

الخوارزميات والتزييف العميق في الإعلام الرقمي Algorithms and Deepfakes in Digital Media

د/ إيمان بهلول

جامعة البليدة 2 لونيبي علي، الجزائر

i.behloul@univ-blida2.dz

<https://orcid.org/0009-0001-0203-260X>

تاريخ الاستلام: 2026/05/06 تاريخ القبول: 2026/05/08 تاريخ النشر: 2026/06/10

ملخص:

تلعب الخوارزميات دورا كبيرا في الإعلام الرقمي من خلال تحديد المعلومات التي تصل إلى الجمهور عبر المنصات لتحقيق التفاعل والانتشار الواسع، مما يسهم في نشر معلومات مغلوطة تمس بالخصوصية الفردية، والأمن المجتمعي خاصة مع تطور تقنيات التزييف العميق التي تجعل التمييز بين الصور والفيديوهات المفبركة والحقيقية أمرا صعبا لجودتها العالية، وقد استدعى هذا تطوير أساليب تعتمد على الذكاء الاصطناعي للكشف عن المحتوى المزيف، كما قامت العديد من المنصات الإعلامية بتطوير خوارزميات تحد من انتشار الأخبار الكاذبة، واستخدام خوارزميات التجميع للتعرف على البوتات التي يديرها كيان واحد، إضافة إلى إنشاء أدوات تمكن الصحفيين والمستخدمين من التحقق من مصداقية الصور والفيديوهات.

الكلمات المفتاحية: خوارزميات؛ تزييف عميق؛ إعلام رقمي؛ ذكاء اصطناعي؛ منصات رقمية.

Abstract:

Algorithms play a significant role in digital media by determining the information that reaches the audience through platforms to achieve interaction and widespread dissemination, which contributes to the spread of false information that affects individual privacy and community security, especially with the development of deepfake technologies that make it difficult to distinguish between fabricated and real images and videos due to their high quality. This has required the development of methods based on artificial intelligence to detect fake content. Additionally, many media platforms have developed algorithms to limit the spread of fake news and used aggregation algorithms to identify bots managed by a single entity, as well as creating tools that enable journalists and users to verify the authenticity of images and videos.

Keywords: Algorithms; Deepfakes; Digital media; Artificial intelligence; Digital platforms.

1. مقدمة

أحدثت الخوارزميات تحولات جذرية في المجال الإعلامي، حيث لم تعد مجرد أدوات تقنية مساعدة، بل أصبحت فاعلا رئيسيا في إنتاج المحتوى وتوزيعه. فقد بات بإمكانها صياغة التقارير الإخبارية في وقت قياسي، مما يخفف العبء عن الصحفيين ويفسح لهم المجال للتفرغ للأعمال التحليلية. وإلى جانب ذلك، تلعب الخوارزميات دور "حارس البوابة" على منصات التواصل الاجتماعي، فهي تتحكم في تدفق المعلومات للجمهور، إذ تقوم خوارزميات الترشيح، المعتمدة على معايير مثل معدلات الإعجاب والمشاركة والتفاعل، بتقديم محتوى يهدف إلى تحقيق أقصى قدر من الجذب والاستقطاب مما يجعل الجمهور داخل فقاعات المعلومات وغرف الصدى. غير أن هذا الدور لا يخلو من مخاطر، إذ يسهل انتشار المعلومات المضللة والأخبار الزائفة، خاصة مع ظهور تقنية التزييف العميق التي تعتمد على الذكاء الاصطناعي لتوليد صور وفيديوهات ذات جودة عالية يصعب تمييزها عن المحتوى الأصلي، مما يجعل عملية التحقق من مصداقيتها أكثر تعقيدا. وفي ظل هذا الواقع، يبرز تساؤل محوري: كيف تسهم الخوارزميات وتقنيات التزييف العميق في إعادة تشكيل المشهد الإعلامي؟ وما انعكاساتها على مصداقية الأخبار؟ وما هي طرق الحد من مخاطرها؟

2. الإطار المفاهيمي: الخوارزميات وتطبيقاتها الإعلامية

1.2. تعريف الخوارزميات

وفقا لقاموس "ويبستر" تعرف الخوارزمية بأنها: *إجراء خطوة بخطوة لحل مشكلة أو لإنجاز هدف ما*. ويستخدم هذا المصطلح في الوقت الحاضر غالبا للإشارة إلى مجموعة القواعد التي تتبعها آلة (وخاصة الحاسوب) من أجل تحقيق غاية محددة. وفي حالة منصات التواصل الاجتماعي مثل فيسبوك، يمكن للخوارزميات - من بين مهام أخرى - أن تحدد أي محتوى سيشاهده المستخدم في وقت معين، وبأي ترتيب، أثناء استخدامه للمنصة. (Presuel & Sierra, 2019, p. 262)

يمكن تعريف الخوارزميات وأشكال الأتمتة الحاسوبية المرتبطة بها من منظور تقني أو اجتماعي. فالتعريفات التقنية، الشائعة في علوم الحاسوب والمعلومات، تؤكد أن الخوارزمية تتبع سلسلة من الخطوات أو القواعد المصممة مسبقا من أجل حل مشكلة. أما التعريفات الاجتماعية، الأكثر شيوعا في دراسات الإعلام والاتصال، فتتركز على ديناميات التفاعل بين الإنسان والآلة، والترتيبات المؤسسية، والظروف البيئية (إلى جانب عوامل أخرى) التي تعطي للخوارزميات شكلها كنتاج اجتماعي، ثقافي ومادي. (Thurman et al., 2019, p. 980,981)

2.2. التطبيقات الخوارزمية في المجال الإعلامي:

شهد مجال الصحافة تغييرات جذرية خلال السنوات الماضية، وتحمل التكنولوجيا الحديثة النصيب الأكبر من المسؤولية عن هذا التحول العميق في صناعة الأخبار. وما يلي هو تحليل ومراجعة لأهم المجالات التي أحدثت فيها التكنولوجيا الحاسوبية تغييرات بارزة في هذا الميدان. وتجدر الإشارة إلى أن استخدام الخوارزميات في الصحافة يمكن أن يتجاوز بكثير الفئات المذكورة هنا، غير أن ما يلي يمثل المجالات الأكثر ارتباطاً بخدمات صناعة الأخبار الحديثة وتتمثل هذه المجالات تحديداً في:

- إنتاج المحتوى الآلي؛ يعد إنتاج الأخبار عبر الخوارزميات واحداً من أبرز وأحدث التحولات في المجال الصحفي. يقوم على تقنيات مثل توليد اللغة الطبيعية (NLG) التي تمكن البرامج من صياغة نصوص إخبارية انطلاقاً من بيانات رقمية منظمة. بدأت التجربة مع برنامج Quakebot في Los Angeles Times عام 2014، وتطورت حتى اعتمدت مؤسسات كبرى مثل *Forbes* و *New York Times* على هذه الأدوات.

الميزة الكبرى هي السرعة وتوفير وقت الصحفيين للمهام الإبداعية والتحقيقية، غير أن هذه الممارسات أثارت جدلاً واسعاً حول إمكانية تهديدها لمستقبل الصحفيين وإحلال الآلة محل الإنسان. ورغم المخاوف، يرى بعض الباحثين أن هذه التقنيات ستعزز الممارسة الصحفية بدل أن تقصمها (Kotenidis & Veglis, 2021, pp. 247, 248).

- تنقيب البيانات؛ في عصر "انفجار البيانات"، أصبح الصحفيون بحاجة إلى أدوات قادرة على فرز كميات هائلة من المعلومات لاستخراج ما هو مفيد، فتتقنب البيانات يمثل المرحلة الأهم في "اكتشاف المعرفة"، إذ يتيح العثور على أنماط واتجاهات خفية في مجموعات البيانات الضخمة (Big Data) التي لا يمكن للبشر التعامل معها يدوياً.

هذه التقنية تساعد الصحفيين في تعزيز قدراتهم البحثية والتحليلية، واكتشاف اتجاهات اجتماعية أو اقتصادية جديدة، وحتى تخصيص المحتوى للجمهور. لكنها تثير بدورها إشكالات أخلاقية تتعلق بمراقبة المستخدمين وخصوصيتهم، ما يستدعي وضع ضوابط تنظيمية واضحة (Kotenidis & Veglis, 2021, p. 249).

- نشر الأخبار؛ تتم اليوم عملية توزيع الأخبار عبر وسائط مثل المجموعات الإخبارية ومحركات البحث ومواقع التواصل الاجتماعي، وجميعها تعتمد على خوارزميات تحدد ما يصل إلى الجمهور. هنا ظهر ما يعرف بـ "الروبوتات الإخبارية" (News Bots) التي تنشر الأخبار وتتفاعل مع المستخدمين لضمان وصول المحتوى لأوسع جمهور ممكن.

لكن هذه الآليات تثير جدلاً حول الشفافية والتحكم، حيث تعمل كـ "حراس بوابة" جدد يتحكمون في تدفق الأخبار، مما قد يؤدي إلى تهميش بعض الأصوات وتفاقم مشكلات الأخبار الكاذبة والمعلومات المضللة. (Kotenidis & Veglis, 2021, p. 250)

-تحسين المحتوى؛ يتمثل هذا المجال في تقديم محتوى إخباري شخصي وموجه لكل مستخدم وفق اهتماماته، من خلال خوارزميات مشابهة لتلك التي تستخدمها محركات البحث. وقد أثبتت هذه الآليات فعاليتها في توفير الوقت وزيادة التفاعل عبر تخصيص الأخبار والخدمات الإعلامية. غير أن هذا التخصيص يثير مخاطر، أهمها فقاعات الترشيح (Filter Bubbles) وغرف الصدى (Echo Chambers) التي تحاصر المستخدم في دائرة ضيقة من الآراء المتشابهة وتحد من تنوع المعلومات، ما قد يؤدي إلى تجزئة الرأي العام. (Kotenidis & Veglis, 2021, p. 251)

3. الخوارزميات والتضليل الإعلامي

خلال العقد الأول من إطلاق تويتر، اقتصر دور الخوارزميات على اقتراح الحسابات المناسبة للمتابعة، فيما ظل الخط الزمني يظهر التغريدات وفق ترتيبها الزمني فقط. غير أن عام 2016 شكل نقطة تحول، إذ بدأت المنصة بالاعتماد على خوارزميات التعلم الآلي لترتيب المحتوى استناداً إلى درجة شعبيته ومدى توافقه مع اهتمامات المستخدم المفترضة، بدلاً من الاقتصار على توقيت النشر. وفي مطلع عام 2019، اتسع نطاق هذا التوجه أكثر، حيث بات تويتر يعرض للمستخدمين تغريدات مختارة آلياً من حسابات يتابعها الأشخاص الذين يتابعهم المستخدم نفسه.

قد تبدو هذه التطورات على تويتر خطوات طبيعية في مسار تحسين تجربة المستخدم، غير أن جوهر المسألة يكمن في أن أي خوارزمية تتحكم في تحديد المحتوى المعروض قد تسهم في تضخيم الأخبار المضللة إلى ما يتجاوز نطاق انتشارها الطبيعي. فغالبا ما تبنى هذه الخوارزميات على هدف تعظيم معدلات التفاعل، ما يعني أن نظريات المؤامرة أو الأخبار المثيرة للجدل مثل: الادعاء بأن فيروس كوفيد-19 سلاح بيولوجي حكومي، أو أن انتخابات عام 2020 قد تم التلاعب بها، فهي تقدم لجمهور أوسع بمجرد أن تسجل مستويات مرتفعة من التفاعل، وهذا النمط لا يقتصر على تويتر وحده، بل يمتد إلى فيسبوك وإلى معظم المنصات الرقمية التي تعتمد على خوارزميات التصنيف والتوصية.. (Giansiracusa, 2021, p. 117)

من أبرز الأمثلة التي تظهر هذه الدينامية حالة صحيفة *Epoch Times*، وهي مؤسسة إعلامية مشبوهة تحولت من صحيفة صغيرة منخفضة التمويل توزع مجاناً في شوارع نيويورك إلى واحدة من أبرز ناشري المحتوى الرقمي في الولايات المتحدة ومصدر رئيسي للمعلومات المضللة اليمينية.

تشير تحقيقات نيويورك تايمز عام 2020 إلى أن تراجع عائدات الصحيفة دفعها لتكثيف إنتاج مقالات قصيرة وسطحية بحثاً عن الانتشار السريع. ومع اقتراب انتخابات 2016، اتجهت المؤسسة بقوة نحو دعم دونالد ترامب ونشر نظريات مؤامرة مرتبطة به. كما كشفت التحقيقات أن *Epoch Times* وضعت استراتيجية سرية لاستغلال خوارزميات فيسبوك لنمو شعبيتها وتمويلها، حيث أطلقت شبكة صفحات في فيتنام مليئة بمقاطع

فيديو ترفيهية ومحتوى مؤيد لترامب، مدعومة ببوتات لرفع التفاعل بشكل مصطنع. هذا التفاعل الوهمي التقطته خوارزميات فيسبوك ليترجم إلى شعبية فعلية، فانتشرت الصفحات بسرعة.

توسعت التجربة لاحقا إلى بقية فروع المؤسسة، ما مكنها من بناء قاعدة جماهيرية بملايين المتابعين. واستمر الارتباط بترامب ليصل إلى حد مشاركة مقالاتها على نطاق واسع من قبله وأفراد أسرته خلال رئاسته، بل واستضافت مسؤولين من إدارته في مقابلات. وبذلك تحولت *Epoch Times* إلى إمبراطورية إعلامية ضخمة مستندة جزئيا إلى استغلال خوارزميات فيسبوك، سواء عبر أساليب محظورة كبوتات التفاعل أو عبر طرق مسموحة لكنها خطيرة أخلاقيا كسرد الأخبار الزائفة في قالب إعلامي متزن ظاهريا. (Giansiracusa, 2021, pp. 178–180)

4. التزييف العميق : التعريف والأنواع

1.4. تعريف التزييف العميق:

تشكل تقنية التزييف العميق (Deepfake) جزءا أساسيا من ثورة الذكاء الاصطناعي، حيث تعتمد بشكل رئيسي على: تعلم الآلة (Machine Learning) والتعلم العميق (Deep Learning) ويتطلب إنشاء التزييف العميق كميات كبيرة من البيانات المرئية والصوتية الأصلية لأغراض التدريب، حيث تقوم خوارزميات معقدة بتحليل هذه البيانات وإنشاء نماذج رقمية دقيقة للأشخاص المراد محاكاتهم.

يتعلم نموذج التزييف العميق من هذه البيانات لإنشاء نسخة رقمية مطابقة للشخص المستهدف، تشمل ملامح وجهه، تعبيراته، حركاته، صوته وتعمل الشبكات العصبية العميقة (Deep Neural Networks) على ربط هذه المعلومات معا لتوليد محتوى جديد يبدو واقعا للغاية، لكنه في الحقيقة مزيف بالكامل.

الشبكات التوليدية التنافسية (GANs): تلعب الشبكات التوليدية التنافسية (Generative Adversarial Networks – GANs) دورا محوريا في عملية التزييف العميق، تتكون هذه الشبكات من نموذجين يعملان معا:

- المولّد (Generator): يقوم بإنشاء المحتوى المزيف (مثل الصور أو مقاطع الفيديو) بناء على البيانات التي تم تدريبه عليها.
- المميّز (Discriminator): يقوم بتقييم المحتوى الذي أنتجه المولّد ويحدد ما إذا كان حقيقيا أو مزيفا.

مع مرور الوقت، يتحسن أداء المولد في خداع المميز، مما يؤدي إلى إنتاج محتوى مزيف أكثر إتقاناً ومطابقة للواقع، وتتكرر عملية التدريب هذه آلاف المرات، مما يؤدي في النهاية إلى تحسين جودة التزييف بشكل متواصل حتى يصبح من الصعب تمييزه عن المحتوى الأصلي. (فرجاني، 2024، p. 91)

2.4. أنواع التزييف العميق:

يمكن تصنيف أنواع التزييف العميق (Deepfakes) إلى عدة أشكال رئيسية:

- إبدال الوجوه (Face-Swap) الشكل الأكثر شيوعاً للتزييف العميق، حيث تستبدل وجوه المشاهير، خصوصاً النساء، في مقاطع إباحية غير توافقية. وقد مثلت هذه الفئة نحو 96% من الفيديوهات المزيفة حتى عام 2019، مدفوعة بوفرة صور وفيديوهات المشاهير التي تستخدم لتدريب الخوارزميات (Giansiracusa, 2021, p. 49).

- تحريك الشفاه (Lip-sync): حيث يعدل الفيديو لجعل حركات الفم متوافقة مع تسجيل صوتي معين. (Nguyen et al., 2022, p. 1)

- إعادة التمثيل (Reenactment) ، فهي تقنية تحاكي تحريك الدمى، حيث تتحرك ملامح شخصية مستهدفة وفق تعابير شخص آخر. المثال البارز هو فيديو باراك أوباما الذي أنتجته BuzzFeed عام 2018، حيث جرى دمج صوت وصورة مزيفين بشكل متقن لإظهار أوباما وكأنه يتحدث بكلمات لم يقلها قط. (Giansiracusa, 2021, p. 50)

- التزييف عبر صورة واحدة (One-shot Deepfake) كما في تطبيق Zao الصيني أو تقنية سامسونج، حيث يكفي إدخال صورة ثابتة واحدة ليتم إدماجها أو تحريكها في مقاطع فيديو بطريقة مقنعة، بفضل خوارزميات مدربة مسبقاً على ملايين المقاطع لتعلم أنماط حركة الوجه (Giansiracusa, 2021, p. 50).

- تحويل النص إلى فيديو (Transcript Editing) ، الذي طور بالتعاون مع Adobe ، تعديل النصوص المرافقة لمقاطع الفيديو بحيث يعاد توليد الصوت وحركات الفم لتتطابق مع النص الجديد. ورغم أن هذه التقنية ما تزال محدودة الاستخدام، إلا أنها تكشف عن مدى خطورة التزييف العميق في إعادة تشكيل الخطاب المرئي والمسموع (Giansiracusa, 2021, p. 51).

5. أساليب وأدوات الكشف عن التزييف العميق باستخدام الذكاء الاصطناعي

1.5. أساليب الكشف عن التزييف العميق

نظرا للتهديدات التي يسببها التزييف العميق، تم تطوير العديد من الأساليب للكشف عن هذا النوع من التلاعب، ويقصد بالأساليب الطرق التي تعتمد عليها خوارزميات الذكاء الاصطناعي للكشف عن التزييف العميق الذي أنتج أصلا بالذكاء الاصطناعي، حيث تعتمد تقنيات الكشف على عدة أساليب مختلفة منها:

- أسلوب تحليل الخصائص البصرية الدقيقة للكشف عن العيوب الدقيقة التي لا تكون مرئية بالعين البشرية، مثل عدم تطابق الإضاءة، وحواف مشوشة أو مشوهة، أو حركات غير طبيعية في الوجه أو العينين، ويعتمد هذا الأسلوب على أدوات التعلم العميق.

- أسلوب الكشف عن التناقضات الصوتية والبصرية من خلال الجمع بين تحليل المكونات الصوتية والمرئية في مقاطع الفيديو لكشف الاختلافات، مثل عدم تطابق الصوت مع حركة الشفاه، وعدم التزامن بين الحركات الجسدية والإيقاعات الصوتية، ومن هذه الأدوات التي تنتج هذا الأسلوب تقنيات دمج النماذج متعددة الوسائط (Multimodal Fusion Models)

- أسلوب تحليل الأنماط الزمنية للفيديو (Temporal Analysis) من خلال تحليل استمرارية الأنماط الزمنية في الفيديوهات المزيفة، إذ تظهر أحيانا انقطاعات غير طبيعية في الحركات، وتغيرات فجائية في الإطارات، ومن بين الأدوات التي تنتج هذا الأسلوب: شبكات LSTM أو GRU التي تركز على معالجة البيانات المتسلسلة.

- أسلوب تحليل معدل رمش العيون والأنماط الحيوية، ويركز هذا الأسلوب على دراسة الأنماط الحيوية البشرية الطبيعية مثل: معدل رمش العيون، والإيماءات الطبيعية، ومن بين الأدوات التي تعتمد على هذا الأسلوب خوارزميات التعلم الآلي التي تقارن الأنماط الطبيعية مع الأنماط المزيفة.

- أسلوب الكشف بالاعتماد على بصمة الفيديو الرقمية (Digital Video Fingerprinting)؛ يستخدم هذا الأسلوب تحليل البصمة الرقمية للمحتوى المرئي للتحقق من مصدره ومصداقيته، ومن الأدوات الشائعة تقنيات Blockchain لحفظ البيانات الأصلية ومقارنتها مع المحتوى المشكوك فيه.

- أسلوب الكشف القائم على تحليل الشبكات العصبية المزيفة، حيث يعمل هذا الأسلوب على تحديد الأنماط التي تولدها الشبكات التوليدية العميقة (Generative Adversarial Networks – GANs)، مثل التشوهات في توزيع البيانات، والبصمات المميزة للشبكات التوليدية، ومن بين الأدوات التي تعتمد هذا الأسلوب خوارزميات GAN-specific Detectors

وكل هذه الأساليب تستخدم تقنيات الذكاء الاصطناعي للكشف عن التلاعبات في المحتوى من خلال تدريب الأنظمة على التعرف على الأنماط التي تشير إلى التزييف، مثل الارتباك أو الإضاءة أو الحركة غير الطبيعية.

(خبريش, 846-845 pp. 2025) إلى جانب التحديات التقنية، تبرز مشكلة اجتماعية ونفسية خطيرة أطلق عليها مصطلح "الفجوة بين الكذب والحقيقة" إذ حتى لو تم إثبات أن مقطعاً معيناً هو تزييف عميق، غالباً ما يكون الضرر قد وقع، لأن الناس يميلون إلى تصديق ما شاهدوه بأعينهم، ومن الصعب تغيير قناعاتهم حتى أمام أدلة دامغة لاحقة، وهذا يطرح الحاجة إلى استراتيجيات وقائية تمنع انتشار هذه المقاطع قبل أن تتحول إلى أدوات تضليل جماعي. (Giansiracusa, 2021, p. 62)

2.5. أدوات الكشف عن التزييف العميق

يعد اكتشاف التزييف العميق (Deepfake) تقنية أساسية لمواجهة الأضرار المحتملة الناجمة عن الاستخدام الخبيث لخوارزميات التعلم العميق. ومع تطور تقنيات التزييف العميق، ازدادت الحاجة إلى أساليب فعالة للكشف عنها. فيما يلي ستة من أفضل أدوات اكتشاف التزييف العميق: (Ayush, 2023)

- منصة Sensity للذكاء البصري: تعد منصة Sensity Visual Intelligence أداة متطورة تستهدف اكتشاف التزييفات العميقة باستخدام تقنيات الذكاء الاصطناعي المتقدم. يمكن للمنصة تحليل الملفات المرئية بدقة عالية جداً واكتشاف أي مؤشرات للتلاعب عبر تقنيات مثل: الرؤية الحاسوبية، التعلم الآلي، ومعالجة اللغة الطبيعية.

تمتاز أيضاً بقدرتها على كشف أصعب أنواع التزييف العميق مثل تبديل الوجوه، وتوليد الصوت، أو تركيب الصور. كما أنها قوية في التحليل الفوري، مما يجعلها أداة لا غنى عنها لمكافحة التزييفات المنتشرة على وسائل التواصل الاجتماعي بسرعة وفعالية.

- منصة WeVerify: هي منصة تعاونية تستخدم لاكتشاف التزييف العميق والتحقق من المحتوى الرقمي. تجمع بين خبرة البشر وخوارزميات الذكاء الاصطناعي لرصد المحتوى المزيف. تستخدم تقنيات مثل التعرف على الوجه، الفحص الجنائي للصور، والبحث العكسي عنها.

كما تستند إلى شبكة من الصحفيين وخبراء التحقق الذين يساهمون في تأكيد النتائج وتوضيحها، مما يمنح المستخدمين فهماً أعمق لسياق المحتوى الذي يتعاملون معه.

- أداة Deepware: تعد أداة متقدمة لاكتشاف التزييف العميق باستخدام خوارزميات التعلم العميق لتحديد ما إذا كانت الصور أو مقاطع الفيديو أصلية أم مزيفة. تعتمد على تقنيات متطورة مثل الشبكات العصبية الالتفافية (CNNs)، وتدريب على قاعدة بيانات ضخمة من المحتوى الحقيقي والمزيف، ما يجعلها قادرة على اكتشاف أدق الفوارق.

تحلل الأداة عناصر متعددة في الصور والفيديوهات مثل: تعابير الوجه، الإضاءة، الظلال، الخلفية، الدقة، ونوع الملف، للكشف عن التلاعب.

-أداة Reality Defender: هي أداة قوية صممت لمواجهة ظاهرة الإعلام المزيف. تجمع بين خوارزميات التعلم الآلي وتقنيات الرؤية الحاسوبية لتحليل مقاطع الفيديو، والتسجيلات الصوتية، والصور عبر دراسة أنماط الوجه والكلام وغيرها.

تمتاز بقدرتها على مقارنة المحتوى المشبوه بقاعدة بيانات كبيرة من الأمثلة الواقعية، ورصد حتى التغييرات الدقيقة غير المرئية للعين البشرية، مثل تبديل الوجوه أو تعديل الصوت ونبرته.

-أداة Duck Duck Goose: أداة فعالة طورت خصيصاً لمكافحة التزييف العميق المنتشر. تستخدم خوارزميات متقدمة لرصد الشذوذ في ملامح الوجه أو الحركات غير الطبيعية مثل حركة الرأس أو العينين، بالإضافة إلى تغييرات الخلفية أو الإضاءة.

-منصة Truepic: هي منصة تقنية حديثة تعتمد على علم الأدلة الجنائية الرقمية وخوارزميات التعلم الآلي للتحقق من أصالة الصور والفيديوهات وكشف التزييف العميق.

تحلل التوقعات الرقمية للمحتوى المرئي وتبحث عن أي تغييرات دقيقة جداً قد لا يلاحظها الإنسان، مثل الأنماط غير الطبيعية في الصور والفيديوهات. وتضمن بذلك الحفاظ على مصداقية وأصالة المحتوى.

تأتي هذه الأدوات استجابة لتزايد التهديدات الناتجة عن التزييف العميق الذي يستغل في نشر الأخبار الكاذبة، تشويه السمعة أو الاحتيال.

6. الحد من الأخبار الزائفة عبر الخوارزميات والمنصات الرقمية

أظهرت الدراسات أن الأخبار الزائفة تنتشر عبر وسائل التواصل الاجتماعي بطريقة تختلف عن الأخبار التقليدية، وهو ما دفع شركات التكنولوجيا والباحثين إلى تطوير خوارزميات وتقنيات متقدمة للكشف عنها. من بين الأمثلة البارزة خوارزمية Fabula التي ركزت على تحليل أنماط التفاعل الشبكي بدلاً من محتوى المنشورات، حيث تعتمد على التعلم العميق لاكتشاف الأخبار الزائفة عبر متابعة مسارات انتشارها. كما طور فريق من Microsoft Research وجامعة ولاية أريزونا نظاماً يجمع بين تحليل المحتوى البسيط (مثل درجة العاطفة والانحياز) ودراسة بيانات المستخدمين، مستخدماً خوارزمية RoBERTa من فيسبوك، مما عزز القدرة على رصد الأخبار المضللة مبكراً قبل انتشارها (Giansiracusa, 2021, pp. 208, 209).

طبقت شركات وسائل التواصل الاجتماعي حلولاً تقنية للحد من انتشار الأخبار الزائفة والمحتوى الضار مثال: فيسبوك يمتلك تطبيق المراسلة واتساب الذي أثبت أنه قناة أخرى لانتشار المعلومات المضللة بشكل سريع، كما كان واتساب وسيلة أساسية لنشر محتوى اليمين المتطرف خلال انتخابات البرازيل عام 2018. ولمواجهة هذا، قلص فيسبوك عام 2018 عدد الأشخاص الذين يمكن إعادة توجيه الرسائل إليهم من 250 إلى 20 شخصاً، ثم جرى تقليص العدد لاحقاً إلى 5 فقط. وقبل انتخابات 2020، خفضت الشركة كذلك الحد الأقصى لإعادة توجيه الرسائل على تطبيق Messenger من 150 إلى 5 فقط.

وفي يناير 2020، أعلن فيسبوك عن خطوة واضحة: حظر تقنية التزييف العميق (Deepfake) على منصته، أي بشكل أدق أعلن عن حظر "الوسائط المضللة المعالجة"، بمعنى أي فيديو "تم تعديله أو تركيبه—بما يتجاوز التعديلات المتعلقة بالوضوح أو الجودة—بطريقة لا يمكن للشخص العادي اكتشافها، وبما قد يضلل المشاهد ليظن أن شخصاً ما قال كلمات لم يقلها فعلياً".

كما شمل الحظر مقاطع الفيديو التي تنتج عبر الذكاء الاصطناعي أو تقنيات التعلم الآلي من خلال الدمج أو الاستبدال أو الإقحام، بحيث تظهر الفيديو على أنه أصلي.

هذه الصياغة الطويلة تهدف أساساً إلى استهداف مقاطع التزييف العميق ذات الطابع الخادع، لكن كشف التزييف العميق، سواء عبر الخوارزميات أو بالوسائل اليدوية، يظل مهمة صعبة وربما يظل كذلك دوماً، مما يجعل تطبيق هذا الحظر على أرض الواقع معقداً. (Giansiracusa, 2021, p. 200)

إضافة إلى ذلك اعتمد فيسبوك في 2019 على تطوير تقنيات الذكاء الاصطناعي لمراقبة المحتوى، من خلال الانتقال من أسلوب التعلم الخاضع للإشراف الذي يتطلب بيانات موسومة يدوياً إلى التعلم الذاتي الإشراف الذي يتيح تدريب خوارزميات أكثر دقة على مجموعات بيانات أضخم، مثل نظام RoBERTa وقد ركزت الشركة على مواجهة الحسابات الوهمية عبر خوارزميات تعلم عميق تستند إلى أكثر من عشرين ألف متغير، تشمل خصائص المستخدم وشبكة أصدقائه، وهو ما مكنها في عام 2020 من حذف أكثر من خمسة مليارات حساب مزيف، مع تقدير أن واحداً من كل عشرين حساباً نشطاً هو حساب وهمي. كما أدخلت فيسبوك نظاماً جديداً لترتيب المشاركات المبلغ عنها بحيث تمنح الأولوية للمحتويات الأخطر والأكثر انتشاراً، بما يتيح استجابة أسرع للتقليل من الأضرار المحتملة (Giansiracusa, 2021, pp. 201–204).

في سبتمبر 2017 نشرت تويتر تدوينة كشفت فيها بعض تفاصيل جهودها بعد انتخابات 2016 للحد من نشاط البوتات على المنصة. في ذلك الوقت كانت أنظمة تويتر الآلية تلتقط حوالي ثلاثة ملايين حساب مشبوه أسبوعياً أي ضعف ما كانت تلتقطه قبل عام من الانتخابات.

كما تستخدم تويتر التعلم الآلي لاكتشاف محاولات تسجيل الدخول المشبوهة—وقد كانت تحظر حينها نحو نصف مليون محاولة يوميا—وذلك عبر مؤشرات تكشف إن كان تسجيل الدخول يتم بطريقة آلية أو مبرمجة. بالإضافة إلى ذلك، تعتمد تويتر على خوارزميات التجميع (Clustering) لاكتشاف مجموعات كبيرة من الحسابات التي أنشأها أو يتحكم بها كيان واحد، رغم أن تفاصيل المتغيرات المستخدمة في هذا التحليل لم تكشف. (Giansiracusa, 2021, p. 204)

7. الخاتمة

يتضح مما سبق أن الخوارزميات لم تعد مجرد أدوات تقنية، بل أصبحت فاعلا رئيسيا في المشهد الإعلامي؛ فقد أسهمت من جهة في تسريع إنتاج المحتوى وتسهيل تداوله على نطاق واسع، لكنها من جهة أخرى ساعدت في تضخيم ظاهرة التضليل الإعلامي عبر التحكم في تدفق المعلومات على المنصات الرقمية. وازدادت التحديات صعوبة مع تطور تقنيات التزييف العميق، حيث لم تعد الصور والفيديوهات أدلة يمكن الوثوق بها بشكل كامل، بل أصبحت قابلة للتلاعب بدرجة يصعب معها تمييز الحقيقي من المزيف، وهذا بدوره أثر على ثقة الجمهور في المحتوى الإعلامي بشكل عام.

في مواجهة هذه التحديات طورت بعض المنصات خوارزميات للحد من انتشار المعلومات الزائفة إضافة إلى ظهور أدوات تعتمد على تقنيات التعلم العميق للكشف عن المحتوى المزيف، غير أن معظمها مدفوع الأجر وهو ما يحد من قدرة الكثير من الصحفيين والمستخدمين من الاستفادة منها، وهو ما يستدعي البحث عن حلول مفتوحة المصدر، وتعاون أوسع بين المؤسسات الإعلامية والمنصات الرقمية، إلى جانب رفع مستوى وعي الأفراد بهذه المخاطر، من أجل الحفاظ على مصداقية المعلومات في الفضاء الرقمي.

8. قائمة المصادر والمراجع

- Ayush. (2023, June 23). 6 Best Deepfake Detection Tools in 2023. *AiBiscuit*. <https://aibiscuit.com/best-deepfake-detection-tools/>
- Giansiracusa, N. (2021). *How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More*. Apress.
- Kotenidis, E., & Veglis, A. (2021). Algorithmic Journalism—Current Applications and Future Perspectives. *Journalism and Media*, 2(2), 244–257.
- Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, 223, 103525.

- Presuel, R. C., & Sierra, J. M. M. (2019). Algorithms and the News: Social Media Platforms as News Publishers and Distributors. *Revista de Comunicación*, 18(2), 261–285.

- Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, Automation, and News. *Digital Journalism*, 7(8), 980–992.
<https://doi.org/10.1080/21670811.2019.1685395>

- خريش، خ. أ. ع. (2025). طبيعة استخدام القائم بالاتصال في وسائل الإعلام الليبية لأساليب كشف التزييف العميق بواسطة الذكاء الاصطناعي (دراسة وصفية تطبيقية في ظل نظرية دافع الحماية). *مجلة بحوث الاتصال*, 863-834.

- فرجاني، ع. (2024). التزييف العميق وتقنيات الخداع الرقمي —دراسة متعمقة في التحديات المهنية والتقنية. *المركز الديمقراطي العربي*, 8(29), 86-102.