا، ما يعني أن استخدامهما محدودًا بالقنوات التي تديرها الشركة وكل الصور مسجلة ومتتبعة.

وظف مختبر الأبحاث "OpenAI" المسؤول عن "Dall-E" و"ChatGPT" مراقبين بشريين لتطبيق القواعد بما في ذلك حظر أي مواد تنتهك الأطفال جنسيًا وحذفوا أي محتوى صريح من بيانات تدريب مولد الصور للحد من تعرضه لتلك المفاهيم.

> انتشار هذه الصور يهدد بإضاعة وقت المحققين الذين يعملون على تحديد الضحايا الحقيقيين من الأطفال

تقول كيت كلونيك أستاذ القانون بجامعة سانت جون: "لا تود الشركات الخاصة أن تكون جزءًا من إنشاء أسوأ محتوى على الإنترنت، لكن ما يخيفني أكثر هو الإصدار المفتوح لتلك الأدوات، حيث يمكن للأفراد أو الشركات غير الموثوقة أن تستخدمهم ثم تختفي تمامًا، وليس هناك طريقة بسيطة ومنظمة للإطاحة بهولاء الفاعلين الفاسدين".

قال أحد محللي سلامة الأطفال إن المستخدمين على منتديات البيدوفيليين في الشبكة المظلمة يناقشون إستراتيجيات إنشاء صور صريحة ومراوغة المرشحات المضادة للإباحية وذلك باستخدام لغات غير الإنجليزية، ما يجعلهم أقل عرضة للقمع أو الكشف.

يقول آفي جاغر، رئيس قسم سلامة الأطفال والاستغلال البشري بمنظمة "ActiveFence"، إن أحد المنتديات الذي يضم 3000 عضو أجرى تصويتًا داخليًا مؤخرًا وقد قال 80% من المستجيبين إنهم استخدموا أو ينوون استخدام أدوات الذكاء الاصطناعي لإنشاء صور تنتهك الأطفال جنسيًا.

ناقش أعضاء المنتدى طرق إنشاء صور ذاتية عن طريق الذكاء الاصطناعي وبناء شخصيات وهمية في عمر المدرسة للفوز بثقة الأطفال، قالت بورتنوف إن فريقها اكتشف حالات استُخدمت فيها صور حقيقية لأطفال منتهكين لتدريب أدوات الذكاء الاصطناعي لخلق صور تُظهر هؤلاء الأطفال في أوضاع جنسية.

تقول يوتا سوراس، المديرة القانونية للمركز الوطني للأطفال المُستغلين والمفقودين – وهو مركز غير ربحي يدير قاعدة بيانات تستخدمها الشركات لتحديد وحظر المواد الجنسية للأطفال – إن فريقها لاحظ زيادة حادة في تقارير الصور المولدة بالذكاء الاصطناعي خلال الأشهر الماضية، وكذلك تقارير لأشخاص يرفعون صورًا جنسية للأطفال على أدوات الذكاء الاصطناعي لتوليد المزيد.

ورغم أنها جزء صغير من 32 مليون تقرير استلمه المركز العام الماضي، فإن انتشار هذه الصور يهدد بإضاعة وقت المحققين الذين يعملون على تحديد الضحايا الحقيقيين من الأطفال.

كما قال مكتب التحقيقات الفيدرالي إنه لاحظ زيادة التقارير المتعلقة بالأطفال الذين حُولت صورهم إلى صور جنسية تبدو حقيقية، تقول سوراس: "بالنسبة لقوات إنفاذ القانون، كيف يحددون أولوياتهم؟ وفي أي شيء يحققون؟ ما وضع هذه الصور في النظام القانوني؟".

قال بعض المحللين القانونيين إن هذه المواد تقع في منطقة قانونية رمادية، فهذه الصور لا تصور أطفالًا حقيقيين تعرضوا للأذى، في عام 2002 ألغت المحكمة العليا بندين من حظر الكونغرس عام 1996 للمواد الإباحية الافتراضية للأطفال، بدعوى أن الصياغة مطاطية وقد تجرم بعض التصوير الأدبي للنشاط الجنسي للمراهقين.

قال المدافعون عن الحظر في ذلك الوقت إن الحكم سيجعل الأمور أصعب في القضايا التي تتضمن إساءة جنسية للأطفال حيث قد يزعم المجرمون أن هذه الصور ليست حقيقية.

> حتى لو كانت هذه الصور غير حقيقية فإنها تمثل ضررًا مجتمعيًا وتساعد في تطبيع الانتهاكات الجنسية للأطفال

لكن القاضي ويليام رينكويست قال: "من مصلحة الكونغرس ضمان القدرة على فرض الحظر على

المواد الإباحية الفعلية للأطفال، ويجب علينا أن نذعن لتلك النتائج التي تقول بأن التطور التكنولوجي القادم سيجعل الأمر مستحيلًا".

يقول دانيل ليونز أستاذ القانون بكلية بوسطن إن الحكم يستحق المراجعة نظرًا للتطور التكنولوجي الهائل في العقدين الماضيين، ويضيف "في ذلك الوقت كان من الصعب إنتاج صور افتراضية لمواد تنتهك الأطفال جنسيًا بطريقة تبدو حقيقية، لكن هذه الفجوة أصبحت ضيقة، وتحولت من قضية تجريبية إلى مشكلة واقعية".

قال مسؤولان من قسم استغلال الأطفال بوزارة العدل إن الصور غير قانونية بموجب القانون الذي يحظر أي صور مصنوعة عن طريق الحاسب الآلي التي تتضمن استغلالًا جنسيًا ولا يمكن تمييزها عن الصور الحقيقية.

وقال رئيس القسم إننا لن نتردد في مقاضاة من ينتج أي صور تستغل الأطفال في أعمال جنسية حتى لو كانت صورًا غير حقيقية، ومن المتوقع مناقشة هذه المشكلة في جلسة تدريب وطنية الشهر المقبل.

بشكل منفصل، تقول مارجريت ميتشيل باحثة الذكاء الاصطناعي التي قادت سابقًا فريق الذكاء الاصطناعي الأخلاقي بغوغل، إن بعض المجموعات تعمل على طرق تقنية لمواجهة تلك القضية.

أحد الحلول – التي تتطلب موافقة حكومية – تدريب نماذج الذكاء الاصطناعي على إنشاء نماذج مزيفة لصور استغلال الأطفال لتتمكن الأنظمة بعد ذلك من كشفها وإزالتها، لكن هذه الفكرة تصاحبها تكلفة نفسية هائلة.

يعمل باحثون آخرون في الذكاء الاصطناعي على أنظمة تماثل تستطيع طباعة رمز على الصور للإشارة إلى منشئيها وذلك للحد من الإساءات، وقد نشر باحثون من جامعة ميرلاند تقنية جديدة لعلامة مائية غير مرئية تساعد في تحديد هوية صانع الصورة ومن الصعب إزالتها.

تقول ميتشل: "هذه الأفكار تحتاج إلى مشاركة صناعية واسعة لتنجح ومع ذلك لن تتمكن من تحديد كل انتهاك، فنحن أشبه بالذين يبنون طائرة في أثناء الطيران بها بالفعل".

بينما تقول سوراس، إنه حتى لو كانت هذه الصور غير حقيقية فإنها تمثل ضررًا مجتمعيًا وتساعد في تطبيع الانتهاكات الجنسية للأطفال، كما أن هذه الصور تُستخدم لإغواء الأطفال في الواقع، وهو سلوك مدمر للغاية".

المصدر: واشنطن بوست

رابط المقال : https://www.noonpost.com/47545/