البرنامج الوطني للذكاء الاصطناعي
NATIONAL PROGRAM FOR ARTIFICIAL INTELLIGENCE

United Arab Emirates

دليل التزييف العميق

يوليو 2021

لقراءة هذا المستند رقميًا

DEEP FAKE GUIDE

المحتوى

www.ai.gov.ae

مقدمة

عملت تقنيات الذكاء الاصطناعي وتعلم الآلة على دعم عدد كبير من التطبيقات في حياة الإنسان، ونحن نرى آثارها في حياتنا اليومية. ولكن التطورات في مجال الذكاء الاصطناعي والرؤية الحوسبية برزت بشكل أوضح في السنوات الأخيرة من خلال التركيز على استخدام خوارزميات الذكاء الاصطناعي لإنشاء مقاطع الصوت والفيديو وتعديلها. وقد أوجدت خوارزميات الذكاء الاصطناعي الحديثة مشكلة لم نواجه مثلها من قبل، وهي قدرتها على إنشاء مقاطع صوت وفيديو حقيقية لم يسجلها أحد (ولم تحدث في الواقع). على سبيل المثال، يمكن لخبراء الذكاء الاصطناعي إنشاء مقاطع صوت وفيديو تظهر فيها شخصيات معروفة ومشاهير أو سياسيون معروفون وهم يؤدون حركات مضحكة أو تصرفات غير لائقة.لا تقتصر إستخدامات هذه التقنية على ذلك، فلها عدة استخدامات أخرى تشمل المؤثرات الخاصة في مجال صناعة الأفلام مثلا.

استُخدِمت هذه التقنيات الجديدة أيضاً في المجال الطبي لإنشاء صوت اصطناعي للمرضى الذين يعانون من تلف في الحبال الصوتية ويؤثر على قدرتهم على الكلام؛ حيث تتخيل خوارزمية الذكاء الاصطناعي صوت المريض وتنشئ ملفات صوتية يبدو لسامعها كما لو أن المريض يتحدث بشكل طبيعي.

المصطلح الأكثر شيوعاً للإشارة إلى أنظمة الذكاء الاصطناعي التي تنشئ مقاطع صوت وفيديو غير حقيقية هو التزييف العميق "Deepfake". يشير هذا المصطلح إلى أن المحتوى مُزيَّف وتم إنشاؤه باستخدام التعلم العميق[1] ولا أساس له من الصحّة رغم أنه يبدو واقعياً ومقنعاً جداً. ولم يسبق أبداً

1 التعلم العميق هو أسلوب تعلم آلي أثبت أنه مفيد للغاية في التطبيقات الصوتية / المرئية بما في ذلك الكشف عن الأشياء والتعرف عليها وتطبيقات تتبع الفيديو.

في تاريخ الذكاء الاصطناعي أن كانت الخوارزميات قادرة على إنشاء محتوى، حيث كانت مجرد أدوات بسيطة يقتصر دورها على تحليل البيانات وتعلمها. لكن التكنولوجيا شهدت تقدماً كبيراً في عام 2014[2]، حيث تم تطوير تقنية معقدة قائمة على التعلم العميق أُطلق عليها اسم الشبكات التوليدية التنافسية "Generative Adversarial Networks" تتعلم من مقاطع الصوت والفيديو ثم تقوم بإنشاء مقاطع مماثلة. وقد أتاح هذا التقدم لأنظمة الذكاء الاصطناعي بإنتاج بيانات جديدة (بيانات غير حقيقية)، وفتحت الشبكات التوليدية التنافسية الباب أمام تطبيقات كثيرة، من بينها ما يُعرف اليوم باسم التزييف العميق "Deepfakes".

عندما ظهرت تقنية التزييف العميق "Deepfakes" للمرة الأولى، كانت عملية معقدة وصعبة التنفيذ. ولكن مع تقدم هذه التقنية، أصبح من السهل على غير المختصين، إنشاء مقاطع صوت وفيديو مُعدَّلة، وأصبحت التقنية أسهل وأرخص وباتت تحتاج إلى قدر أقل من البيانات والقدرة الحوسبية، كما أصبح من الممكن إنشاؤها بشكل أسرع بكثير. وقد ساهم ذلك في زيادة انتشار مقاطع الصوت والفيديو المُعدَّلة للأغراض الترفيهية من ناحية، ولكنه ساعد من ناحية أخرى على استغلال هذه التقنيات لإنشاء مقاطع صوت أو فيديوهات أو أخبار مسيئة و مزيفة ومحتوى يُطلَق عليه اسم التزييف العميق "Deepfakes"، غالباً ما يُستخدم في التنمّر الإلكتروني.

نظراً لصعوبة الكشف عن هذه التقنية وتحديدها، فإن هذا الدليل يُعرِّف بوضوح مشكلة التزييف العميق "Deepfakes" ويبين تدابير الحماية منها، وكيفية إبلاغ السلطات المختصة بحالات التزييف العميق "Deepfakes".

2 استنادًا إلى إم آي تي تكنولوجي ريفيو و بي بي سي متاح على:
https://www.technologyreview.com/2014/12/29/169759/2014-in-computing-breakthroughs-in-artificial-intelligence/
and https://www.bbc.co.uk/teach/ai-15-keymoments-in-the-story-of-artificial-intelligence/zh77cqt

تعريف المشكلة

التزييف العميق "Deepfakes" هو شكل جديد لمشكلة قديمة تتعلق بتوزيع محتوى مُزيَّف. سابقا، كان المحتوى المُزيَّف من إنتاج أشخاص يستخدمون أدوات للتلاعب بالصور والصوت، ولكنه اليوم يُنتج باستخدام تقنية الذكاء الاصطناعي وتعلم الآلة، مما يجعله أقرب بكثير إلى الواقع. ويسمح استخدام "تقنية التزييف العميق" بإنشاء محتوى (فيديو وصوت) يتم من خلاله انتحال شخصيات أخرى وتقديم معلومات مُزيَّفة عن سلوكهم و أنشطتهم و البيئة المحيطة بهم.

ويمكن أن تمثل تقنية التزييف العميق "Deepfakes" تهديداً حقيقياً عندما تُستخدم كأداة لإنشاء وتوزيع مقاطع صوت ، فيديوهات و معلومات زائفة عن أفراد ومسؤولين وشخصيات معروفة تقول وتفعل أشياء لم تحدث أبداً، كما يمكن لهذه الصور و الفيديوهات أن توضع خارج موقعها لتبدو كمواقف وأحداث حقيقية.

يجري عادة استغلال تقنية التزييف العميق "Deepfakes" لأغراض مسيئة، منها على سبيل المثال.

- الإضرار بسمعة الأفراد والدول.
- التلاعب بالرأي العام بقصد التأثير على حدث سياسي أو إعاقة عمل الحكومة.
- زعزعة الثقة باستخدام واقع مُزيَّف يمكن حدوثه.
- خلق أدلة ملفقة للتأثير على أحكام القضاء.

انتشر في السنوات الأخيرة عددٌ من مقاطع الفيديو المُزيَّفة تعرف من خلالها ملايين الأشخاص حول العالم على هذه التقنية. وكان أثر ذلك كبيراً جداً لأن الأشخاص الذين تعرضوا لهذا التنمّر كانوا مشاهير وسياسيين وشخصيات عامة أخرى. فمن السهل جد إنشاء صور و فيديوهات مُزيَّفة للمشاهير لأن جميع البيانات المطلوبة متاحة في مختلف المصادر الإعلامية.

أدى التطور السريع في تقنية التزييف العميق "Deepfakes" إلى انتشارها بشكل أكبر وتسهيل الوصول إليها واستخدامها وجعلها أرخص وأسرع. حيث باتت تتوفر أدوات أحدث وأقوى تحتاج إلى قدر أقل من البيانات والقوة الحوسبية، مما أدى إلى زيادة دقة عمليات التزييف العميق وعدد أدواتها. ويؤثر التزييف العميق "Deepfakes" على الأفراد ويضر باستقرار المؤسسات ومصالح الدول، ويشكّل تهديداً للسمعة من خلال تزوير السلوك والأنشطة والأحداث، حيث يُظهر تورط شخص أو عدة أشخاص في فعل لم يحدث أبداً. قد تستخدم هذه التقنيات سلبا ضد دول معينة خلال وضع أجندة عامة للتأثير على المجتمع والتلاعب بالرأي العام، خاصة قبل حدث معين. وقد يمتد هذا الأثر فيشكّل تهديداً لسمعة الدول ويتسبب في تشويش العلاقات الدولية والدبلوماسية إذا لم تقم الحكومات المعنية بالاستجابة للواقعة على الفور.

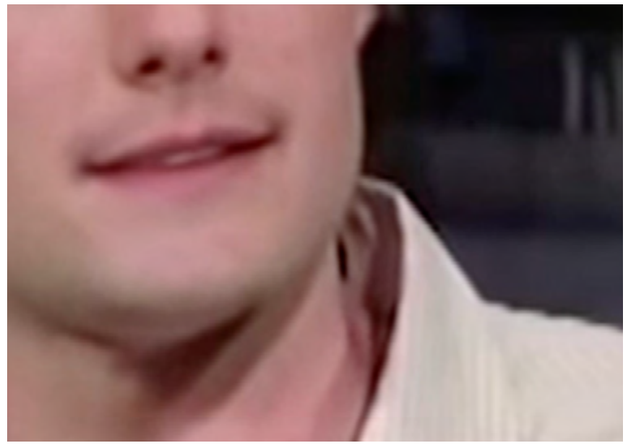
تحظر القوانين المحلية الحالية التنمّر الإلكتروني والإيذاء المتعمد وانتحال الشخصية. كما توجد معايير محددة تتعلق بإنتاج ، توزيع ، نشر أو بث محتوى إعلامي يحتوي على أخبار مُزيَّفة ويتعدى على خصوصية الأفراد ويخالف الأعراف والتقاليد الإسلامية.

إلا أن بعض الدول تفتقر إلى هذه التشريعات وتواجه انتشاراً كبيراً في ممارسات "التزييف العميق" Deepfakes التي تسيء للأنظمة السياسية وتضر بخصوصية الأفراد وحياتهم الشخصية. لذلك، بدأت الدول تصدر تشريعات تحظر التزييف العميق "Deepfakes"، ولكن هذه التشريعات قد لا تُنفذ على أرض الواقع وقد لا تكون كافية لردع مثل هذه الممارسات.

استبدال الوجه: تعمل تطبيقات استبدال الوجه Face Swapping على معالجة بيانات وجه المستخدم، ثم تستبدل وجوه المشاهير بوجه المستخدم عن طريقة الاستعانة بخوارزميات الذكاء الاصطناعي

هذه صورة شخص غير حقيقي بإستخدام تقنية الشبكات التوليدية
Dec 2019 - Karras et al. and Nvidia
(https://thispersondoesnotexist.com)

ما هو التزييف العميق “DEEPFAKES”؟



أصبح من السهل جداً إنتاج المحتوى باستخدام تقنية التزييف العميق "Deepfake"، وكما نلاحظ هنا بات من الصعب الآن أن نحدد ما إذا كان الشخص الموجود في الفيديو هو الشخصية الحقيقية أم شخصاً مزيفاً.

من الناحية القانونية، يعرّف التزييف العميق "Deepfakes" على أنه مقطع فيديو «تم إنشاؤه بقصد الخداع، ويبدو أنه يصور شخصاً حقيقياً يقوم بفعل لم يحدث في الواقع»[1]. كما يمكن تعريف التزييف العميق "Deepfakes" على أنه محتوى مرئي أو صوتي أو كلاهما تم التلاعب به باستخدام الذكاء الاصطناعي وتقنية برمجيات متقدمة لتزييف حقيقة الأفراد والأشياء والأماكن والأحداث. ويبدو هذا المحتوى المُزيَّف قريباً من الواقع، وقد يجد عامة الناس صعوبة في اكتشافه.

التلاعب بالصور والصوت ومقاطع الفيديو ليس بالأمر الجديد، وهو معروف منذ سنوات. ولكن مع تطور التقنيات الحديثة في مجال الذكاء الاصطناعي وتعلم الآلة وتسويق البرامج، أصبح من السهل إنشاء محتوى مُزيَّف باستخدام برامج وتطبيقات منتشرة بكثرة على أجهزة الحاسوب والأجهزة المحمولة.

يمكن تصنيف المحتوى المُزيَّف إلى فئتين رئيسيتين على النحو التالي:

1 **التزييف السطحي Shallowfakes**:

أ- مقاطع فيديو ذات حركة بطيئة: وهي مقاطع فيديو استُخدم فيها برنامج لتعديل الفيديو لإبطاء سرعة الكلام دون تغيير طبقة الصوت. وقد يكون القصد من ذلك هو الإشارة إلى وجود خلل في الشخص المستهدف من خلال الفيديو أو التشديد على كلمات معينة أو نبرة الصوت لتزييف وجهات نظر محددة ولترك انطباعاً خاطئاً لدى الجمهور.

ب- تغيير التواريخ والمواقع: التلاعب بالتواريخ والمواقع لتظهر مقاطع الفيديو على أنها حديثة وفي أماكن مختلفة، مما يؤدي إلى انتشار أخبار كاذبة تضر بسلامة المجتمع والأفراد.

2 **التزييف العميق Deepfakes**: وهو عملية يجري فيها استبدال الوجه "Face Swapping" باستخدام تقنيات الذكاء الاصطناعي وتعلم الآلة من خلال تدريب خوارزميات الذكاء الاصطناعي على الصور المستخرجة من شبكات منفصلة، ثم إعادة بناء الوجه الجديد وإنشاء الفيديو المطلوب. كما يمكن تنفيذ العملية نفسها لإنشاء مقاطع صوتية.

1 مشروع قانون التزييف العميق في ولاية إلينوي في الولايات المتحدة الأمريكية: ILCS 5/29-21 10

يتطلب التزييف العميق عالي الجودة تدريبًا متكررًا ومزيدًا من البيانات

مع تطور التقنيات الحديثة في مجال الذكاء الاصطناعي وتعلم الآلة وتسويق البرامج، أصبح من السهل إنشاء محتوى مُزيَّف أو الوصول إليه باستخدام برامج وتطبيقات منتشرة بكثرة على أجهزة الحاسوب والأجهزة المحمولة بطرق متعددة.

تطبيقات التلاعب بالوجه: تطبيق Faceswap (لنظامي iOS وAndroid) هو تطبيق لتحرير الصور مدعوم بتقنية تعلم الآلة. لا أحد من هؤلاء الأشخاص حقيقي، فقد تم اختلاق كل هذه الصور باستخدام تقنيات الذكاء الاصطناعي والشبكات التوليدية التنافسية.

أنواع المحتوى المُزيَّف

ترتبط الاستخدامات الأكثر شيوعاً لتقنية «التزييف العميق» Deepfakes بما يلي:

1 **المحتـوى المـرئي:** ويقصـد بذلـك اسـتخدام تقنيـة التزييـف العميـق "Deepfakes" في إنشـاء الصـور ومقاطـع الفيديـو.

a. تبديـل الوجـه باسـتخدام خوارزميـات التشـفير وفـك التشـفير Encoder/Decoder Algorithms لتركيب الخريطة الرقمية Digital Map لوجه شخص معين على وجه شخص آخر.

تستخدم خوارزمية التشفير آلاف الصور لدراسة ملامح الوجه لدى شخصين مختلفين، ثم تكتشف أوجه التشابه بينهما وتختصرها إلى ميزات مشتركة وتضغط الصور. بعد ذلك يتم تدريب خوارزمية ذكاء اصطناعي ثانية تسمى بخوارزمية فك التشفير على كيفية استعادة الوجوه من الصور المضغوطة. وبما أن الوجهين مختلفان، تتم برمجة الخوارزمية الأولى لاستعادة وجه الشخص الثاني. ولتبديل الوجهين، يتم تزويد تعليمات فك التشفير "Decoder Algorithm" ببيانات الصور المشفرة الخاصة بالوجه الآخر.

b. التلاعب بالوجه مثل تعديل تعابيره ومزامنة الشفاه باستخدام الشبكات التوليدية التنافسية.

تستخدم هذه الطريقة خوارزميتين للذكاء الاصطناعي، حيث يتم إدخال بيانات عشوائية في الخوارزمية الأولى تعرف باسم خوارزمية التوليد لتحويلها إلى صورة. ثم تُضاف هذه الصورة المُصطنعة ضمن سلسلة من الصور الحقيقية لبعض المشاهير على سبيل المثال، ويتم إدخالها في الخوارزمية الثانية المعروفة باسم خوارزمية التمييز "Discriminator". في البداية، لا تبدو الصور التي يتم إنتاجها على أنها صور وجوه، إلا أن تكرار العملية عدة مرات وإجراء التعديلات بناءً على الملاحظات على الأداء يؤدي إلى تحسين أداء خوارزميتي التمييز "Discriminator" وخلق الصور الجديدة "Generator" . وبعد تنفيذ عدد كافٍ من الدورات والملاحظات، تبدأ الخوارزمية في إنتاج وجوه واقعية تماماً لأشخاص غير حقيقيين.

2 **المحتـوى الصـوتي:** ويُقصد به بشكل رئيسي تركيب الصوت وتعديله إما عن طريق إنشاء ملف صوتي يتضمن حديثاً مزيفاً بنفس صوت الشخص لكنه لم يقله في الحقيقة، أو عن طريق التحكم بنبرة صوت الشخص لإظهار شعور أو سلوك غير حقيقي.

تشكل تقنية التزييف العميق "Deepfakes" مخاطر كبيرة بسبب التلاعب بالحقائق وتشويه السمعة من خلال بث هذه الرسائل على قنوات إعلامية مختلفة دون الكشف عن مصدرها.

ورغم أن هذه التقنيات معقدة بالنسبة لغير المختصين في الذكاء الاصطناعي، فإن شبكة الإنترنت توفر العديد من الأدوات والتطبيقات التي تسمح لأي كان بإنشاء محتوى تزييف عميق بشكل فوري على هواتفهم و حواسبيهم.

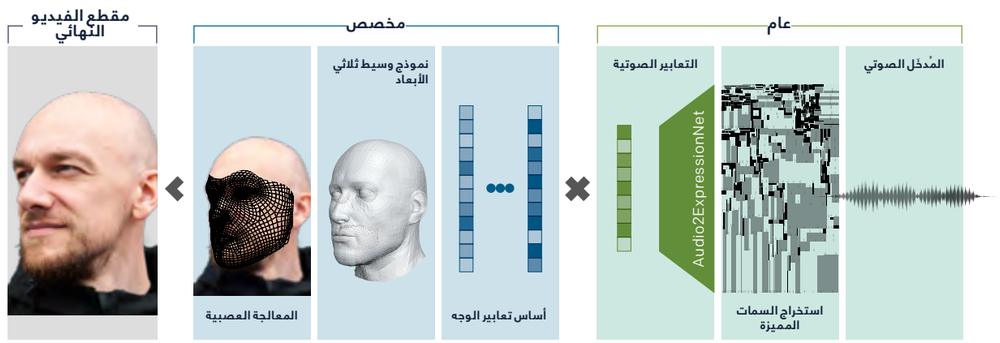
الاستخدامات الإيجابية للمُحتوى المُزيَّف

الاعتقاد الشائع هو أن التزييف العميق أداة غير آمنة ولكن من المهم الإشارة إلى أنها مجرد أداة لديها استخدامات جيدة وغير جيدة. هناك العديد من الاستخدامات الإيجابية لتطبيقات التزييف العميق في مختلف المجالات، بعض الأمثلة تشمل:

1 **التطبيقات الطبية:**
- إنشاء صور طبية جديدة مثل صور الرنين المغناطيسي لأغراض التدريب.
- إنشاء ملفات صوتية قائمة على حركة أعضاء النطق للمرضى الذين فقدوا قدرتهم على الكلام بسبب أورام السرطان والأمراض الأخرى التي تؤثر على الحبال الصوتية.

2 **الترفيه:**
- صناعة الأفلام والإعلانات: تحسين المحتوى وإنشاء مؤثرات بصرية خاصة لصنع مشاهد متحركة والتلاعب بالوجه.
- مذيع الأخبار من خلال مقدم الأخبار الافتراضي.

3 **خدمة العملاء:**
- المساعد الافتراضي: هناك اتجاه إلى استخدام المساعد الافتراضي بالصوت والصورة لتقديم خدمة العملاء في مراكز الاتصال.



يتم استخدام خوارزميات صناعة الأصوات والتلاعب في تعابير الوجه لإنشاء مقاطع فيديو تبدو حقيقية للشخص المستهدف.

كيف أحمي نفسي وأسرتي؟

تعتمد تقنية التزييف العميق "Deepfakes" على جمع كمية كبيرة من البيانات لتدريب نظام الذكاء الاصطناعي على صنع مقاطع صوت وفيديو مُزيَّفة. ويمكن أن تكون هذه البيانات على شكل:

صور للأشخاص

مقاطع فيديو

مقاطع صوتية

وغالباً ما يقوم الأشخاص بنشر هذه البيانات على نطاق واسع في وسائل التواصل الاجتماعي. بشكل عام، كلما ازداد حجم البيانات التي يحصل عليها صانع مقاطع الصوت أو الفيديو المُزيَّفة، كلما تحسنت جودة هذه المقاطع وأصبحت أقرب إلى الواقع. وكما هو الحال في أي تقنية تعتمد على الخوارزميات، تحتاج تقنية التزييف العميق "Deepfakes" باستخدام الذكاء الاصطناعي إلى التدريب لكي يتحسن أداؤها. لذلك يجب أن يدرك الناس أن زيادة استخدامهم لهذه التطبيقات المختلفة وزيادة الصور التي ينشرونها على الإنترنت تعني أنهم يسمحون بتدريب النظام بشكل أكبر على تقديم مقاطع صوت وصور مُزيَّفة قريبة جداً من الواقع.

كلما زاد حجم البيانات التي يحصل عليها صانع مقاطع الصوت أو الفيديو المُزيَّفة، كلما تحسنت جودة هذه المقاطع وأصبحت أقرب إلى الواقع.

من المهم تعريف الأطفال الصغار بمخاطر كشف هويتهم (بالصوت والصورة) على الإنترنت حيث يمكن أن يقعوا ضحايا لتقنية التزييف العميق "Deepfakes".

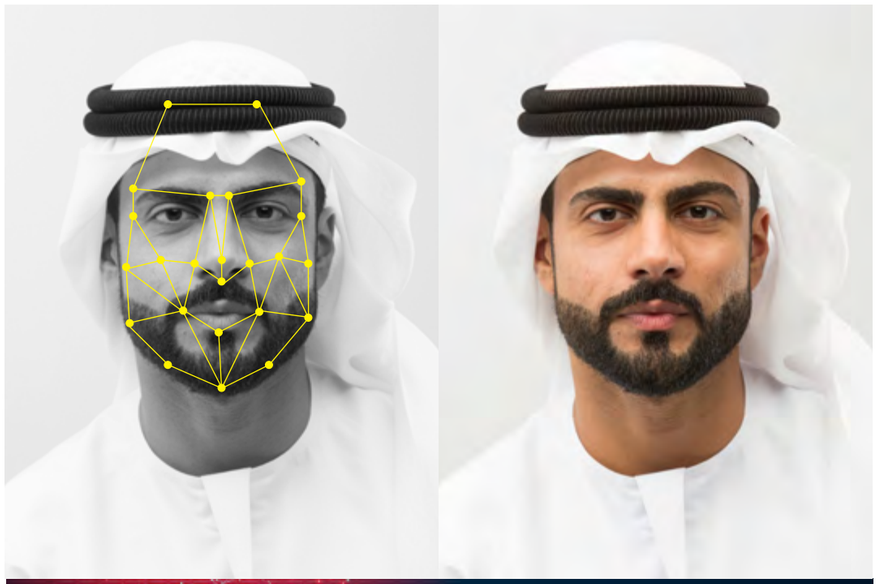
كيف يمكنني اكتشاف "التزييف العميق" DEEPFAKES؟

من الممكن اكتشاف التزييف العميق "Deepfakes"، إلا أن هناك بعض أنظمة الذكاء الاصطناعي المتطورة التي يمكن استخدامها لجعل المحتوى المُزيَّف أقرب إلى الواقع، مما يزيد من صعوبة اكتشافه. ويمكن اكتشاف معظم مقاطع الفيديو المُزيَّفة من خلال ما يلي:

1. حركات وجه الشخص الفوضوية وغير المنتظمة.
2. حدوث تغير مفاجئ في الإضاءة الموجهة إلى الشخص.
3. تغير لون البشرة أثناء المقطع.
4. رمش العين بشكل متكرر أو عدم رمشها على الإطلاق.
5. عدم تطابق حركة الشفاه مع الكلام المسموع.
6. تشوه في المنطقة المحيطة بالوجه.

وعلى الرغم من إمكانية الفحص الدقيق لملامح الأشخاص الذين يظهرون في مقاطع الفيديو، إلا أن هذا قد يستغرق وقتاً طويلاً ونتيجته غير موثوقة. وتعتمد الطريقة الأفضل لاكتشاف المحتوى المُزيَّف على تنفيذ فحص منهجي للكشف عن التزييف العميق "Deepfakes" باستخدام أدوات قائمة على الذكاء الاصطناعي يجب تحديثها بشكل منتظم.

يتم إجراء العديد من الأبحاث حول استخدام الذكاء الاصطناعي لاكتشاف التزييف العميق "Deepfakes".



مجلس جودة الحياة الرقمية

تأسس مجلس جودة الحياة الرقمية بناء على توجيهات سمو الشيخ سيف بن زايد آل نهيان، نائب رئيس مجلس الوزراء ووزير الداخلية في دولة الإمارات العربية المتحدة، ويتكون من عضوية عدد من الجهات الاتحادية والمحلية. ويعد البرنامج الوطني للذكاء الاصطناعي أحد الأعضاء الرئيسيين وله دور فعال في تمكين جودة الحياة الرقمية من خلال مجال الذكاء الاصطناعي والاقتصاد الرقمي وأيضاً مجالات العمل عن بعد. يهدف المجلس إلى خلق مجتمع رقمي آمن في دولة الإمارات، وتعزيز هوية إيجابية ذات تفاعل رقمي هادف.

للتعرف أكثر على دور مجلس جودة الحياة الرقمية يرجى زيارة الموقع الالكتروني:

www.digitalwellbeing.ae

جودة الحياة الرقمية
DIGITAL WELLBEING

منشورات أخرى

لقراءة المنشورات الأخرى الصادرة عن البرنامج الوطني للذكاء الاصطناعي، يرجى مسح الباركود .

دليل الذكاء الاصطناعي	دليل التعاملات الرقمية	إستراتيجية الإمارات للذكاء الاصطناعي 2031	تقرير البنية التحتية لأجهزة الذكاء الاصطناعي في دولة الإمارات العربية المتحدة
			

www.ai.gov.ae

دليل التزييف العميق