«التزييف العميق» يخرج عن السيطرة؟ هل فات الأوان؟

يوماً بعد يوم، يغرق الفضاء الرقمي بمقاطع فيديو وصور ورسائل صوتية تبدو حقيقية، ولكنها في الواقع نتاج أدوات ذكاء اصطناعي بارعة في التقليد والخداع. فمع التطور المتسارع لتقنيات الذكاء الاصطناعي أصبحت القدرة على تزوير الوقائع أكثر سهولة وخطورة.

وما كان يتطلب في السابق، تقنيات متقدمة واستوديوهات متخصصة، أصبح اليوم في متناول أي شخص يملك حاسوباً واتصالاً بالإنترنت. فقد بات تزوير الحقيقة ممكناً ببضع نقرات فقط. وهذه السهولة في التلاعب بالصور والفيديوهات، ساهمت مؤخراً في زجّ مشاهير في مواقف لم يعشوها، وإظهار شخصيات عامة وهي تُدلي بتصريحات لم تصدر عنها قط. فضلاً عن تشويه سمعة أفراد عاديين من خلال صور عارية أو مقاطع فيديو مفبركة لا تمتّ للواقع بصلة.

ورغم أن الحكومات حول العالم بدأت تستشعر خطورة هذه الموجة، وتُسرّع في محاولات ضبطها قانونياً وتكنولوجياً، إلا أن تقنية «التزييف العميق» أو الـ Deepfake التي تغذيها أدوات الذكاء الاصطناعي تتحرك بخطى أسرع بكثير من قدرة الحكومات على اللحاق بها. حيث تُظهر بيانات شركة Signicat للتحقق من الهوية، أن محاولات الاحتيال باستخدام مقاطع الفيديو المزيّفة، ازدادت بأكثر من ٢٠ ضعفاً في السنوات الثلاث الماضية. في حين حذّر مكتب التحقيقات الفيدرالي (FBI) مؤخراً، من أن استخدام المجرمين للذكاء الاصطناعي التوليدي لتنفيذ مخططات مالية احتيالية، أصبح منتشراً على نطاق واسع.

كيف انطلقت تقنية «التزييف العميق»؟

بحسب تقرير أعدته «بلومبرغ» واطلع عليه موقع «اقتصاد سكاي نيوز عربية»، فقد كانت تقنية التزييف العميق في البداية حكراً على الأكاديميين والباحثين، ولكن في عام ٢٠١٧ قام أحد الأشخاص بابتكار خوارزمية لإنشاء مقاطع فيديو مزيّفة، باستخدام شيفرة برمجية مفتوحة المصدر، ونشرها عبر موقع Reddit، حيث قامت إدارة الموقع بحظر هذا المستخدم لاحقاً، إلا أن ما قام به كان قد انتشر على نطاق واسع.

وفي بداياتها، كانت تقنية «التزييف العميق» تعتمد على وجود فيديو حقيقي وأداء صوتي أصلي، إضافة إلى مهارات تحرير متقدمة لتحقيق نتائج مقنعة، ولكن مع بروز أنظمة الذكاء الاصطناعي التوليدي في أواخر عام ٢٠٢٢، أصبح من الممكن لأي مستخدم إنتاج صور ومقاطع فيديو مزيفة بدرجة عالية من الإتقان، وذلك بمجرد إدخال أوامر نصية بسيطة، حيث أن كل ما يتطلبه الأمر اليوم هو تحديد المشهد أو الشخصية أو السياق، ليتكفّل الذكاء الاصطناعي بإنشاء المحتوى المطلوب خلال ثوانٍ.

كيف تُصنع مقاطع الفيديو المفبركة؟

غالباً ما تُصمَّم مقاطع الفيديو المُفبركة باستخدام خوارزمية ذكاء اصطناعي يتم تدريبها على تسجيلات حقيقية لشخص معيّن، في عملية تُعرف باسم «التعلّم العميق» أو Deep Learning، ومن خلال هذه التقنية، يصبح بالإمكان استبدال عناصر في الفيديو الأصلي كوجه الشخص مثلاً بمحتوى مُركّب، دون أن يبدو التعديل واضحاً للعين المجردة.

وتزداد خطورة هذه المقاطع عندما تُستخدم بالتوازي مع تقنيات استنساخ الصوت، حيث يُدرّب الذكاء الاصطناعي أيضاً على نبرة صوت الشخص وطريقة نطقه، ليبدو وكأن الكلام المنطوق في الفيديو صادراً عنه بالفعل، لتكون النتيجة مقطع فيديو يبدو حقيقياً من حيث الصورة والصوت، لكنه في الواقع زائف بالكامل.

كيف يمكن اكتشاف عمليات التزوير؟

بات اكتشاف التزوير الرقمي أكثر تعقيداً مع إطلاق شركات الذكاء الاصطناعي أدوات توليد متقدمة، إلا أن بعض المؤشرات لا تزال تكشف أحياناً عن الطابع الاصطناعي للمحتوى، فقد تظهر تشوهات جسدية واضحة، مثل وجود أطراف في غير مواضعها أو يد تحتوي على عدد غير طبيعي من الأصابع.

كذلك، يمكن ملاحظة تباين غير متناسق في ألوان الأجزاء المُعدلة مقارنة ببقية الصورة، أو وجود خلل في تزامن الصوت مع حركة الفم في مقاطع الفيديو المُزيّفة، في حين تبقى التفاصيل الدقيقة مثل خصل الشعر، وتعابير الفم والظلال، من التحديات التي لم تتقنها بعد أنظمة «التزييف العميق»، ما يؤدي أحياناً إلى ظهور حواف مشوّهة أو غير طبيعية للعناصر داخل الصورة أو الفيديو.

أمثلة بارزة على التزييف العميق

في أغسطس ٢٠٢٣، قام قراصنة من الصين بنشر صور مُعدّلة لحرائق الغابات في جزيرة ماوي بولاية هاواي، زاعمين أنها نتيجة لاختبار «سلاح طقس» سري تابع للولايات المتحدة، في محاولة لنشر الفوضى وتعزيز نظريات المؤامرة. وفي مايو من العام نفسه، تراجعت الأسواق الأميركية بشكل مؤقت بعد انتشار صورة مفبركة تُظهر انفجاراً قرب مبنى البنتاغون، قبل أن يتضح لاحقاً أنها صورة مزيفة تم إنشاؤها باستخدام الذكاء الاصطناعي.

وفي يناير ٢٠٢٤، تعرّضت نجمة البوب العالمية تايلور سويفت لحملة تشويه ممنهجة، بعد انتشار صور مُلفّقة لها على نطاق واسع، ما أثار موجة غضب بين معجبيها وقلقاً رسمياً وصل إلى البيت الأبيض. أما خلال الحملة الرئاسية الأميركية لعام ٢٠٢٤، فقد انتشر مقطع فيديو مفبرك يُظهر المرشحة الديمقراطية كامالا هاريس وهي تصف الرئيس جو بايدن بـ«المُخرِف»، وتقول إنها لا تعرف شيئاً عن إدارة البلاد، حيث حصد هذا الفيديو عشرات الملايين من المشاهدات، مثيراً جدلاً واسعاً حول قدرات الذكاء الاصطناعي على التأثير في الرأي العام والعملية الانتخابية.

ما الذي يُبذل لمكافحة التزييف العميق؟

في ١٩ مايو ٢٠٢٥، وقّع الرئيس الأميركي دونالد ترامب قانون « Take It Down» أو «قانون إزالة المحتوى»، الذي يُجرّم المواد الإباحية المُولّدة بالذكاء الاصطناعي دون موافقة الطرفين، ويُجبر شركات التواصل الاجتماعي على حذف هذا المحتوى عند الطلب.

كما أنه في العام الماضي، منعت لجنة الاتصالات الفيدرالية الأميركية، الشركات من استخدام الأصوات المُولّدة بالذكاء الاصطناعي في المكالمات الآلية، وذلك بعد أن تلقى سكان نيو هامبشاير مكالمة آلية، قبل الانتخابات التمهيدية الرئاسية للولاية، بدا فيها صوت بايدن وكأنه يحثّهم على البقاء في منازلهم و«ادخار أصواتهم لانتخابات نوفمبر».

أما في أوروبا، فيُلزم قانون الذكاء الاصطناعي للاتحاد الأوروبي المنصات، بتصنيف مقاطع الفيديو المُزيّفة بعمق، وقد طبّقت الصين تشريعاً مُماثلاً في عام ٢٠٢٣.

الأسواق تحت تهديد

ويقول الكاتب والمحلل التكنولوجي المختص بالذكاء الاصطناعي ألان القارح، في حديث لموقع «اقتصاد سكاي نيوز

عربية»، إن الخطر الحقيقي لا يكمن فقط في قدرة «التزييف العميق» على إنتاج صور ومقاطع فيديو غير موجودة، بل في احتمال أن تصل هذه التقنية إلى مستوى من الإتقان يجعل من المستحيل تقريباً على العين البشرية، وحتى على بعض البرامج التكنولوجية اكتشاف حقيقتها، وعندها سيدخل العالم في مرحلة «الحقيقة البديلة» أي الحقيقة القابلة للتلاعب بالكامل، محذراً من أن المحتوى المزيف عالي الجودة، يمكنه التسبب في تقلبات في الأسواق إذا ما نُشر في توقيت معين، فمثلاً قد يتم نشر فيديو مزيف لرئيس بنك مركزي يُعلن فيه عن رفع مفاجئ للفائدة، كما تم نشر مقطع فيديو يبدو فيه رئيس شركة كبرى وهو يعلن إفلاس شركته.

الفئات الأكثر تعرّضاً

ويؤكد القارح أن ما تم ذكره ليس مجرد سيناريوهات افتراضية، بل هي احتمالات واقعية في ظل التطور السريع لأدوات الذكاء الاصطناعي التي لا تتطلب موارد هائلة أو مهارات تقنية فائقة، مشدداً على أن السياسيين وقادة الأعمال والمشاهير هم في صدارة الفئات المعرضة للخطر من «التزييف العميق»، نظراً لتوفر كمّ هائل من التسجيلات والصور العائدة لهم عبر شبكة الإنترنت، ما يُسهّل على الأنظمة توليد محتوى مزيف عنهم بدقة عالية، ولكن يجب أيضاً عدم تجاهل الخطر الذي تشكله هذه الأدوات على الأفراد العاديين، خصوصاً لناحية استخدام وجوههم في تركيب صور مُخلة بالآداب دون علمهم أو موافقتهم.

تحركات لمواجهة الأسوأ

ويشرح القارح أن «التزييف العميق» بات يُستخدم أيضاً في إنشاء مستندات مكتوبة، وتوقيعات رقمية وحتى محادثات نصية مُقنعة، فهناك حالات تم فيها استخدام دردشات مزورة عبر البريد الإلكتروني، لخداع شركات وتحويل أموال إلى حسابات وهمية، لافتاً إلى أن بعض الحكومات بدأت تتحرّك لسنّ قوانين تحظر المحتوى الذي تم تركيبه عبر «التزييف العميق» ولكن تبقى قدرة التشريعات على مواكبة سرعة التطور التقني محدودة، والأسوأ من ذلك أنه بمجرد انتشار مقطع مزيف عبر الإنترنت، فإن احتواءه يصبح شبه مستحيل، حتى لو ثُبت لاحقاً أنه غير حقيقي.

وبحسب القارح فقد تسارعت مؤخراً الجهود التي تقوم بها شركات التقنية الكبرى والمؤسسات البحثية، لتطوير أدوات متقدّمة قادرة على كشف عمليات «التزييف العميق»، إذ أطلقت شركات مثل مايكروسوفت وغوغل وميتا، أدوات تعتمد على الذكاء الاصطناعي، مصممة لتحليل الإشارات الرقمية الخفية التي تخلّفها عمليات التزوير، مثل التناقضات في الظلال، أو التكرار غير الطبيعي للبكسلات، وفي السياق ذاته بدأت بعض الوكالات الحكومية والمعاهد الأكاديمية، ببناء قواعد بيانات لتدريب أنظمة الكشف عن التزييف، وجعلها أكثر دقة مع الوقت، مشيراً إلى أن عام ٢٠٢٤ شهد أيضاً إطلاق تحالفات تقنية عالمية تهدف إلى وضع علامات مائية رقمية «ديجيتال ووترماركس» داخل الصور والفيديوهات المُولّدة بالذكاء الاصطناعي، بحيث تُصبح قابلة للتتبع.

ثغرة «الإنكار المعقول»

ويكشف القارح أن النقطة الحرجة في «التزييف العميق» هو أنه لا يهدد الحقيقة فقط، بل يزعزع الثقة في الصور والفيديوهات والتسجيلات الصوتية الحقيقية، وكل ما اعتدنا اعتباره دليلاً قاطعاً على الواقع، حيث أن هذه الثغرة باتت تُستغل بشكل متزايد من قبل جهات وأفراد، عبر ما يُعرف بـ «الإنكار المعقول» أو Plausible Deniability، إذ يمكن لأي شخص الآن أن يشكك في صحة مقطع فيديو حقيقي بمجرد الادعاء بأنه مُفبرك، وبهذا تصبح الأدلة البصرية محل تشكيك دائم، حتى في القضايا الجنائية أو الفضائح السياسية.

ويختم القارح حديثه بالإشارة إلى أن «التزييف العميق» ليس مجرد تطور تقني، بل هو اختبار لقدرة المجتمعات على حماية الحقيقة، إذ يجب على كل مستخدم للإنترنت، أن يكون هو خط الدفاع الأول، عبر التوقف للحظة قبل تصديق أو مشاركة فيديو أو صورة مشكوك بصدقيتها، مؤكداً أن العالم لن يتمكن تماماً من إيقاف الانعكاسات السلبية «للتزييف العميق»، ولكن يمكنه أن يُقلل من تأثيره.