SIPAQ
رابطة محترفي تكنولوجيا المعلومات والاتصالات السودانيين بدولة قطر
SUDANESE ICT PROFESSIONALS ASSOCIATION IN QATAR

هل ما نراه على الإنترنت حقيقي فعلًا؟ التزييف العميق يغيّر قواعد الثقة...

(قراءة متعمقة في التزييف العميق والتحديات التي يواجهها الفضاء الرقمي)

عبـــدالرحمن بابكر

خبير أنظمة أمنية

ahussein19926@gmail.com

بتاريخ : 01 فبراير 2026

1. مقدمة

في إبريل من العام 2023، وفي ولاية أريزونا الأمريكية، تلقت سيدة تدعى جينيفر دي ستيفانو اتصالا هاتفيا من رقم مجهول، سمعت على الطرف الآخر صوتا مطابقا تماما لصوت ابنتها بريانا ذات الخمسة عشرة ربيعا وهي تبكي وتستغيث، بعد ثوان، انتقلت المكالمة إلى رجل آخر هدد الأسرة وطالب بفدية لقاء إطلاق سراح الابنة، سرعان ما كشف الوالدان الخدعة بعد التواصل مباشرة مع ابنتهما التي كانت بأمان، ليتبين أن ما حدث لم يكن سوى اختطاف وهمي تم عن طريق استنساخ الصوت عبر تقنيات الذكاء الاصطناعي، وجمعت هذه التقنية بيانات صوت الابنة من مصادر مفتوحة (على الأرجح من تسجيلات منشورة على الإنترنت أو مكالمات سابقة)، ثم أنتجت مقطعا صوتيا مزيفا يحاكي صوتها بدقة. فتحت الحادثة الباب واسعا أمام التوعية بمخاطر الجريمة السيبرانية المعتمدة على الذكاء الاصطناعي والتي بدأت تأخذ أنماطا جديدة وغير معروفة، وعلى رأسها التزييف العميق DeepFake، الذي أصبح يمثل تهديدا متناميا للأفراد والمؤسسات والمجتمعات في هذا العصر الرقمي.

التزييف العميق هو مصطلح يشير إلى تقنيات تقوم بتوليد محتوى صوتي أو مرئي مزيف عالي الدقة يحاكي الواقع بشكل يصعب تمييزه، وتستفيد هذه التقنيات من خوارزميات الذكاء الاصطناعي التوليدي التي باتت قادرة على إنشاء مقاطع فيديو وصوت وصور تبدو حقيقية تماما، ومع تزايد سهولة استخدام هذه الأدوات وانتشارها، أصبحت مخاطر إساءة استخدامها كبيرة، بدءا من عمليات الاحتيال والابتزاز وانتهاء بنشر المعلومات المضللة والتشهير.

في هذا التقرير، نستعرض مفهوم الذكاء الاصطناعي التوليدي وأبرز تقنياته، وكيف يسهم في ظهور التزييف العميق، ثم نناقش الأنماط والمؤشرات التي تكشف المحتوى المزيف، وطرق رصده تقنيا، كما نسلط الضوء على إحصائيات عالمية حديثة حول التزييف العميق تبين مدى انتشار هذه الظاهرة خلال الأعوام 2023–2024، قبل أن نختم ببيان الجهود الدولية المبذولة لمكافحتها.

2. وصف خوارزميات الذكاء الاصطناعي التوليدي

يعرف **"الذكاء الاصطناعي التوليدي"** بأنه أحد فروع الذكاء الاصطناعي الذي يركز على إنشاء محتوى جديد كليا، ويستطيع الذكاء الاصطناعي التوليدي إنتاج نصوص، صور، أصوات، فيديوهات، أو حتى نماذج ثلاثية الأبعاد **3D Models** اعتمادا على أنماط **patterns** تعلمها أثناء عملية التدريب باستخدام بيانات ضخمة.

توجد عدة تقنيات متوفرة حاليا ضمن إطار الذكاء الاصطناعي التوليدي لإنتاج أنماط مختلفة من المخرجات وفقا للنموذج يعد أبرزها أدناه:

1. **الشبكات التوليدية التنافسية Generative Adversarial Networks**

تعتمد هذه التقنية على وجود نموذجين عصبيين يتنافسان معا – أحدهما مولد (Generator) والآخر مميز (Discriminator)، يقوم المولد بإنتاج صور أو أصوات أو نصوص، بينما يحاول المميز كشف المواد المزيفة وتمييزها من الحقيقية، يستمران في عملية تدريب تنافسية حيث يتحسن أداء كل منهما تدريجيا؛ فالمولد يتقن إنتاج محتوى أكثر واقعية، والمميز يطور قدرته على اكتشاف

التزييف، تستمر هذه العملية حتى يصل المحتوى المنتج إلى درجة من الواقعية يصعب معها تمييزه عن الحقيقة.

2. **النماذج اللغوية الضخمة Large Language Models**

تعد النماذج اللغوية الضخمة LLMs أحد أبرز فروع الذكاء الاصطناعي التوليدي المتخصصة في توليد النصوص وفهم اللغة الطبيعية، حيث تعتمد أحدث هذه النماذج على بنية المحولات (Transformers) التي تستخدم آلية الانتباه الذاتي (Self-Attention) لتحليل سياق الكلمات بشكل متوازي، وأتاحت هذه البنية تجاوز قيود النماذج القديمة كسلاسل الشبكات العصبية المتكررة (RNN) وذاكرة المدى القصير الطويل (LSTM)، وكانت النتيجة هي نماذج لغوية قادرة على إنتاج نصوص مطولة ومنطقية تشبه ما يكتبه الإنسان.

3. **نماذج الانتشار Diffusion Models**

نماذج ذكاء اصطناعي توليدي لانتاج عدة أنماط من المخرجات كالصور والنصوص والوسائط الصوتية وغيرها، وتقوم النماذج المذكورة بإضافة الضوضاء noise إل البيانات وتعلم إزالتها أثناء عملية التدريب، ويعد تدريب النماذج المذكورة أكثر استقرارا مقارنة بنماذج الشبكات التنافسية التوليدية GAN، كما يمكن لهذه النماذج إنتاج مخرجات بجودة فائقة.

4. **حقل الاشعاع العصبي Neural Radiance Field**

تعد حقول الاشعاع العصبي NeRF نمطا حديثا من نماذج التوليد القادرة على إنشاء صور وأصوات ونصوص بواقعية عالية، وتقوم فكرة هذه النماذج على إضافة ضوضاء عشوائية random noise تدريجيا إلى البيانات التدريبية ثم تعلم إزالة هذه الضوضاء لإعادة تكوين البيانات الأصلية.

خلال التدريب، يتعلم النموذج كيفية الانتقال من حالة الضجيج إلى الحالة الواضحة، مما يمكنه لاحقا من توليد محتوى جديد انطلاقا من الضوضاء فقط، وتمتاز نماذج الانتشار باستقرار أعلى في التدريب مقارنة بالشبكات التوليدية التنافسية، وقادرة على إنتاج مخرجات فائقة الجودة يصعب تمييزها، (من أمثلتها نماذج توليد الصور الحديثة مثل Stable Diffusion وغيرها).

5. **النماذج الهجينة Hybrid Models**

هي نماذج توليدية تجمع بين أكثر من تقنية ذكاء اصطناعي توليدي بهدف تحقيق نتائج أكثر واقعية أو لتجاوز حدود كل تقنية على حدة بهدف الاستفادة من إيجابيات كل تقنية على حدة، على سبيل المثال، يمكن دمج نموذج انتشار مع نموذج إشعاع عصبي (Diffusion + NeRF) لإنتاج فيديوهات 3D مزيفة لتبدو مقنعة بشكل غير مسبوق، أو دمج نموذج لغوي ضخم مع نموذج رؤية حاسوبية لإنشاء محتوى متعدد الوسائط يتفاعل فيه النص مع الصورة أو الفيديو بشكل ذكي.

3. أنماط الفيديوهات المنتجة عبر الذكاء الاصطناعي

على الرغم من الواقعية المتزايدة للمحتوى المنتج عبر نماذج الذكاء الاصطناعي التوليدي، إلا أنه لا يزال بالإمكان كشف العديد من المقاطع المزيفة عبر ملاحظة بعض المؤشرات والعلامات التحذيرية في الصور والفيديوهات والصوتيات، حيث يعتمد كشف التزييف عادة على مزيج من الفحص البصري أو السمعي الدقيق وتحليل السياق والسلوكيات المرافقة للمحتوى، فيما يلي أبرز المؤشرات التي قد تدل على أن المحتوى مفبرك وغير أصلي:

1- مؤشرات بصرية

- **تشوه في ملامح الوجه:** ظهور اختلالات أو تشوهات حول الفم أو العينين خصوصا أثناء الكلام، مثل تكسر في حواف الوجه أو اختفاء بعض المعالم للحظات من الصورة.
- **رمش وحركة عين غير طبيعيين:** قد تومض عيون الشخص بمعدل أقل أو أكثر من الطبيعي، أو تبدو حركة العين غير متزامنة بشكل صحيح مع حركة الرأس وتعبيرات الوجه.
- **إضاءة وظلال غير متسقة:** عدم تناسب ظلال الوجه مع مصدر الضوء في المشهد، أو وجود بريق ولمعان غير منطقي على البشرة لا يتوافق مع بيئة التصوير الحقيقية.
- **عيوب في الانتقال والحركة:** ملاحظة اهتزازات أو تقطيع في الصورة عند حركة الشخص أو التغير المفاجيء في الخلفية، مثل اندماج الجسم بالخلفية أو اختفاء أطراف عند الانتقال

السريع، كذلك قد تظهر الخلفية ضبابية أو تتغير التفاصيل فيها بشكل مفاجئ لا تتسق مع حركة الكاميرا.

- **تفاصيل غير منطقية في الأطراف:** غالبا ما تكشف التفاصيل الدقيقة كأصابع اليدين أو الأذنين عن التزييف، إذ قد تكون غير متناسقة في الشكل أو العدد (كمثال شائع: مواجهة صعوبة في توليد شكل أصابع اليد الطبيعية)، التحسن الأخير في جودة النماذج قلل هذه العيوب، لكنها تظل من العلامات الممكنة لكشف المحتوى المزيف.

2- **مؤشرات صوتية**

- **نبرة ثابتة وغياب الانفعالات الطبيعية:** قد يلاحظ أن الصوت المزيف ذو نبرة رتيبة أو متكررة الإيقاع، ويفتقر للتغيرات الطبيعية في طبقة الصوت والقوة والتي تنتج عن التنفس أو الانفعال العاطفي لدى المتحدث الحقيقي، قد يبدو الصوت كأنه روبوتي قليلا بدون تلك الفوارق البشرية الدقيقة، وقد تطورت النماذج المعاصرة بشكل كبير لتنتج مقاطع صوتية متضمنة العاطفة وبشكل لا يمكن تمييزه على الإنتاج الطبيعي.
- **عدم التطابق مع حركة الشفاه:** في حالة الفيديوهات، قد يوجد تأخير طفيف أو عدم اتساق بين الصوت وحركة فم الشخصية، على سبيل المثال، تسمع كلمة بينما حركة الشفاه لا تطابقها تماما بالوقت الحقيقي، مما يشير إلى أنه صوت مركب وليس نابعا من المتحدث الحقيقي.\

3- **مؤشرات سلوكية أو سياقية:**

- محتوى غير منطقي أو صادم: غالبا ما يحاول محتوى التزييف العميق استغلال عنصر الصدمة أو الاستعجال لدفع المشاهد لتصديقه أو تداوله بسرعة، لذا قد تجد أن مضمون الفيديو أو التسجيل غير معقول أو مفبرك دراميا.

4. طرق كشف التزييف العميق

مع تنامي خطر التزييف العميق، تتعدد وسائل كشفه وتقع على عاتق المستخدمين مسؤولية التحقق من صحة ما يشاهدونه أو يسمعونه، فيما يلي أهم طرق اكتشاف المحتوى المزيف:

1. الملاحظة البشرية والتحليل السياقي: الخطوة الأولى لكشف التزييف هي الحس النقدي للمحتوى، ويجب الانتباه إلى المؤشرات البصرية والسمعية غير المنطقية المذكورة أعلاه، والتحقق من سياق المحتوى، على سبيل المثال، هل من المعقول أن يقول الشخص أو الجهة ما يعرض في الفيديو؟ وهل يتوافق التاريخ والمكان مع الأحداث المذكورة؟ غالبا ما يساعد التحقق من مصادر أخرى موثوقة (كالتغطية الإخبارية أو التصريحات الرسمية) في كشف عدم صحة المقاطع المثيرة للشك، فالمحتوى المزيف كثيرا ما يكون معزولا عن سياق معروف، أو يتضمن أحداثا مستبعدة تستوجب التحقق،
2. فحص البيانات الوصفية وخصائص الملف: يمكن للتفاصيل التقنية أن تكون مؤشرا إضافيا، على سبيل المثال، قد تفتقر الفيديوهات أو المقاطع الصوتية المزيفة إلى البيانات الوصفية (Metadata) الأصلية التي تسجلها أجهزة التصوير أو التسجيل في العادة (مثل معلومات الكاميرا أو الموقع

والتوقيت)، حيث أن عدم وجود هذه البيانات أو وجود بيانات مخالفة للمنطق (كأن يدعي المقطع أنه صور بكاميرا غير متوافقة مع جودته) قد يشير إلى حدوث تعديل أو توليد اصطناعي.

لكن ينبغي التنبيه أن غياب البيانات الوصفية ليس دليلا قاطعا على التزييف، لذا يجب اعتباره مجرد قرينة مساعدة ضمن مجموعة قرائن أخرى.

3. أدوات الكشف التزييف: تتوفر على الإنترنت أدوات مجانية ومدفوعة للمساعدة في تحليل الوسائط وكشف زيفها، تقوم هذه الأدوات بتحليل كل إطار في الفيديو أو كل عينة صوتية بحثا عن أي تشوهات أو أنماط رقمية مريبة، على سبيل المثال، يمكنها اكتشاف التغييرات البكسلية التي لا يلتقطها نظر الإنسان، أو كشف الفروقات في توقيت حركة الشفاه مقارنة بالصوت، وقد كثفت المؤسسات البحثية وشركات التكنولوجيا في الآونة الأخيرة جهودها لتطوير تقنيات كشف تلقائي للتزييف العميق، من الأمثلة البارزة أداة Sensity AI التي تحلل الصور والفيديو على مستوى البكسل لاكتشاف أي إشارات للتلاعب الرقمي، إذ تعمد خوارزميتها إلى رصد الاختلافات البنيوية الدقيقة في درجات الألوان والحدود والتي يصعب على نماذج التزييف توليدها بشكل متقن، كذلك ابتكرت شركة إنتل أداة FakeCatcher القادرة على كشف الفيديوهات المزيفة (بالزمن الحقيقي) عبر تحليل نمط تدفق الدم في وجه الشخص الظاهر في الفيديو، تعتمد هذه التقنية على حقيقة أن نبض الدم وتغير لونه تحت الجلد (في الجبهة والخدين مثلا) يحدث بشكل طبيعي لدى البشر الأحياء ويظهر في الفيديو الحقيقي (وهو أمر تعجز خوارزميات التزييف الحالية عن محاكاته بدقة).

ظهر أيضا العديد من البرمجيات الأكاديمية والتجارية التي تمكن المؤسسات (وحتى الأفراد) من رفع أي فيديو أو ملف صوتي مشبوه ليتم التحقق منه باستخدام نماذج كشف مدربة خصيصا لهذا الغرض.

على الرغم من ذلك، فإن معركة الكشف والتزييف هي سباق مستمر، فمع كل تحسن يطرأ على خوارزميات التزييف العميق وقدرتها على إنتاج تفاصيل أكثر إقناعا (مثل تحسن توليد ملامح اليد والأصابع التي شكلت سابقا نقطة ضعف للتزييف)، تزداد صعوبة مهمة أدوات الكشف التي يجب أن تواكب هذا التطور، ولا يزال مجال مكافحة التزييف العميق نشطا في مراكز الأبحاث؛ إذ من المتوقع أن نشهد مستقبلا ظهور أشكال جديدة وأكثر تعقيدا من التزييف، تقابلها أيضا أساليب كشف مبتكرة، لذا يتعين على الجهات المعنية والمستخدمين الاستمرار في تحديث معارفهم وأدواتهم لكسب هذه المعركة التقنية المتسارعة.

5. حقائق وأرقام حول التزييف العميق

تشير الأرقام إلى تنامي سريع في حجم وانتشار التزييف العميق عالميا خلال العامين الأخيرين، وفيما يلي بعض المؤشرات الرقمية التي تسلط الضوء على واقع التزييف العميق عالميا في الفترة 2023–2024، مع بيان مدى انتشار هذه الظاهرة وعدد ضحاياها وأنواع الجرائم المرتبطة بها:

المؤشر العالمي الرئيسي	القيمة/الوصف	المصدر
الزيادة العالمية في حالات احتيال التزييف العميق (2023–2022)	أكثر من 10 أضعاف (ارتفاع بنسبة **1000%** تقريبا في سنة واحدة)	بيانات شركة Sumsub (2023)
توقعات مستقبلية لنسب المحتوى المولد عبر تقنيات التزييف العميق	**90%** من المحتوى المتوفر في الشبكة العنكبوتية قد يكون مولدا بحلول العام 2026	تقرير البرلمان الأوروبي (2025)
نسبة ضحايا حيل التزييف العميق الصوتي الذين تكبدوا خسائر	**77%** من الذين استهدفوا برسائل صوتية مزيفة أفادوا بفقدان أموال نتيجة ذلك	دراسة McAfee للأمن الرقمي (2023)
نسبة حالات التزييف العميق ضمن جرائم الاحتيال بالعملات الرقمية (2023)	**88%** من حالات الاحتيال عبر التزييف العميق استهدفت قطاع العملات المشفرة، تليها **8%** في قطاع التكنولوجيا المالية	تقرير أمني (2024)
حصة محتوى التشهير من فيديوهات التزييف العميق	يقدر بأكثر من **95%** من إجمالي فيديوهات التزييف العميق المنتشرة على الإنترنت، وغالبية الضحايا نساء (بينهن آلاف الشخصيات المشهورة)	تقرير Sensity للتهديدات (2020) "مشتق من دراسة للبرلمان الأوروبي بشأن التعامل مع التزييف العميق في السياسة الأوروبية (2021)"
استهداف شخصيات عامة بفيديوهات التزييف العميق بهدف التأثير	ما لا يقل **عن 500 ألف مقطع صوتي ومرئي مزيف** تمت مشاركته عبر مواقع التواصل الاجتماعي في 2023	تقرير DeepMedia (2023)

هذه الأرقام تكشف مدى الانتشار والضرر الذي بلغه التزييف العميق عالميا، فقد تصاعدت وتيرة الهجمات والتزويرات اعتمادا على هذه التقنية بشكل هائل خلال فترة قصيرة، فعلى سبيل المثال، تشير البيانات إلى أن معدل حوادث التزييف العميق قفز إلى عشرة أضعاف في عام واحد، بل سجلت بعض المناطق زيادات غير مسبوقة (كما في الولايات المتحدة حيث زادت حوادث الاحتيال بالتزييف العميق بنسبة 1740% خلال 2022 (هذا التصاعد يعني أن آلاف المقاطع المزيفة تنتج وتبث عبر الإنترنت سنويا)[1]، بل وتوقعت وكالة الشرطة الأوروبية يوروبول أنه بحلول 2026 قد يصبح 90% من المحتوى الرقمي على الإنترنت مولدا بصورة اصطناعية[2]، وهي نسبة صادمة تعكس خطورة الموقف إذا لم تتخذ إجراءات فعالة.

على مستوى ضحايا التزييف العميق، يتضح أن المخاطر تطال كلا من الأفراد والمؤسسات على حد سواء، إذ يكشف مسح لشركة McAfee الأمنية أن حوالي ربع المستخدمين (26%) صادفوا خدعة عبر التزييف العميق خلال 2024، وأن قرابة 10% من الناس وقعوا فعلا ضحية لمثل هذه الخدع وفقدوا أموالا أو بيانات، ومن أبرز تلك الخدع استخدام الصوت المزيف لاستدراج الضحايا: فقد بينت دراسة أن واحدا من كل عشرة أشخاص تلقى فعليا رسالة أو اتصالا بصوت مستنسخ يطلب مساعدة مالية طارئة، وأن 77% ممن وصلتهم هذه الرسائل انطلت عليهم الحيلة وتكبدوا خسائر مالية، وفي كثير من الأحيان، تجاوزت الخسائر لكل ضحية ألف دولار أمريكي، بل إن 7% من الضحايا خسروا مبالغ ما بين 5 آلاف إلى 15 ألف دولار نتيجة تلك العمليات الاحتيالية[3].

بالنظر إلى أنواع الجرائم المرتبطة بالتزييف العميق، يمكن تصنيفها في ثلاث فئات رئيسية: الاحتيال المالي، والتشهير، والتلاعب السياسي والإعلامي، حيث تشمل الفئة الأولى جرائم النصب والابتزاز وسرقة الهوية التي تهدف للربح المادي، وقد أصبحت تقنيات التزييف العميق أداة مفضلة للمحتالين الإلكترونيين حول العالم، على سبيل المثال، فإن 88% من حالات التزييف العميق المكتشفة في 2023 كانت مرتبطة بعمليات احتيال في قطاع العملات الرقمية[4]، وشهدت نفس الفترة ارتفاعا حادا في استخدام المقاطع المزيفة لاجتياز إجراءات التحقق الأمني في الخدمات المالية (كإظهار فيديو مزيف للوجه لاجتياز أنظمة التعرف على العميل عن بعد)، بشأن الفئة الثانية، تعلقت الفئة بالتشهير وانتهاك الخصوصية، وتشير الدراسات إلى أن هذا النوع يشكل غالبية المحتوى المزيف على الإنترنت، إذ تفوق نسبته 95% من إجمالي فيديوهات التزييف العميق المنتشرة،، حيث أن غالبية الضحايا هنا من النساء[5]، بينهن شخصيات مشهورة تعرضت صورهن وفيديوهاتهن للتلاعب لإنتاج محتوى زائف، وهذا النوع يمثل جريمة تشهير واعتداء على الخصوصية قد يسبب أضرارا نفسية ومعنوية جسيمة.

أما الفئة الثالثة فهي التلاعب السياسي والإعلامي، حيث تستخدم الفيديوهات والصوتيات المزيفة لنشر أخبار أو مشاهد كاذبة بغرض التأثير على الرأي العام أو سمعة خصوم سياسيين، وقد رصد بالفعل تداول مكثف لهذا المحتوى؛ حيث أنه ووفقا لديب ميديا DeepMedia الرائدة في مجال تقنيات كشف التزييف العميق، فقد تمت مشاركة نحو 500 ألف مقطع فيديو أو صوت مزيف عبر وسائل التواصل الاجتماعي في عام 2023 معظمها يستهدف شخصيات سياسية معروفة لتضليل الجمهور[6].

باختصار، ترسم هذه الإحصائيات صورة مقلقة لتفشي التزييف العميق عالميا خلال الفترة الأخيرة، فالتقنية تنتشر بوتيرة سريعة تفوق قدرات التوعية والمواجهة التقليدية، مخلفة وراءها ضحايا في مختلف الدول والفئات، وعليه، أصبح من الضروري أن تتكاتف الجهود الدولية – قانونيا وتكنولوجيا وتوعويا – للحد من هذه الظاهرة وحماية المجتمعات من آثارها الخطيرة.

6. جهود مواجهة التزييف العميق

أدركت الحكومات والمؤسسات حول العالم خطورة التزييف العميق المتصاعد، فسارعت خلال السنوات الأخيرة إلى اتخاذ خطوات قانونية وتقنية للحد من انتشاره ومحاسبة المسؤولين عنه، فيما يلي عرض لأبرز جهود الدول والمؤسسات في مواجهة هذه الظاهرة:

1. **إطارتنظيمي وتشريعي عل المستوى الدولي:** تعمل العديد من الحكومات على تحديث قوانينها لسد الثغرات المتعلقة بالتزييف العميق، ففي الاتحاد الأوروبي، تم إدراج معايير خاصة للتعامل مع المحتوى الاصطناعي في قانون الذكاء الاصطناعي الشامل (AI Act) الذي أقر في 2025، حيث يتطلب هذا القانون بشكل صريح وضع ترميز واضح أو علامة مميزة على أي محتوى تم توليده بواسطة الذكاء الاصطناعي لإعلام المستخدمين بحقيقته، على سبيل المثال، إن استخدمت شركة إعلانا صوتيا لشخصية مشهورة مقلد بواسطة الذكاء الاصطناعي، فعليها تضمين تنويه صريح بأن الصوت اصطناعي، كما فرض الاتحاد الأوروبي عبر قانون الخدمات الرقمية (DSA) مسؤوليات أكبر على منصات التواصل الاجتماعي لمراقبة وإزالة المحتوى المزيف الضار بسرعة، خاصة ذلك المتعلق بالانتخابات أو السلامة العامة، وبموجب هذا القانون، قد تواجه الشركات غرامات تصل إلى 6% من إيراداتها العالمية إذا تقاعست عن الحد من انتشار التزييف الضار على منصاتها.
2. **في المملكة المتحدة**، صدر قانون السلامة على الإنترنت (OSA) في 2023، الذي يجرم صراحة إنشاء أو مشاركة صور وفيديوهات مزيفة دون موافقة أصحابها، مع عقوبات بالسجن تصل إلى سنتين، يلزم القانون أيضا المنصات الإلكترونية باتخاذ تدابير استباقية لحماية المستخدمين من المحتوى

المولد بالذكاء الاصطناعي إذا كان هدفه إيذاءهم أو تضليلهم، تحت طائلة غرامات تصل إلى 10% من الدخل السنوي للشركة.

3. **في الولايات المتحدة**، وعلى الرغم من غياب قانون فيدرالي شامل حتى وقت قريب، اعتمدت معظم الولايات تشريعات خاصة بها لمواجهة التزييف العميق، حتى منتصف عام 2025، كانت أكثر من 45 ولاية أمريكية قد سنت قوانين تتعلق بالتزييف العميق، تركز بعض هذه القوانين على مكافحة التزييف في سياق الانتخابات (منع نشر فيديوهات مزيفة لمرشحين قبيل التصويت) والبعض الآخر يستهدف المحتوى المزيف والتشهير، فعلى سبيل المثال، لدى ولايات ككاليفورنيا وتكساس وفرجينيا قوانين تسمح لضحايا التشهير بمقاضاة من ينتجها أو ينشرها للحصول على تعويضات مالية وأوامر قضائية لوقف انتشارها، وعلى المستوى الفيدرالي، أقر الكونغرس الأمريكي حديثا قانونا يعرف باسم قانون "أوقفه" أو "TAKE IT DOWN" Act لعام 2023، ووقعه الرئيس ليصبح ساريا في مايو 2025، يجرم هذا القانون نشر أو التهديد بنشر أي محتوى مزيف (سواء كان عميق التزييف أم حقيقي) دون إذن الضحية، والأهم أنه يلزم منصات الإنترنت بوضع آلية سريعة لتمكين الضحايا من الإبلاغ عن تلك المواد المزيفة والمطالبة بإزالتها خلال 48 ساعة من الإبلاغ، مع فرض سلطة إنفاذ ذلك على هيئة التجارة الفيدرالية، علاوة على ذلك، هناك مشروعات قوانين أخرى قيد النظر في الكونغرس مثل قانون DEEPFAKES للمساءلة الذي يهدف إلى توفير حق مقاضاة فيدرالي لضحايا المحتوى المزيف الضار، وكذلك قانون وسم المحتوى الاصطناعي AI Labeling Act الذي يجبر مطوري أدوات التوليد الاصطناعي على تضمين وسوم أو علامات مائية واضحة في الصور

والفيديوهات المنتجة، هذه التحركات التشريعية تعكس تنامي الإدراك الرسمي لخطورة التزييف العميق على الأمن القومي وحقوق الأفراد.

4. **القارة الآسيوية وجهود الصين ومحيطها**: اتخذت بعض دول آسيا والمحيط الهادئ زمام المبادرة في سن تنظيمات صارمة للتزييف العميق، الصين على وجه الخصوص تبنت لوائح شاملة دخلت حيز التنفيذ في أوائل 2023، تلزم هذه القواعد أي مزود لخدمات المحتوى الاصطناعي بوضع علامة مائية أو تنويه على الصور والفيديوهات المولدة، كما تحظر استخدام التزييف العميق في أي أغراض غير مشروعة مثل الاحتيال المالي أو الإضرار بالأمن القومي أو سمعة الأفراد، وألزمت السلطات الصينية المنصات بالتحقق من الهوية الحقيقية للمستخدمين قبل تمكينهم من استخدام أدوات إنشاء المحتوى العميق، لمنع إساءة الاستخدام مجهول المصدر، وتفرض الصين عقوبات صارمة على المخالفين تشمل الغرامات العالية بل وحتى الحجز (التوقيف) في بعض الحالات الخطرة، دول أخرى في آسيا مثل كوريا الجنوبية وسنغافورة وضعت مسودات تنظيمية مشابهة تركز على شفافية المحتوى وتحميل المنصات مسؤولية إزالة التزييف الضار بسرعة، لضمان عدم استغلال التقنيات الجديدة في التضليل الإعلامي أو الجرائم الإلكترونية.

7. المخرجات

ختاما، يمكن القول إن مواجهة التزييف العميق أصبحت مسؤولية جماعية عالمية تتطلب تضافر الأدوات التقنية مع الأطر القانونية ونشر الوعي المجتمعي، فعلى الرغم من التقدم الملحوظ في رصد المحتوى المزيف ووضع قوانين رادعة، يواصل محترفو التزييف تطوير أساليبهم بوتيرة متسارعة، لذا

ستبقى معركة التزييف والكشف مستمرة في المستقبل المنظور، ويتوقع أن يشهد المستقبل نمطا جديدا من أنماط التزييف ووسائل وأساليب كشفها، والجدير بالذكر أنه كلما انتشرت تقنيات الذكاء الاصطناعي التوليدي أكثر في حياتنا اليومية، تعاظمت الحاجة لوضع حواجز أمان رقمية تكفل استفادة البشرية من فوائد هذه التقنيات مع تقليل جوانبها السلبية.

إن تحقيق هذا التوازن الدقيق هو التحدي الأبرز الذي يواجهنا في عصر ثورة المعلومات والذكاء الاصطناعي، والمؤكد أن الوعي والمعرفة هما خط الدفاع الأول، فكلما فهمنا هذه التقنيات وآليات عملها، سهل علينا أن نميز الحقيقة من التزييف ونحصن مجتمعاتنا من خدع العصر الرقمي.

References

1. Sumsub (2023) *Global deepfake incidents surge tenfold from 2022 to 2023, Sumsub research reveals*, Sumsub Newsroom, 20 September. Available at: https://sumsub.com/newsroom/sumsub-research-global-deepfake-incidents-surge-tenfold-from-2022-to-2023/
2. European Parliament Research Service (2025) *The rise of deepfakes and the threat to democracy*, Briefing No. 775855. Available at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI(2025)775855_EN.pdf
3. McAfee (2023) *Artificial imposters: Cybercriminals turn to AI voice cloning for a new breed of scam*, McAfee Blogs, 4 April. Available at: https://www.mcafee.com/blogs/privacy-identity-protection/artificial-imposters-cybercriminals-turn-to-ai-voice-cloning-for-a-new-breed-of-scam/
4. Security.org (2024) *Deepfake statistics: Everything you need to know about AI-generated content*. Available at: https://www.security.org/resources/deepfake-statistics/
5. European Parliament Research Service (2021) *Tackling deepfakes in European policy*, Study No. 690039. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf
6. Bradley Arant Boult Cummings LLP (2024) *Audio deepfakes: Cutting-edge tech with cutting-edge risks*, 11 January. Available at: https://www.bradley.com/insights/publications/2024/01/audio-deepfakes-cutting-

مهندس/ عبدالرحمن بابكر
خبير أنظمة

مهندس/ عبدالرحمن بابكر - خبيرأنظمة أمنية بخبرة تتجاوز ١٠ سنوات في مجال الأمن المادي والذكاء الاصطناعي، حاصل على بكالوريوس الهندسة الكهربائية من جامعة قطر، محترف في تصميم وتنفيذ أنظمة الأمن المادي، وتحليل البيانات، وأنظمة الذكاء الاصطناعي المحلية والسحابية، إضافة إلى الأنظمة المدمجة والتقنيات الناشئة. حاصل على عدة شهادات دولية في أمن المعلومات، والذكاء الاصطناعي، وتحليل البيانات، وإدارة المشاريع.