

التزوير العميق بقيادة الذكاء الاصطناعي يندر بمستقبل مشؤوم



ترجمة وتحرير نون بوست

عمل الباحثون على تطوير برنامج جديد قائم على الذكاء الاصطناعي يدعى "ديب فايكس" لتزييف الفيديوهات من خلال تعديل أقوال شخص ما لجعله يقول ما ترغب فيه عن طريق كتابة ما تريد أن يقوله بكل بساطة. وتمثل هذه الخطوة قفزة نوعية أخرى في تقنية لا تعود بأي فائدة اجتماعية فعلية سوى مساعدة إستوديوهات إنتاج الأفلام على توفير الأموال. وتسمح الحكومات الاستبدادية في جميع أنحاء العالم بإساءة استخدام هذه التقنية بهدف تشويه سمعة الأصوات المعارضة وممارسة الأفعال الانتقامية من خلال نشر مقاطع فيديو إباحية وتأجيج الصراع الاجتماعي.

استمرّ العلماء في كل من جامعتي ستانفورد وبرنستون ومعهد ماكس بلانك للمعلوماتية وأدوبي للأبحاث في العمل على تقديم حجج أولية تفسّر السبب وراء انعدام ثقة الأشخاص في تكنولوجيا تطوير منصة برمجيات جديدة تسمح للمرء بتعديل النسخة الأصلية لنص يخصّ خطاباً مسجّلاً ألقاه شخص ما بهدف تغيير الفيديو وتحريف أقوال هذا الشخص.

تعمل هذه التكنولوجيا عن طريق مطابقة الوحدة الصوتية أو ما يعرف بالفونيم، وهي عملية معالجة اللغة الطبيعية للأصوات المختلفة مع دراسة الجانب المرئي وملامح الوجه الناتجة عن هذه الوحدات الصوتية كلما تحدّث الشخص. وفي المرحلة التالية، تُوظف تقنية الذكاء الاصطناعي لفبركة الفيديو الجديد بناءً على النسخة المعدّلة ومطابقة صوت المتحدث مع النص المعدّل.

وفقاً للباحثين: "يضع النهج القائم على تعديل النص الأسس من أجل تعزيز الأدوات المستخدمة في التحرير خلال مرحلة ما بعد إنتاج الأفلام. وغالبًا ما تتطلب مشاهد الحوار المصورة إعادة ضبط التوقيت أو التحرير بناءً على تغييرات نصية صغيرة، ما يستدعي حاليًا عملاً يدوياً شاقاً. وتضمن تقنية التحرير سهولة التكيّف على مستوى المحتوى المرئي والمسموع للفيديو والموجّه نحو فئات مستهدفة محددة. فعلى سبيل المثال، يمكن تعديل مقاطع فيديو الإرشادات مع جماهير من مختلف الخلفيات الاجتماعية أو يمكن تعديل مقطع فيديو لسرد القصص مع أطفال من مختلف الفئات العمرية بالاستناد إلى تعديلات

نصية. باختصار، تهدف الجهود المبذولة إلى تطوير سرد القصص“.

من المهم أن نستمر في تطوير تقنيات علم الأدلة الجنائية وبصمات الأصابع وتقنيات التحقق (الرقمية وغير الرقمية) لكشف الفيديو المزور الذي وقع التلاعب به

في المقابل، يعترف الباحثون بأنه يمكن لبعض الجهات الفاعلة الفاسدة محاولة استخدام هذه التكنولوجيا لأغراض دنيئة، كما يقدمون بعض الحلول المحتملة لهذه المشكلة الحتمية. في هذا الصدد، كتب الباحثون: “تزداد مخاطر ارتكاب الانتهاكات لاسيما عندما تُطبَّق على وسيلة اتصال تُعتبر أحياناً دليلاً موثقاً على الأفكار والنوايا. ونحن نقر بأن الجهات الفاعلة ذات النوايا السيئة قد تستخدم هذه التقنيات لتزوير البيانات الشخصية والتشهير بالشخصيات المرموقة، ونحن نشعر بالقلق إزاء هذا الخداع وسوء الاستخدام“.

في سياق متصل، أضاف هؤلاء الباحثين: “نعتقد أنه من المهم أن يظهر الفيديو المركب باستخدام أدواتنا على أنه محتوى مركب. وفي الحقيقة، قد يكون الفيديو المفبرك واضحاً وفقاً للسياق الذي وضع فيه (على سبيل المثال، في حال كان الجمهور يشاهد فيلماً خيالياً)، ويمكن الإشارة إلى ذلك مباشرة في الفيديو أو عبر العلامة المائية. ونعتقد أيضاً أنه من الضروري الحصول على إذن من الممثلين عند القيام بأي تغيير على المشاهد قبل مشاركة النسخة النهائية من مقطع الفيديو مع قاعدة جماهيرية كبيرة“.

في نهاية المطاف، من المهم أن نستمر في تطوير تقنيات علم الأدلة الجنائية وبصمات الأصابع وتقنيات التحقق (الرقمية وغير الرقمية) لكشف الفيديو المزور الذي وقع التلاعب به. وقد تقلل تدابير الحماية هذه من إمكانية سوء الاستخدام وتسمح لنا باستخدام تقنيات تحرير الفيديو بشكل مبتكر على غرار تقنياتنا. كما ينبغي علينا أخذ مثل هذه التدابير في كل مرة بعين الاعتبار.

أولاً، إن الفكرة القائلة إن تقنية “ديب فايكس” تقدم نفسها بشكل واضح على هذا النحو تكشف عن فهم ساذج للدعاية. ففي الحقيقة، كان ستالين ليجد مثل هذه الطرق جذابة وكانت أي حكومة استبدادية أو حركة سياسية تستخدم هذه التكنولوجيا في المستقبل لاستهداف المنشقين والمعارضين السياسيين والأقليات العرقية أو الدينية.

ثانياً، في حال تمكن الذكاء الاصطناعي من إنشاء مقاطع فيديو مفبركة، فيمكن لشبكة ذكاء اصطناعي إزالة العلامة المائية بطريقة أسهل. أما ثالثاً، فمن المؤكد أن الموافقة على تعديل خطابك كممثل يعتبر أمراً إيجابياً في صناعة الأفلام، لكن تتغير المفاهيم في حال شرع شخص في إنشاء محتويات مفبركة وغير مشروعة. ومن جهتها، لا تسعى الجهات الفاعلة التي تقوم بمثل هذه الأعمال الضارة إلى الحصول على الموافقة على إنشاء محتوى مزيف وإباحي لأحد المشاهير أو التشهير وتشويه الآخرين وفقاً لأجندة معينة.

لطالما اعتبرت صناعة التكنولوجيا منقذاً لوقتنا وأموالنا وخير مثال لروح المبادرة، لكن التصدعات التي لحقت بصورتها العامة بدأت في الانتشار

أخيراً، إن إنشاء أدوات لاكتشاف عمليات التزوير لا يجدي نفعاً، لاسيما في ما يتعلق بالدعاية. وعلى الرغم من أن “بروتوكولات حكماء صهيون” تعد عملية تزوير عمرها أكثر من قرن يعود تاريخها لروسيا القيصرية سيئة السمعة، إلا أنها لا تزال تستخدم بنجاح لتأجيج معاداة السامية في جميع أنحاء العالم حتى يومنا هذا. فكم سيمضي من الوقت قبل أن يبدأ تداول مقاطع فيديو عبر الإنترنت لجورج سوروس يتحدث فيها عن استخدام دماء الأطفال المسيحيين في صناعة الحلويات اليهودية؟ وهل من المفترض تصديق أن عملية إثبات فبركة مقاطع الفيديو هي كل ما نحتاجه لمنع هذه المحتويات من إلحاق ضرر

كبير؟

سوف تتسبب هذه التكنولوجيا في تصفية الشعوب بأكملها

لطالما اعتبرت صناعة التكنولوجيا منقذا لوقتنا وأموالنا وخير مثال لروح المبادرة، لكن التصدعات التي لحقت بصورتها العامة بدأت في الانتشار. ويوشك سائقو أوبر وليفت على القيام بتمرد مفتوح بخصوص أجورهم المتدنية. ويحاول موقع فيسبوك بشكل يائس ومثير للسخرية أن يعيد تقديم نفسه على أنه "شبكة تركز على الخصوصية" حتى في الوقت الذي يواصل فيه نشر البيانات الشخصية لمستخدميه في العالم.

يمكن للباحثين الذين يعملون على عمليات التزوير باستعمال الذكاء الاصطناعي إعلامنا بشكل مباشر أن الحل لإساءة استخدام الدعاية لتقنياتهم يتمثل في العلامة المائية أو الوثوق في تقنية أخرى من أجل إنقاذنا من مخاطر التكنولوجيا التي يقومون بإنشائها.

لقد كشفت الانتخابات الرئاسية الأمريكية لسنة 2016 للعالم أن تكنولوجيا المعلومات منخفضة التكلفة يمكن أن ينجر عنها تأثيرات كبيرة، فضلا عن القيمة الدعائية المخيفة لتقنية ديب فايكس في جميع أنحاء العالم. وقد أثار مقطع فيديو صدر ليلة رأس السنة لرئيس الغابون، علي بونغو أونديمبا، أزمة سياسية بعد أن طعن البعض من منتقديه، بما في ذلك الجيش، في مصداقيته وأعتبروه مفركا ويهدف إلى إخفاء الوضع الراهن للرئيس.

إن تحديد ما إذا كان الفيديو مزيفا أم لا قد لا يكون أمرا سهلا، وبصراحة إن هذا ليس موضوعنا. في الواقع، إن حقيقة وجود هذه التقنية لها تأثيرات سلبية، لأنه يمكن بسهولة استخدام مقاطع الفيديو المزيفة لتشويه سمعة بعض المعارضين أو دعم القادة الفاشلين أو نشر معلومات خاطئة أخرى، حيث يمكن للناس عدم تصديق بعض الفيديوهات التي قد لا تعجبهم والاعتقاد بأنها مزيفة في حين أنها حقيقية. وفي المستقبل، ستكون عملية إثبات ما إذا كان الفيديو مفركا أم لا مفيدة للغاية، وستكون العملية مشابهة لعملية التحقق من الأخبار المضللة.

صنع فيديو مزيف يظهر فيه زعيم أو شخصية دينية أو عرقية أقلية وهو يعترف بارتكابه لبعض الجرائم الفظيعة سيكون دافعا للتحريض على ارتكاب إبادة جماعية

يطالب الباحثون الذين طوّروا هذه التقنية الجديدة بجميع أنواع الضمانات التي ينبغي وضعها لحماية العالم من عواقب هذه التقنية، لكنهم بينما يسارعون إلى تطوير هذه التقنية من أجل إنتاج مقاطع فيديو مزيفة، فإنه لا يبدو أن هناك من يهتم لهذه الضمانات.

في المقابل، يثق مطورو برنامج "ديب فايك" أن معظم سكان العالم يتحكمون في مصادر المعلومات الخاصة بهم، مما يعني أنهم قادرون على اكتشاف ما إذا كان الفيديو الذي شاهدوه للتو على التلفزيون الحكومي مزيفا أم لا، كما يبدو أنه ليس لديهم أي مشكل في القيام بذلك بمفردهم.

لا يمكن الحد من عواقب هذه التقنية. ففي سنة 1994، أسفر سقوط طائرة تقل رئيس رواندا عن مقتل جميع من كانوا على متنها. وعلى خلفية ذلك، ألقت حكومة الهوتو في رواندا اللوم على جيش من متشددى التوتسي مما تسبب في إبادة جماعية استمرت مئة يوم راح ضحيتها مليون شخص من شعب التوتسي. كما تسببت الصراعات المختلفة في البلقان في مقتل أكثر من مئة ألف شخص منذ تفكك يوغوسلافيا. وقد كانت الحكومات تخلق منذ زمن طويل افتراءات وأكاذيب للتحريض على المذابح ضد اليهود والأقليات المضطهدة الأخرى.

يمكن لمثل هذه التوترات والمشاكل التي تحدث الآن في جميع أنحاء العالم أن تتسبب في أي لحظة في اندلاع حرب أو إبادة جماعية، فصنع فيديو مزيف يظهر فيه زعيم أو شخصية دينية أو عرقية أقلية وهو

يعترف بارتكابه لبعض الجرائم الفظيعة سيكون دافعا للتحريض على ارتكاب إبادة جماعية. وحتى يومنا هذا، لا أحد يعرف من هو الشخص المسؤول عن إسقاط طائرة الرئيس الرواندي، لكن الاتهامات هي التي تسببت في حصول هذه المجازر.

لا تشبه جميع القصص الأخرى قصة مجزرة 1994 في رواندا، فأحيانا لا يكون الاتهام كافيا. لكن ما الذي يدفعك لتوجيه الاتهامات في حين أن لديك اعترافا؟ لا شيء قادر على منع تقنية التزوير من أن تكون النسخة الجديدة من بروتوكولات حكماء صهيون للقرن الحادي والعشرين. ومن المؤكد أن هذا النوع من التزوير قد أصبح الآن أمرًا حتميًا وقد يلقي الكثير من الناس حتفهم بسببه لا محالة.

في حال تلقى الأشخاص المشاركون في صنع هذه التكنولوجيا التعليمات المناسبة في مجال الأخلاقيات أو حتى في مجال التاريخ، فإن هذه التكنولوجيا لم تكن لتتجاوز مرحلة الاقتراح

في مقاله الذي نشره موقع "ماثر جونز"، سلط علي بريلا ند الضوء على الجدل الذي أثاره الفيديو المزيف للرئيس الغابوني، حيث كتب: "في الوقت الذي ركزت فيه وسائل الإعلام على السيناريوهات الرهيبة التي يمكن استخدامها ضد الولايات المتحدة ودول غربية أخرى، حذر الخبراء من أن تلك المصادمات يمكن أن تسبب الكثير من الفوضى في البلدان النامية، التي غالبًا ما تكون حكوماتها هشّة وأغلب سكانها لا يعرفون كيفية استخدام التكنولوجيا الرقمية".

في بعض الحالات، قد يتسبب صنع مقطع فيديو مزيف يظهر فيه الرئيس التنفيذي لشركة أمازون جيف بيزوس وهو يعلن عن تقاعده في ارتفاع أسهم الشركة، في حين يمكن أن تتسبب بعض المعلومات الخاطئة في بعض البلدان إلى حدوث انقلابات وحالات صراع وعنف مما قد يهدد استقرار دول بأكملها.

في مقال نُشر مؤخرًا في صحيفة "نيويورك تايمز"، تعتقد أستاذة الفلسفة في جامعة يورك في تورنتو الدكتورة ريجينا ريني أن الأوان قد فات بالفعل للعودة وهي ترى أنه يجب علينا أن نجهّز أنفسنا لمعالجة الفيديو كما لو أننا نتعامل مع شهادة أو أدلة وثائقية. وأضافت ريني أنه "من الواضح أن المعلومات الحالية حول الأخبار المزيفة ليست سوى لمحة عما سيحدث لاحقًا عندما تصبح الأصوات والصور وليس الكلمات فقط عرضة للتلاعب من قبل أي شخص لديه جهاز كمبيوتر جيّد نسبيًا".

الأخلاق

في حال تلقى الأشخاص المشاركون في صنع هذه التكنولوجيا التعليمات المناسبة في مجال الأخلاقيات أو حتى في مجال التاريخ، فإن هذه التكنولوجيا لم تكن لتتجاوز مرحلة الاقتراح، وأي شخص دافع عنها كان ليترد من قبل أولئك المديرين الذين تلقوا تعليمًا أخلاقيًا مناسبًا. ونظرًا لأن جميع المسؤولين في الشركات والمؤسسات التي طورت هذه التكنولوجيا لا يعتقدون أن مثل هذه التقنيّة خطيرة بما يكفي لمنع تطويرها، فلعل الوقت قد حان لكي يتدخل المجتمع ويمنح هؤلاء الأشخاص التعليمات الضرورية.

يبدو أن المجتمع مبني وفق مستوى معين من الثقة، إلا أن هذه التكنولوجيا تدمر ببساطة جزءًا من تلك الثقة المكتسبة بصعوبة بين الطرفين والتي لا يمكن إصلاحها بسهولة

من جهة أخرى، يبدو أن الباحثين المعنيين يقدرّون هذه النقطة، إذ كتبوا التالي: "نعتقد أن إجراء محادثة عامة قوية أمر ضروري لإنشاء مجموعة من اللوائح التنظيمية والقوانين المناسبة التي من شأنها الموازنة بين مخاطر إساءة استخدام هذه الأدوات وأهمية الاستخدامات الإبداعية والتوافقية". ومع ذلك، لم يكن من المفترض أن تصل الأمور إلى هذا الحد، لأن "حالات الاستخدام الإبداعية والتوافقية" تتضاءل أهميتها بسبب الضرر الهائل الذي يمكن أن يحدثه "إساءة استخدام" تكنولوجيتها.

من هذا المنطلق، يستطيع أي شخص عقلائي أن يستنتج أنه لا يمكن لأي فائدة قد تمنحها هذه التكنولوجيا للصناعة تبرير الضرر الذي ستلحقه بالمجتمع. علاوة على ذلك، يبدو أن صناعة التكنولوجيا

خالية من الأشخاص العقلانيين، بما أنهم لا يستطيعون منع أنفسهم من التدخل ما إن تعترضهم إحدى الفئات الاجتماعية التي يمكنهم "تعطيلها".

تكشف هذه التكنولوجيا أنه لم يعد بإمكاننا الوثوق بما نراه بأعيننا. وقد يحرص المرء على أهمية أن نكون يقظين دائمًا في مواجهة المعلومات الخاطئة، ولكن هذا الأمر غير مجده كذلك. فمن يريد أن يقضي كل لحظة من يومه في قراءة المعلومات أو عرضها أو تحليلها في ظلّ هذا النوع من الشكوك الدائم؟ علما بأن هذا الأمر يتطلب مستوى تعليمي معين وطاقة عقلية لا يمتلكها الجميع ولا يستطيع أي أحد الحفاظ عليها إلى الأبد.

في نهاية المطاف، يبدو أن المجتمع مبني وفق مستوى معين من الثقة، إلا أن هذه التكنولوجيا تدمر ببساطة جزءا من تلك الثقة المكتسبة بصعوبة بين الطرفين والتي لا يمكن إصلاحها بسهولة، في حال وجدت في المقام الأول. ولكن، في حال كنت تعمل على تعديل بسيط لبرنامج نصي وكنت بحاجة إلى إعادة تصوير 15 ثانية من الحوار، فإن هذه التكنولوجيا جديرة بالتكلفة الاجتماعية التي ندفعها من أجلها.

المصدر: إنترستنج إنجينيرينغ