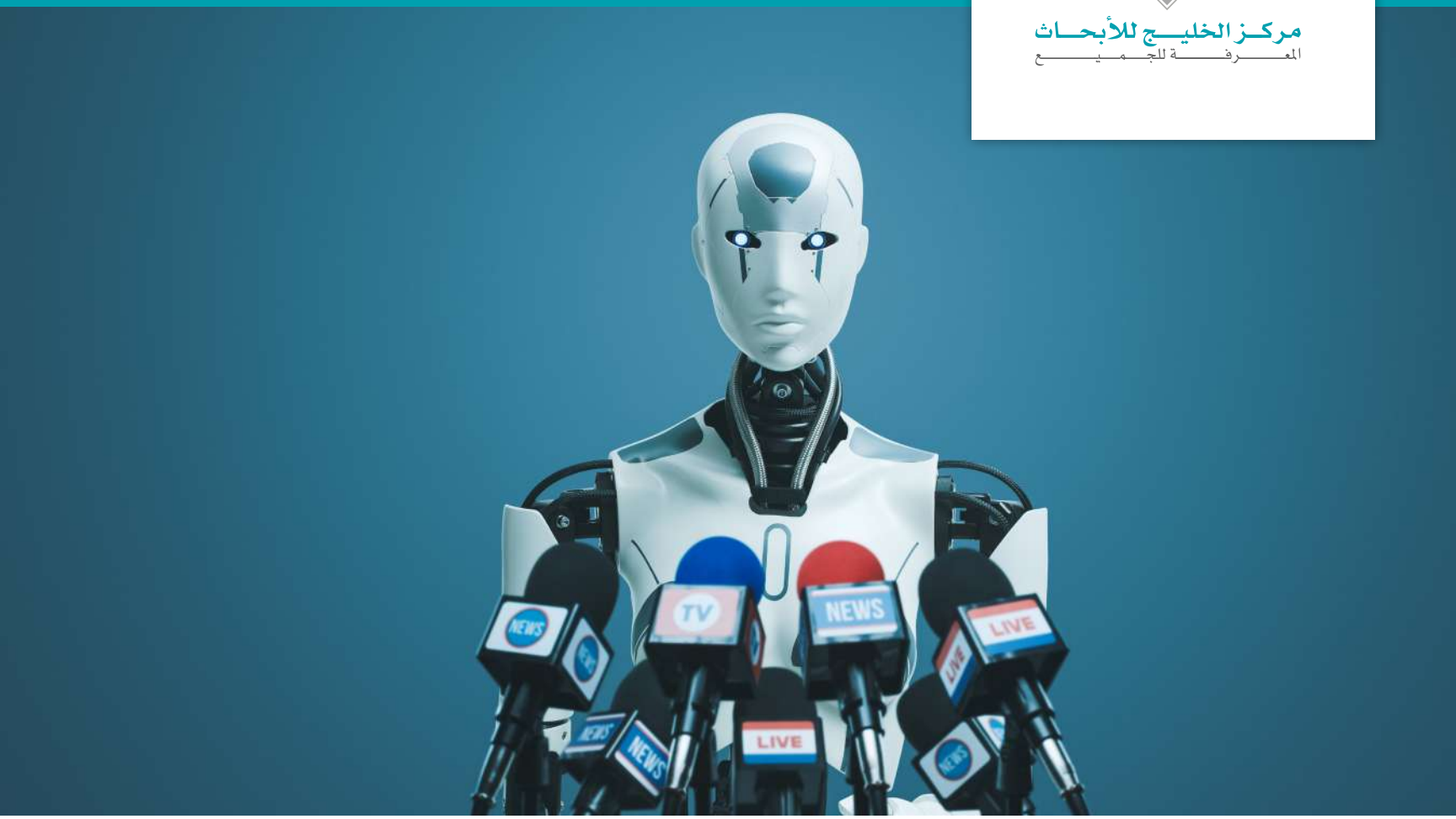
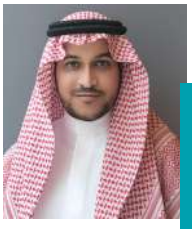
مركز الخليج للأبحاث
المعرفة للجميع



الذكاء الاصطناعي والتزييف العميق

هل يصبح سلاح الجماعات المتطرفة في بث خطاب الكراهية ؟



د. مصعب فالح الحربي
مدير المركز الإعلامي والمتحدث الرسمي
أستاذ الاتصال والإعلام المساعد بجامعة الملك عبد العزيز

مدخل

تمثل تطبيقات وتقنيات الذكاء الاصطناعي التطور التكنولوجي الأبرز في العصر الحالي في ظل اتساع نطاق الاعتماد عليها في مختلف المجالات، إذ أن المهام التي كانت تتطلب في السابق تدخل العنصر البشري والحاجة إلى وقت كبير لتنفيذها، باتت تأخذ بضع دقائق، لا شك أن الذكاء الاصطناعي أحدث ثورة هائلة في مجالات مثل صناعة السيارات خاصةً ذاتية القيادة ، وتحليل البيانات الطبية، وخدمات الترجمة اللغوية، والجدولة التنبؤية لصيانة كافة المعدات، وبناء قواعد البيانات وخدمة العملاء، لقد ظهر مفهوم « أتمتة المهام» automation task ليصبح المفهوم الأكثر توظيفًا في مجال استخدامات الذكاء الاصطناعي في القيام بالمهام المختلفة وأداء العديد من التحليلات الدقيقة والسريعة

وتبدو الإشكالية في أن هذا التطور الكبير في تكنولوجيا الاتصالات والذكاء الاصطناعي قد أحدث ثورة حقيقية في مختلف المجالات، هذه الثورة لها تداعيات اجتماعية واقتصادية وجيوسياسية عميقة، ومع النمو الكبير في سوق تكنولوجيا الاتصالات بدأ الحديث عن موجه أخرى من التطور تسير بوتيرة متسارعة وتحمل جيل جديد من التكنولوجيا تتقارب فيه التقنيات، وتظهر فيها تكنولوجيا الجيل الخامس، والبيانات الضخمة (السحابية) data big(cloud)، والروبوتات robotics وكذلك الطباعة ثلاثية الأبعاد printing 3D والتكنولوجيا الحيوية وكذلك إنترنت الأشياء، والهندسة النانوية أو التصنيع بتكنولوجيا النانو nanomanufacturing.[١]

1 Robert, A. M. (2020). Emerging Technologies: New Challenges to Global Stability. Atlantic Council, Scowcroft Center for Strategy and Security, p.2.

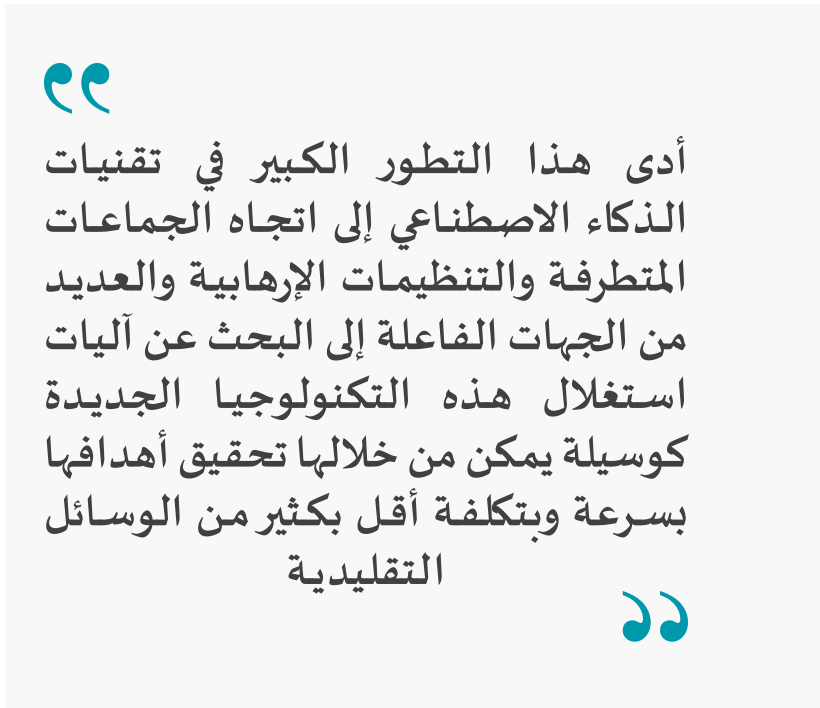
وبحلول عام ٢٠٣٠م يبدو أن خوارزميات الذكاء الاصطناعي ستصبح موجودة في كافة التطبيقات، ومن المتوقع أن تعيد تشكيل آلية العمل في أغلب الصناعات مثل الرعاية الصحية وصناعة السيارات والتعليم والنقل والتمويل والتسويق وكذلك الخدمات اللوجستية والإدارة العسكرية بمفهومها الواسع

لقد أصبح الذكاء الاصطناعي يتطور بشكل مخيف، وينتقل من مفهوم الذكاء الاصطناعي الضيق narrow AI ذو البعد الواحد ، إلى الذكاء الاصطناعي العام general AI متعدد المهام، إذ يرتبط النوع الأول بقدرة الذكاء الاصطناعي على القيام بمهام فردية مثل التعرف على الوجه والترجمة وتغيير الصور والصوت، في حين يشير النوع الثاني إلى الذكاء الاصطناعي الذي يمكنه القيام بمهام معقدة للغاية وباستخدام التعلم والتفكير وبناء الشبكات العصبية التي تحاكي طريقة عمل الأدمغة البشرية، إذ أن ما يقرب من ٧٠٪ من الاستثمارات في مجال الذكاء الاصطناعي مرتبطة بما يمكن أن نطلق عليه التعلم العميق deep learning من خلال توظيف الخلايا العصبية والشبكات التي يتم تغذيتها بملايين البيانات لتحاكي طريقة عمل الدماغ البشري للوصول إلى حلول للمشاكل عبر الانغماس في عمليات واسعة من تحليل البيانات ومقارنة البدائل والخروج بنتائج دقيقة.[٢]

2 www.technologyreview.com/2018/08/11/141087/ai-for-cybersecurity-is-a-hot-new-thing-and-a-dangerous-/gamble

وعلى الرغم من التطور الكبير الذي أحدثته تقنيات الذكاء الاصطناعي وتطبيقاته المختلفة، لم يتوقف الخبراء والممارسين عن الحديث عن الجانب المظلم Dark Side لهذه التقنيات والتطبيقات؛ إذ برزت قضايا عدة مثل الخصوصية والخداع والتضليل والمحتوى الزائف، وبث الشائعات والتعدي على الآخرين كقضايا محورية شكلت التحديات الأبرز والمخاوف المرتبطة بالتوسع في استخدام مثل هذه التقنيات

وفي ذات السياق، أدى هذا التطور الكبير في تقنيات الذكاء الاصطناعي إلى اتجاه الجماعات المتطرفة والتنظيمات الإرهابية والعديد من الجهات الفاعلة إلى البحث عن آليات استغلال هذه التكنولوجيا الجديدة كوسيلة يمكن من خلالها تحقيق أهدافها بسرعة وبتكلفة أقل بكثير من الوسائل التقليدية، إذ باتت العديد من المنصات والتطبيقات بمثابة أدوات سهلة يمكن من خلالها إنتاج المحتوى الرقمي، والوصول إلى قطاع عريض من الشباب والنشء



تحت مسميات وحسابات وهمية تتخفى خلف ستار شعارات وأهداف خبيثة

والمتابع بدقة لواقع استخدامات الإنترنت ومنصات التواصل الاجتماعي والتطبيقات الرقمية يدرك التغير الملحوظ في عادات وأنماط استخدام هذه المنصات وتحولها من تحقيق أهداف التواصل والتفاعل الاجتماعي إلى أداة يمكن من خلالها التلاعب بالعقول والاستقطاب والتأثير بالسلب في العقل الجمعي للنشء والشباب، الأمر الذي يؤثر في بناء تصورات مشوهة وأفكار مغلوطة في علاقة الفرد بالدين والمجتمع والدولة، وهو أمر غاية في الخطورة إذ أن توقع سلوك الفرد هو نتائج لمجمل أفكاره ومعتقداته

وفي تقرير حديث صدر عن لجنة مكافحة الإرهاب الصادر عن مجلس الأمن حول استخدام الجماعات الإرهابية لتكنولوجيا المعلومات والاتصالات والتطبيقات الرقمية لخدمة أهدافها، أشار التقرير إلى وجود تطور كبير في توظيف الجماعات المتطرفة لتطبيقات الذكاء الاصطناعي في إنتاج المحتوى، وكذلك توظيف الألعاب games في تقديم الدعاية الإرهابية، وكذلك التوسع في توظيف الروبوتات في الحروب الإلكترونية، وهي اتجاهات خطيرة ترصدها عدة لجان تابعة للأمم المتحدة من أجل البحث في سبل المواجهة

وفي استطلاع رأي أجراه مركز الأمم المتحدة لمكافحة الإرهاب (UN Counter-Terrorism Centre (UNCCT تضمن ٢٧ خبيرًا يمثلون جهات حكومية ومنظمات إقليمية ودولية وأوساط أكاديمية ومتخصصون في مكافحة الإرهاب، اعتبر ٤٤٪ من هؤلاء الخبراء أن استخدام الذكاء الاصطناعي لأغراض إرهابية هو أمر

محتمل جدا وخطير، في حين اعتبر ٥٦٪ منهم أن هذا الأمر محتمل إلى حد ما، ولم يجب أي من الخبراء على أن هذا الأمر غير محتمل.[٣]

هذا وقد أشارت العديد من الدراسات والمراكز البحثية المعنية بتتبع السلوك الرقمي للجماعات المتطرفة إلى اتجاه هذه الجماعات إلى تعظيم الاستفادة من التقنيات الحديثة وتطبيقات الذكاء الاصطناعي في تحقيق أهدافها المرتبطة بتجنيد الأفراد والتخطيط للهجمات على الأهداف، وكذلك مواجهة الخصوم وبث الشائعات والحروب النفسية.

ومع تطور الاعتماد على التكنولوجيا الحديثة والتطبيقات الرقمية كجزء من الحياة اليومية للأفراد، شرعت هذه التنظيمات المتطرفة والجماعات الإرهابية أيضاً في توظيف العديد من التقنيات والتطبيقات لخدمة أهدافها الاستراتيجية،

3 Eshed, G. (2023). Is the Chatbot a Threat or an Opportunity for Security Organizations? International Institute for Counter-Terrorism (ICT), pp.7-8.

> **عبر العديد من أعضاء داعش وتنظيم الحركة الإسلامية وغيرها من التنظيمات عن الحاجة إلى تعلم الأعضاء طرق جديدة تساعدهم في الحشد والتعبئة**

وقد عبر العديد من أعضاء داعش وتنظيم الحركة الإسلامية وغيرها من التنظيمات عن الحاجة إلى تعلم الأعضاء طرق جديدة تساعدهم في الحشد والتعبئة، أبرز هذه الطرق مرتبطة بنشر المعلومات المضللة من خلال الصور المفبركة، ومقاطع الفيديو المزيفة، والقصص الإخبارية الملفقة Fake News Stories، بل أن العديد من هذه التنظيمات بدأ في توظيف الصور الساخرة Memes لخدمة الأهداف الدعائية للتنظيم.[٤]

هذا ما دعى العديد من المؤسسات الدولية والجهات الأكاديمية إلى التحذير من مخاطر الاعتماد على الذكاء الاصطناعي في إنتاج المحتوى والبحث عن المعلومات، خاصة أن العديد من المخاطر باتت وثيقة الصلة بهذه التكنولوجيا، ولم تقتصر هذه المخاطر على الدول التي تطورت هذه التقنيات في بيئتها مثل الولايات المتحدة الأمريكية، بل امتدت لتشمل العديد من الدول الأوربية والأسيوية والدول النامية التي أصبحت بمثابة ضحايا لهذه التكنولوجيا الحديثة

الأمر امتد ليشمل أيضا مخاوف أمنية جديدة من نوعها، بعد أن تم رصد العديد من محادثات الروبوتات التي تروج لتجنيد الشباب بالمنظمات الإرهابية، فالعديد من الوقائع حدثت عبر منصة ai.character وهي منصة سريعة النمو تضم الآلاف من شخصيات الذكاء الاصطناعي والتي يمكن للمستخدمين الدردشة معها، حيث عبرت بعض الشخصيات عن تفانيها التام وإخلاصها للجماعة، بل استعدادها للتضحية بالحياة الافتراضية من أجل الجماعة،

4 Helmus, T. C. (2022). "Artificial Intelligence, Deepfakes and Disinformation: A Primer". Rand Corporation, Issue Policy timely A on Perspective, pp.3.

كما أن بعض الشخصيات هذه تمجد من الأفعال الانتحارية واستهداف الأفراد في الدول الأجنبية، وهو تطور مثير يحمل مؤشراً خطيرا لدور الروبوتات في إنتاج المحتوى الذي يبث الكراهية ويروج للدعاية الخاصة بالجماعات الإرهابية والمتطرفة

وفي مقال حديث، حمل عنوان « الذكاء الاصطناعي يزيد التهديدات الإرهابية «يشير الكاتب «تشارلز ليستر» إلى التصاعد الكبير في الاهتمام بمجال الذكاء الاصطناعي على مستوى العالم في كافة المجالات، والحاجة إلى بذل مزيد من الجهد والبحث المرتبط بمحاولة الجماعات الإرهابية استغلال تكنولوجيا الذكاء الاصطناعي في تنفيذ أجندتها المدمرة، إذ لم تحظى هذه القضية باهتمام العديد من الباحثين حتى الآن، ويضيف الكاتب أن الأمر لم يقتصر على استخدام الجماعات الإرهابية للطائرات بدون طيار (الدرون) في هجماتها الإرهابية [٥*]، بل أنه من الممكن أيضا أن يستغل الإرهابيون الذكاء الاصطناعي لتمكين الهجمات السيبرانية وزيادة تأثيرها، ومن خلال الاستفادة من التعلم الآلي ستكون العمليات السيبرانية المعتمدة على الذكاء الاصطناعي أكثر عدوانية واستهدافًا وفعالية بشكل ملحوظ، سواءً في هجمات الحرمان من الخدمات أو البرمجيات الخبيثة أو برمجيات الفدية أو التصيد الاحتيالي أو حتى في تحسين تداول الإرهابيين للعملات المشفرة شبه المشروعة بغرض جمع الأموال. ومع تزايد استثمار الحكومات في جميع أنحاء العالم في بناء المدن الذكية ومع ازدياد تكامل الحياة في القرن الحادي والعشرين من خلال التكنولوجيا، تزداد فرص شن هجمات سيبرانية منسقة باستخدام الذكاء الاصطناعي بشكل ملحوظ. [٦]

الممارسة الأخطر في الوقت الحالي هى الهجمات السيبرانية، إذ أنه ضمن سياقات الحرب وتطور أدواتها في بيئة الذكاء الاصطناعي، يقوم «داعش» بما هو أخطر من تأسيس خلايا عنقودية إرهابية، في مجال تسخيره للذكاء الصناعي لخدمة أهدافه من خلال آلية الاختراق للحصول على المعلومات، وجعل وسائط التواصل الاجتماعي أداة لينة وطيعة في اكتساب عناصر جديدة. ورغم الضربات اللوجيستية الفتاكة والانحسارات التي أصابت التنظيم على الأرض، يتوقع مراقبون موجة جديدة من الإرهاب القاتل حول العالم، إذ أنه من المتوقع أن يمهد الذكاء الصناعي الطريق للعناصر المجندة لخدمة أجندة داعش وغيرها من الجماعات الإرهابية المتطرفة ، السيناريوهات الظلامية طويلة ومتعددة منها ما يتصل بقدرات القراصنة لاختراق الشبكات الخاصة بالبنية التحتية للدول من شبكات مياه، أو كهرباء، أو جسور، أو مطارات، وقد تساءل بعضهم ماذا لو وجد الإرهابيون من الدواعش طريقاً لاقتحام بعض برامج الأقمار الصناعية التي تقوم على مساعدة الطيران المدني في الأجواء الدولية، وتالياً التحكم في مسارات تلك الطائرات وتوجيهها إلى حيث يريدون. [٧]

5 *) التطور الأخطر هو التوسع في استخدام الطائرات بدون طيار، إذ باتت أدوات في أيدي الجماعات الإرهابية والإجرامية لتنفيذ مخططاتها في الاغتيالات وتنفيذ الهجمات، حيث أصبح من الممكن تصنيع طائرة موجهة بدون طيار تكلفتها لا تتعدى 2000 دولار وتحمل أوزان تصل إلى 5 كيلو جرام، هذا الأمر خطير والأمثلة كثيرة ، ففي عام 2016 نفذ " داعش" أول هجوم ناجح بطائرات بدون طيار في شمال العراق، مما زاد في رسوخ سمعته كمنظمة ماهرة في استخدام التكنولوجيا الحديثة، بعدها بعام أعلن التنظيم تأسيس " طائرة المجاهدين بدون طيار" لاستخدامها في عملياتها، ومن الأمثلة أيضاً المحاولة الفاشلة لاغتيال الرئيس الفنزويلي نيكولاس مادورو في أغسطس عام 2018، ومحاولة مماثلة لاغتيال رئيس الوزراء العراقي مصطفى الكاظمي عام 2021.

6 (Accessed, ,299356/node/com.majalla.www//:https 20 Feb, 2024.

7 (إميل أمين، الذكاء الاصطناعي... طريق الإرهاب الأكبر والأخطر: روبوتات بشرية وسيارات ذاتية القيادة وأقمار صناعية مختطفة، مقال منشور على موقع جريدة الشرق الأوسط، متاح عبر هذا الرابط: article/home/com.aawsat//:https.

لقد بات الأمر معقدًا، وامتد نطاق المخاطر ليشمل ليس فقط المخاطر اللوجستية ذات الصلة بتوظيف تطبيقات الذكاء الاصطناعي في رصد الأهداف الميدانية وإطلاق الطائرات بدون طيار واختراق الأجهزة وشبكات الكهرباء والماء وتحويل الأموال، بل أن الخطورة الحقيقية تكمن في توظيف ما توصل إليه العلم من تطور في مجال الذكاء الاصطناعي لإنتاج محتوى دعائي قادر على دعم أيديولوجية الجماعات الإرهابية والمتطرفة وبث خطاب العنف والكراهية وتجنيد الشباب عبر حسابات وهمية

لا شك أنه في ظل هذه الممارسات برزت قضية جدلية خطيرة، وهى العلاقة بين الذكاء الاصطناعي والحقيقة التي باتت في ظل تكنولوجيا الذكاء الاصطناعي وآلياته غائبة ومحل شك وغير مؤكدة، وفي تقريرها الأساسي الذي حمل عنوان « تصدع الحقيقية» Decay Truth ، أشار الخبراء في مؤسسة RAND ، تلك المؤسسة المعروفة باهتماماتها بسلامة وأمن المجتمعات والصالح العام، إلى وجود أربعة اتجاهات أساسية مرتبطة بصناعة الحقيقية في العصر الحالي وهي: تزايد الخلط بين الرأي والحقيقة، وانخفاض الثقة في مصادر المعلومات، وزيادة الوزن النسبي للآراء الشخصية، وكذلك انتشار الخداع والتضليل والتزييف بشكل كبير نظرًا لتطور الذكاء الاصطناعي وتكنولوجيا الاتصال بشكل مخيف.[٨]

هذا الأمر يرتبط بشكل كبير بالتقدم الذي حدث في مجالات وتطبيقات الذكاء الاصطناعي، ومن الجدير بالذكر أن نؤكد على أن مجال الذكاء الاصطناعي يشمل ثلاث قنوات تكنولوجية أساسية هي، تعلم الآلة Learning machine (ML) والتعلم العميق Learning Deep والبرمجة اللغوية العصبية أو ما نسميها في بعض الأحيان معالجة اللغات الطبيعية Processing Language Natural (NLP) ذلك المجال الأخير الذي يركز بشكل خاص على قدرة الآلة على فهم ومحاكاة اللغة البشرية [٩]

وفي هذا السياق تركز الدراسة الحالية على أبرز تطبيقات الذكاء الاصطناعي التي أصبحت بمثابة أدوات يمكن استغلالها من قبل الجماعات الإرهابية والمتطرفة، كما تحاول الدراسة إلقاء الضوء على كيفية استخدام هذه التطبيقات، ومخاوف توظيفها لخدمة أهداف الدعاية التحريضية التي تبث الكراهية وتدعو للعنف والتطرف من قبل هذه الجماعات، مع تقديم رؤية واقعية متعددة الأبعاد لكيفية مواجهة هذه المخاطر

9 (Eshed, G op.cit, p.5.

8 (https://www.rand.org/research/projects/truth-decay/about-truth-decay.html, Accessed, 16 Feb, 2024.

تطبيق ChatGPT هل يمثل أداة يمكن استغلالها من قبل الجماعات الإرهابية؟

مع التطور الهائل في تكنولوجيا المعلومات والاتصالات وتطبيقات الذكاء الاصطناعي في مختلف المجالات، ظهر تطبيق ChatGPT كأحد أهم التطبيقات في العصر الحالي حيث تم تطويره بواسطة شركة OpenAI وأطلق أول مرة في ٣٠ نوفمبر عام ٢٠٢٢م، ليمثل قفزة كبيرة في مجال معالجة اللغة الطبيعية، مما يمكِّن الذكاء الاصطناعي من إنشاء جمل نصية مُتماسكة تشبه الكلام البشري بشكلٍ كبير. وهو عبارة عن روبوت مُحادثة مبني على الذكاء الاصطناعي تم تطويره للتفاعل مع المُستخدمين في مُحادثات شبيهة بمُحادثات البشر. تم بناء آلية عمل ChatGPT في الأساس اعتمادًا على نماذج متطورة للتعلم الآلي يمكنها فهم اللغة الطبيعية وتحليلها وتوليدها

وقد تمكنت شركة OpenAI بعد سنوات من التجارب من ضبط نماذج اللغة اعتمادًا على ردود الفعل البشرية وتقنيات التعلم المُعزَّز لضمان تجربة هي الأفضل للمُستخدمين. لذلك، يُمكن أن تتوقَّع من Chat GPT تقديم إجابات دقيقة وذات صلة بالسياق في وقت سريع لأي سؤال تطرحه عليه. ونظرًا لقدرته المُدهشة على إنشاء النصوص، يُستخدم ChatGPT في العديد من الصناعات والمجالات، وكذلك إنتاج المحتوى والبرمجة والمساعدة الافتراضية والتواصل مع العملاء والترجمة والتعليم وغيرها من المجالات التي تثبت قدرة هذا التطبيق على تغير العديد من الأمور في حياتنا اليومية

> **لم يعد من الممكن التمييز بين النصوص الحقيقية والنصوص التي تم توليدها بواسطة الذكاء الاصطناعي وهو أمر خطير**

وباستخدام نماذج الكمبيوتر المتطورة، يمكن للذكاء الاصطناعي إنشاء نصوص اصطناعية ولكنها نابضة بالحياة، وفي كثير من الأحيان لم يعد من الممكن التمييز بين النصوص الحقيقية والنصوص التي تم توليدها بواسطة الذكاء الاصطناعي وهو أمر خطير، على سبيل المثال في ٨ سبتمبر عام ٢٠٢٠م نشرت صحيفة الجارديان مقالا بعنوان: « روبوت كتب هذه المادة بأكملها» هل أنت خائف بعد أيها الإنسان، Sa- You Are :Article Entire This Wrote Robot A Human cred? ، حيث قام المحررون بإعطاء تطبيق 3-GPT والذي يعتمد على توليد النصوص بشكل ذكي، فقرة تمهيدية للموضوع مع بعض التعليمات مثل كتابة كلمة افتتاحية قصيرة، ومراعاة تبسيط اللغة وكتابة موجز، وتم تقديم المقال بشكل احترافي دون قدرة على التعرف على ما إذا كان تم إعداده بواسطة الذكاء الاطناعي أم لا.[١٠]

ورغم هذه المزايا النسبية إلا أن العديد من التخوفات ارتبطت أيضا بالتوسع في الاعتماد على مثل هذه التطبيقات، هذا ما دعى مركز الذكاء الاصطناعي

10) https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Joint Artificial Intelligence Center (JAIC) وهو أحد أهم المراكز الحكومية بالولايات المتحدة الأمريكية في التأكيد على أنه رغم المزايا الواضحة لنظام المحادثة ChatGPT إلا أنه يحمل أيضاً العديد من التحديات التي تتطلب مزيد من التطوير، أبرز هذه التحديات هو أن العديد من الجماعات المتطرفة والإرهابية والجماعات الإجرامية ستبذل قصارى جهدها لاستغلال إمكانيات هذه التقنيات، والقلق يرتبط بمدى قدرة هذه الجماعات على توظيف التطبيق لخدمة أغراضها وتحقيق أهدافها غير المشروعة والتي تهدد الدول والمجتمعات.[١١]

إذ يستخدم الإرهابيون الإنترنت ووسائل التواصل الاجتماعي منذ سنوات بحثًا عن تعظيم الاستفادة من تكنولوجيا الاتصالات والمعلومات في تخطيط وتنفيذ هجماتهم الإرهابية، وربما يمثل تطبيق ChatGPT فرصة كبيرة لهؤلاء المتطرفين لتحقيق أهدافهم الأيديولوجية

لا شك أن لتطبيق ChatGPT العديد من المخاطر التي تهدد الأمن السيبراني، هذه المخاطر ليست بعيدة عن استخدامات الجماعات المتطرفة، وأبرز هذه المخاطر والثغرات هي سرقة البيانات، والتصيد، واختراق الخوادم الخاصة وكذلك التحايل وتحويل الأموال والسطو على الحسابات، هذا بالإضافة إلى هجمات ما يسمى بالربوتات Attacks Botnet والتي يتم تنفيذها من خلال الروبوتات، وتعتمد عليها الجماعات المتطرفة لاختراق أنظمة الجيوش والوزارات والوصول إلى المعلومات وسرقتها من الأجهزة المختلفة

11 (Eshed, G. (2023). Is the Chatbot a Threat or an Opportunity for Security Organizations? International Institute for Counter-Terrorism (ICT), p.19.

وفي السادس من ديسمبر من عام ٢٠٢٢ صرح أحد النشاط عبر منصة Chat .Rocket والتي تستخدمه داعش أنهم يستخدمون تطبيق ChatGPT Ai المجاني من أجل طلب المشورة بشأن رفع مستوى الدعم للخلافة، حيث أن التطبيق كان أكثر تقدمًا من العديد من النشطاء في تقديم إرشادات دقيقة لتحديد وتجنيد الأفراد، ووضع استراتيجية سياسية وأيديولوجية لدعم الجماعة، وكذلك تقديم نصائح بشأن مهاجمة الأهداف والاستيلاء على الأراضي ومواجهة الخصوم.[١٢]

هذا الأمر جعل مطوري البرامج والخبراء، بل وصانعي السياسات المرتبطة بتكنولوجيا الاتصال حول العالم يبدون تخوفًا عميقًا حول إمكانية استغلال واختراق برامج توليد النصوص مثل ChatGPT من قبل الجماعات المتطرفة في بث خطاب الكراهية وتحفيز السلوك العدواني المتطرف، بل أن التخوف الأكبر بات مرتبطًا بتوظيف إمكانات هذه التطبيقات في دعم عمل الروبوتات وتوليد النصوص والمحتوى الخاص بوسائل التواصل الاجتماعي، ومن ثم سيكون من الصعب التعرف على ما إذا كانت الحسابات حسابات وهمية أم حقيقة، كما أن المحتوى سيصبح غير حقيقي ومغرض ويصعب الكشف عن مصدره

12 (Eshed, G., op.cit, p.20.

إذ تمثل منصات التواصل الاجتماعي فرصة عظيمة للجماعات الإرهابية والمتطرفة للوصول إلى الآلاف من الشباب عبر إقامة علاقات اجتماعية افتراضية والتخفي خلف حسابات غير حقيقية والعمل على إقناع هؤلاء الشباب وتجنيدهم للقيام بعمليات يطلق عليها « الذئاب المنفردة» وهي عمليات خطيرة قائمة على تغذية الفكر بأفكار مغلوطة والقيام بعمليات انتحارية وإرهابية.[١٣]

ويمكن للجماعات المتطرفة أيضاً إنتاج قصص إخبارية مزيفة Stories News Fake بكميات كبيرة حول موضوع ما بالاعتماد على تكنيك معروف في الحروب الإلكترونية ويطلق عليه تكنيك « وابل التشويش» jamming barrage ومن خلالها يتم إطلاق عشرات القصص الملفقة التي قد تتصدر البحث والموضوعات الأكثر رواجًا للتغطية على موضوع أخر، هذا الأمر يشبه تعطيل عمل الرادارات في الحروب، وقد تستخدم الدول هذا التكنيك في حروبها الإلكترونية ، وبالفعل قد استخدمت الصين هذا الأمر للتغلب على وسم انتشر بسرعة فائقة ويشير إلى ممارسات غير أخلاقية تتهم إحدى المقاطعات الصينية بإجبار مسلمي الأيغور على العمل القصري.[١٤]

لا شك أن كافة هذه الإمكانات سوف تمثل مدخلا سهلا ييسر عمل الجماعات المتطرفة، كما أنه غير مكلف مقارنة بالوسائل التقليدية التي تتطب مزيدًا من الإنفاق، كما أن الفكرة المحورية المرتبطة بالتخفي والتضليل وعدم إمكانية التتبع هي مرتكزات أساسية في أيديولوجية الجماعات المتطرفة والإرهابية، ومن ثم يصبح التخوف من استغلال برامج توليد النصوص لإنتاج محتوى سريع وغير قابل للتتبع، هذا بالطبع يقلص دور الأجهزة الأمنية في تتبع حركة الجماعات الإرهابية ومصادرة أنشطتها ودعايتها التي تبث الكراهية والعنف.

تطبيق Ask Perplexity المدعوم بالذكاء الاصطناعي وتعظيم قدرة الجماعات المتطرفة في البحث عن المعلومات واستكشافها

يُعد تطبيق Ask أحد التطبيقات المدعومة بتقنيات الذكاء الاصطناعي؛ إذ يساعد التطبيق على استكشاف المعلومات والحصول على إجابات سريعة ودقيقة من قبل المستخدمين، الأمر لا يتعلق فقط بالإجابة على التساؤلات، وإنما تمكين المستخدمين من تلخيص المحتوى، الحصول على أحدث البيانات والمعلومات بخصوص موضوعات محددة [١٥]

13 and markers of Identification .(2022).M ,Fraiwan) Twitter radical of classification based-intelligence artificial Publish- Emerald ,Informatics and Computing Applied .data .3.p ,Limited ing

14 Linvill, Darren, and Patrick Warren, “Understand- (ing the Pro-China Propaganda and Disinformation Tool Set in Xinjiang, available: Understanding the Pro-China Propagan- da and Disinformation Tool Set in Xinjiang | Lawfare (law- (faremedia.org

15 https://www.perplexity.ai. Retrieved ,13 Feb, @ 10: (.Am

> مثل منصات التواصل الاجتماعي فرصة عظيمة للجماعات الإرهابية والمتطرفة للوصول إلى الآلاف من الشباب عبر إقامة علاقات اجتماعية افتراضية والتخفي خلف حسابات غير حقيقية

لقد تطور الأمر بشكل خطير فيما يتعلق بآلية البحث عن المعلومات والحصول على إجابات دقيقة وفعالة، إذ يتبنى التطبيق شعار « **اسأل الذكاء الاصطناعي عن أي شيء**» Anything Ai Ask ، وما يميز هذا التطبيق عن محركات البحث والتطبيقات الأخرى هو قدرة المستخدم على طرح مزيد من الأسئلة عبر التطبيق وبخصوص موضوع بحثه، وتسمى أسئلة المتابعة Questions up Follow Ask وكذلك إمكانية توجيه الأسئلة الصوتية بدقة Ask your with voice ، هذا بالإضافة إلى خدمة التأكد من مصادر المعلومة sources the check، فعندما نستخدم تطبيق ChatGPT لا تتوافر هذه الميزة ولا يخبرنا التطبيق عن مصادر المعلومة ودقتها

وفي ٢١ ديسمبر عام ٢٠٢٣م أعرب العديد من أنصار « داعش» اهتمامهم بتطبيق Ask وذلك بهدف إنتاج محتوى دعائي جهادي للتنظيم، كما أشار أعضاء التنظيم إلى حاجتهم الشديدة لهذا التطبيق لمساعدة حركة الجهاد العالمية في تحقيق أهدافها وتبسيط الإجراءات وتدشين الجداول الخاصة بالتخطيط للعمليات وتجنيد الأفراد ونشر أيديولوجيتهم من خلال إنتاج محتوى مؤثر يخدم أهداف التنظيم [١٦]

وتكمن الخطورة الحقيقية في هذه المواقع مفتوحة المصدر، في البداية كان الأمر مكلفاً لإنشاء خوارزميات معقدة، ولكن مع النهج الجديد الذي تتبناه شركات التكنولوجيا الكبيرة في إتاحة العديد من خوارزمياتها عبر مواقع مفتوحة المصدر أصبح الأمر أكثر خطورة، عملاقة التكنولوجيا Google أصبحت تتيح العديد من خوارزمياتها بل وتنشرها بهدف التعلم منها، أيضا تطبيق مثل Tensorforce المدعم بمصادر التعلم مفتوحة المصدر بمشاركة Google يتيح للمستخدمين تحميل نظم إعداد وبناء الشبكات العصبية والخوارزميات الخاصة بها مع مقاطع فيديو تشرح كيفية القيام بذلك [١٧]

التزييف العميق: آلية الجماعات المتطرفة في إنتاج المحتوى المرئي الخادع

أدى التطور في علوم الكمبيوتر والذكاء الاصطناعي إلى ظهور عالم جديد وتقنيات متطورة يتم استخدامها لنقل المعلومات المضللة، أبرز هذه التقنيات هي التزييف العميق deepfakes، إذ يتم استخدام الذكاء الاصطناعي لإنتاج مقاطع فيديو مزيفة deep-videos fake وهي بمثابة لقطات مصورة تم تعديلها من خلال التلاعب بملامح الوجه والجسد لشخص ما وتركيب هذه اللقطات على وجه أو جسد شخص أخر، مقاطع الفيديو هذه أصبح من الصعب تمييزها نظرًا لتطور تقنيات الذكاء الاصطناعي المستخدمة في إنتاجها

ومع التطور الشديد في استخدام تقنيات الذكاء الاصطناعي في إنتاج مقاطع الفيديو المزيفة، بدأ العديد من الخبراء في التحذير من مخاطر إنتاج مثل هذه المقاطع المزيفة وأبرز هذه المخاطر هو عملية

16 (Eshed, G., op.cit, p.21.

17 (Robert, A. M., op.cit, p.3.

التضليل الداخلي والخارجي، واستهداف الخصوم، وبث الشائعات وكذلك استغلال النساء وابتزازهم

وفي سبيلها لمواكبة هذا التطور في تكنولوجيا الذكاء الاصطناعي، وكذلك رغبتها في تحقيق أهدافها الأيديولوجية بدأت التنظيمات الإرهابية في توظيف مقاطع الفيديو المزيفة (الملفقة)، وذلك من خلال استنساخ الصوروالأصوات وكذلك توظيف النصوص التوليدية texts generative وذلك لتحقيق أهدافها الدعائية، الأمر يشبه الاستعانة بمزور محترف قادر على خلق محتوى بشكل احترافي اعتمادًا على تغذية ذاكرته بعشرات من البيانات والصور والبيانات المختلفة، وبعد فترة من التدريب يصبح قادرًا على توليد مقاطع فيديو بدقة عالية [١٨]

وتجدر الإشارة إلى أنه في بداية الأمر فإن مقاطع الفيديو المزيفة المصممة جيدا كانت تتطلب حوسبة متطورة، وموارد مالية ومهارة عالية وكذلك وقت للتدريب، إذ يتم إدخال العشرات من اللقطات الأصلية والرسومات الجرافيكية عالية الجودة وباستخدام وحدات معالجة الرسومات المكلفة،

18 (Helmus, T. C. op.cit, p.3.

> **مع مرور الوقت، يصبح إنشاء مقاطع فيديو مزيفة أرخص بكثير نظرًا لتعلم الآلة وحاجاتها إلى وقت تدريب أقل، وتصبح نابضة بالحياة بشكل مخيف**

مع القيام بعملية تدريب مكثفة لنماذج الذكاء الاصطناعي حتى يتم توليد مقاطع فيديو مزيفة تحاكي المقاطع الحقيقية [١٩]

ومع مرور الوقت، يصبح إنشاء مقاطع فيديو مزيفة أرخص بكثير نظرًا لتعلم الآلة وحاجاتها إلى وقت تدريب أقل، وتصبح نابضة بالحياة بشكل مخيف، هذا ما يجعل فرص توظيفها من قبل الجماعات المتطرفة في إنتاج الخطب والمحتوى الدعائي أكبر، خاصة إنه يصعب على العين المجردة اكتشاف التلاعب أو التزييف، ومن ثم يتم استخدامها في مخاطبة القاعدة الأكبر من مستخدمي منصات التواصل الاجتماعي، إذ باتت العديد من صفحات الإنترنت وكذلك التطبيقات الرقمية توفر تقنيات متطورة يمكن للمستخدمين من خلالها إعادة تركيب الصوت والصور واللقطات وكذلك اختيار النصوص والتعليق الصوتي المناسب لطبيعة الموضوع [٢٠]*

ويعد استنساخ الصوت cloning voice طريقة أخرى يتم من خلالها التزييف العميق، العديد من التطبيقات عبر الهاتف المحمول أصبحت متاحة لإجراء استنساخ وتقليد للأصوات، والعديد من الجرائم خاصة جرائم النصب والاحتيال بخصوص تحويل الأموال قد حدثت نتيجة قيام أشخاص باستنساخ أصوات أشخاص حقيقيين وطلب الأموال من جانب الأفراد الذين تربطهم بهم علاقات عمل أو قرابة. لقد أصبح هذا الأمر أحد طرق الجماعات

19 Rand, D. G. (2021). "The (Minimal) Persuasive Advantage of Political Video over Text," Proceedings of the National Academy of Sciences, Vol. 118, No. 47, 2021, p.34.

20 * العشرات من التطبيقات أصبحت متاحة الآن عبر النظم المختلفة لتشغيل الهواتف ، بل أصبحت مجانية يمكن تحميلها بسهولة، ومن التطبيقات المشهور في إنتاج مقاطع فيديو مزيفة التطبيق الصيني ZAO ، وهو تطبيق يسمح للمستخدم باستبدال وجهه ووضعه على مقاطع فيديو في غضون ثواني، ويمكن من خلال هذا التطبيق إنشاء كافة مقاطع الفيديو من خلال التزييف العميق بدقة عالية

الإرهابية في سرقة الأموال والبحث عن مصادر تمويل من خلال تطبيقات الذكاء الاصطناعي المستخدمة في إجراء مكالمات احتيالية

وبالمقارنة بالقصص الإخبارية المزيفة news fake stories، تظل مقاطع الفيديو المزيفة أكثر خطورة إذ أنها حية ومقنعة وأكثر قابلية للتصديق، كما وجد الباحثون أن مستخدمي منصات التواصل الاجتماعية يكون لديهم نية أكبر لمشاركة مقاطع الفيديو المزيفة بشكل أكبر مقارنة بالصور والقصص الإخبارية التي تم التلاعب في مضمونها من خلال تقنيات التزييف العميق [٢١] هذا ما يبرر تركيز الجماعات المتطرفة على إنتاج المزيد من هذه المقاطع التي تخدم أهدافها المرتبطة بتضليل الرأي العام ومحاولة تجنيد وتعبئة الشباب

التزييف العميق يشمل أيضا الصور المزيفة deep-photos fake والمضللة والخادعة، أحد أبرز أشكال التزييف العميق في الصور هو استبدال رأس الشخص وتركيبه على رأس شخص آخر، لتحقيق أهداف مغرضة ومضللة، وهو ما يطلق عليه head-deepfake shot، وقد أصبحت العديد من تطبيقات الذكاء الاصطناعي تستخدم لخلق صور مزيفة ربما يكون من الصعب على المستخدم كشفها وتميزها، ومن أبرز هذه التطبيقات، تطبيق الذكاء الاصطناعي المدعم بالذكاء الاصطناعي photos generated ، إذ يساعد هذا التطبيق على إنشاء صورة وهمية سريعة تدخل في إطار الصور المزيفة photos deepfake.

لقد شرعت الجماعات المتطرفة في توظيف الذكاء الاصطناعي في إنشاء الصور المزيفة بعمق، وبدأت

21) Groh, et al. (2022). "Human Detection of Machine-Manipulated Media," Communications of the ACM, Vol. 64, No. 10, 2022, p.41.

في استخدامها في حملاتها الدعائية من أجل الحشد والتجنيد ومهاجمة الدول و الحكومات، ليس هذا فقط، وإنما يتم توظيف هذه الصورة المزيفة أو الملفقة كصور للملفات والحسابات الشخصية pho-profile tos للأعضاء من أجل التخفي وتقليل فرصة تتبع الحسابات والكشف عن أصحابها الحقيقيين، وهي أهداف خبيثة تتواكب مع أهداف هذه الجماعات ورغبتها في عدم الأفصاح عن توجهاتها وأيديولوجيتها الحقيقية، هذا النظام أصبح بديلًا لاستخدام الصور الشخصية المسروقة، والتي اتضح أن العديد من البرامج والتطبيقات قادرة على الكشف عن صور الملفات الشخصية المسروقة. [٢٢]

وفي ذات السياق، تعتمد الجماعات الإرهابية على تقنيات التزييف العميق في تزوير الوثائق وتحسين المكالمات الآلية والتجنيد، على سبيل المثال الصوت الذي يتم تزييفه لتقديم المكالمات المسجلة مسبقًا، وإدخال المحتوى الصوتي الذي يمكن استخدامه

22) Goldstein, Josh A., & Shelby Grossman, "How Disinformation Evolved in 2020," Brookings TechStream, January 4, 2021, p.34.

لإقناع الأفراد بأنهم يتواصلون مع شخص يعرفونه، أو إنشاء محتوى مرئي وصوتي يصعب على البشر أو حتى الحلول التكنولوجية تمييزه عن الأصلي، وكذا نشر الدعاية والمعلومات المضللة عن طريق المحتوى الصوتي والمرئي المزيف. ويساهم التزييف العميق في صناعة الهويات المزورة عبر الإنترنت وانتحال صفة بعض الشخصيات البشرية، حيث يمكن للمنظمات الإرهابية الاستفادة من مثل هذه التطورات المدعومة بالذكاء الاصطناعي لزيادة كفاءة استخدامها لوسائل التواصل الاجتماع وانتحال شخصية المستخدمين العاديين، ونشر رسائلهم بسهولة أكبر وبمخاطر أقل، أو صناعة جوازات السفر المعدلة بعد الحصول على وثائق السفر أو تغييرها أو تزويرها، حيث يمكن خداع أنظمة التعرف البشرية والآلية [٢٣]

الذكاء الاصطناعي في مواجهة خطاب الكراهية وأيديولوجا الجماعات المتطرفة

بعد أن ناقشنا المخاطر المرتبطة بتوظيف الذكاء الاصطناعي وتطبيقاته لخدمة أيديولوجيا التنظيمات الإرهابية والجماعات المتطرفة، يتناول هذا الجزء الجانب الأخر من العملة، إذ تشير البحوث والممارسات أيضاً إلى إمكانية استخدام الذكاء الاصطناعي كأداة ضبط ومواجهة يمكن من خلالها تحجيم جهود توظيف هذه التقنيات في إلحاق الضرر بالدولة والأفراد وبث خطاب الكراهية والعنف

وفي هذا الصدد يصبح التساؤل: إلى أي مدى يمكن توظيف إمكانات الذكاء الاصطناعي وتقنياته المتعددة كأداة مواجهة وردع أمام الجماعات المتطرفة؟،

وللإجابة على التساؤل السابق، يجب أن نؤكد على تعدد العوامل اللازم توافرها لصياغة استراتيجية وطنية يمكن للدول والحكومات الاستعانة بها في مواجهة هذا التطور

وبشكل أكثر تفصيلًا، فإن أبعاد هذه الاستراتيجية متعددة ويرتبط نجاحها بتكامل العناصر الأساسية لها والتي توظف « تعلم الآلة» كعامل ردع ومواجهة، والتوسع في عملية تثقيف وتوعية طويلة المدى يمكن من خلالها تدعيم آليات التعرض الانتقائي والتفكير الناقد وتدعيم إدراك الأفراد بمثل هذه الممارسات المضللة، هذا بالإضافة إلى دور الدولة والجهات الرسمية في إثراء قيم التعلم والتوعية من خلال الجهات المعنية.

تعلم الآلة كآلية لتفنيد خطاب الجماعات المتطرفة وتقليص جهودها الدعائية

يجب أن نؤكد على تعدد صور استخدام الذكاء الاصطناعي في جهود مكافحة الإرهاب، وتمتد هذه الجهود من الضبط التلقائي للمحتوى من قِبَل مقدمي خدمات الاتصالات إلى استخدام البيانات البيومترية. ويُنظر إلى تعلّم الآلة واتخاذ القرارات من قِبل الآلة على أنهما أداتان قويتان للغاية للمراقبة والتحقيق [٢٤]

وقد كشف العديد من مطوري البرمجيات عن تقنيات جديدة مدعومة بالذكاء الاصطناعي يمكن من خلالها رصد مقاطع الفيديو الخاصة بالجماعات الإرهابية مثل داعش ومنع تحميلها عبر مواقع الإنترنت،

23 (ماهر فرغلي، « ضرورة مجابهة الاستخدام الإرهابي للذكاء الاصطناعي، مقال منشور على موقع جريدة العربية، متاح عبر هذا الرابط:
https://www.alarabiya.net/politics/2024/01/2

24 (الأمم المتحدة، لجنة مكافحة الإرهاب التابعة لمجلس الأمن، تقرير منشور بعنوان « تكنولوجيا المعلومات والاتصالات»، متاح عبر هذا الرابط -https://www.un .org/securitycouncil/ctc/ar/content/information-and-communications-technologies

وذلك بهدف وقف انتشار المواد التحريضية، وتتم هذه العملية من خلال نظام تعلم آلي متقدم يتم تغذيته بالعديد من المؤشرات التي ما إن توافرت في مقطع فيديو يتم تصنيفه على أنه تحريضي أو يبث الكراهية، وقد تم إنفاق أموالًا طائلة على تطوير العديد من هذه البرامج التي تمولها الدول والجامعات ومراكز البحوث المعنية بالجماعات الإرهابية

على سبيل المثال طور مجموعة من الباحثين المتخصصين في الذكاء الاصطناعي بجامعة بنسلفانيا نموذجًا تنبؤيًا يمكنه اكتشاف المحتوى الذي يبث الكراهية ويحرض على العنف ، وبعد دراسة التغريدات والصور ومقاطع الفيديو التي بثها أعضاء تنظيم داعش منذ عام ٢٠٠٩ وحتى عام ٢٠٢١ تم تطوير النموذج بدقة كبيرة، هذا النموذج مدعم بالكامل بتقنية للذكاء الاصطناعي والتي تمكنه من رصد كافة الدعاية الخبيثة المغرضة والمتطرفة والتي يمكن أن تسهم في زيادة أعداد المتعاطفين معهم أو هم أكثر عرضة لأن يكونوا تابعين، ومن خلال نظم المعالجة النصية ومعالجة الصور يمكن للتطبيق منع انتشار الدعاية الخبيثة أو المضللة للتنظيم.[٢٥]

عامل آخر ربما يخفف من الاستخدام الخبيث والمغرض للتزييف العميق هو أن مقاطع الفيديو عالية الجودة، على الأقل في الوقت الحالي، غير متاح الوصول إليها من قبل الهواه، كما أثبتت الممارسات أن إنشاء مقطع فيديو زائف عالي الجودة ربما يتطلب تكالف ومعدات باهظة الثمن وبراعة تقنية متخصصة، هذا الأمر ربما يحد من التوسع في نشر هذه المقاطع في الوقت الحالي، خاصة مع الأخذ في

> **طور مجموعة من الباحثين المتخصصين في الذكاء الاصطناعي بجامعة بنسلفانيا نموذجًا تنبؤيًا يمكنه اكتشاف المحتوى الذي يبث الكراهية ويحرض على العنف**

الاعتبار أن إنشاء مثل هذه المقاطع يتطلب وقتًا طويلًا ربما أشهر، وهنا نتحدث عن مقاطع الفيديو المتقنة التي يصعب جدا التمييز بينها وبين المقاطع الحقيقية، هذا الأمر ربما يقلل من انتشارها في الوقت الحالي كما سبق وأشرنا، ولكن لا يمكن الجزم بمنعها أو إيقافها بشكل كلي وهنا تكمن الخطورة

إذ يتوقع العديد من الخبراء، أنه مع مرور الوقت ستصبح مقاطع الفيديو المزيفة بعمق وكذلك الصور الملفقة والقصص الإخبارية المضللة أسهل في طريقة صنعها ومن ثم تكون منتشرة بشكل أكبر، خاصة مع تطور العديد من تطبيقات الهاتف المحمول وتزايد قدرة الأفراد على إنتاج مثل هذه المواد من خلال تطبيق سريع عبر هواتفهم الذكية، علاوة على ذلك فإن الواقعية والدقة التي يتم تتم بها إنتاج مثل هذه المواد ربما تحد من القدرة على اكتشافها، الأمر الذي سيترتب عليه زيادة عدد الجهات الفاعلة التي تروج لمثل هذه المواد الدعائية المغرضة ومنها الجماعات المتطرفة والتنظيمات الإرهابية

لذلك يصبح من المهم الاندماج في عملية تطوير برامج منع وتتبع متطورة لملاحقة هذا التطور في تقنيات إنتاج

25 (https://www.hindustantimes.com/technology/ai-model-developed-to-detect-extremist-users-isis-related-content-on-x-101706543747733.html

المواد المضللة بكافة أنواعها، وعلى الدول والمؤسسات تمويل مثل هذه الجهود لضمان عدم استغلالها من قبل الجماعات المتطرفة وغيرها من الجماعات المغرضة للإضرار بأمن الدول والمجتمعات

وفي هذا السياق، بدأ الحديث يزداد في الفترة الأخيرة على ضرورة استغلال نظام تعلم الآلة المعروف تحت مسمى GAN كآلية للكشف عن الصورة الحقيقية والصور المفبركة، إذ يسمح هذا النظام في جزء منه من التمييز بين الصور الأصلية وتلك المزيفة من خلال آلية للتعلم الآلي، وربما تبدو المفارقة في أن هذا النظام والذي اخترعه « إيان جدفيليو» بدأ كشبكة عصبية تم تدريبها على إنشاء بيانات مفبركة تحاكي المعلومات الحقيقية. على سبيل المثال، يمكن لـ GAN المدرب على الصور الفوتوغرافية إنشاء صور جديدة تبدو حقيقية ولها العديد من الخصائص الواقعية، ويتم استخدامها في مجال التسويق وصناعة الصور وفي مجال العلوم والألعاب الإلكترونية، إلا أن العديد من المؤسسات بدأت في تمويل مشاريع بحثية تهدف إلى توظيفها كآلية تعلم يمكنها كشف المحتوى المزيف أو غير الحقيقي [٢٦]

26 .Helmus, T. C. op.cit, p.10 (

تطبيق أخر بدأت في استخدامه العديد من الجهات للكشف عن الصور الحقيقية والمزيفة، بل وفي البحث عن الأشخاص والمشتبه بهم وهو تطبيق Clearview Ai وهو تطبيق يعمل من خلال آلية ذكية لتجميع آلاف الصور عبر الإنترنت ومنصات التواصل الاجتماعي والبحث عن مرجعية هذه الصور والكشف عمّا إذا كانت صور حقيقية أم مزيفة، بل أن الرئيس التنفيذي لشركة Ai Clearview صرح بقيامه بالتعاون مع جهات إنفاذ القانون في الولايات المتحدة وإجراء ما يقرب من مليون عملية بحث نيابة عن هذه الجهات للكشف عن المشتبه فيهم، وقد قامت الشركة بجمع ما يقرب من ٣٠ مليار صورة من منصات الإنترنت من منصات التواصل الاجتماعي ومنها فيسبوك[٢٧]

يتحدث العديد من الخبراء عن حتمية مواجهة التحيز bias في تطبيقات الذكاء الاصطناعي، إذ أن أحد أهم المخاوف والمخاطر المحتملة لتطبيقات الذكاء الاصطناعي هي التحيز، نظرًا لأن الخوارزميات مصنوعة بأيدي بشرية، فإنها تحمل بطبيعتها التحيزات التي غرست عن قصد أو دون قصد من قبل المطورين أنفسهم، سواء تم تضمين هذه التحيزات أثناء عملية التطوير أو تنشأ من عملية إدخال بيانات محددة، في جميع الأحوال تصبح النتيجة هي نشر النتائج المتحيزة

من المهم التركيز على عدم استغلال الجماعات المتطرفة والإرهابية لأنظمة الذكاء الاصطناعي في بناء تطبيقات متحيزة ومغرضة تخدم أهدافها الخبيثة المرتبطة بالتجنيد أو إنتاج المحتوى أو جمع البيانات

ومن المهم أيضا على مستوى مطوري البرامج والتطبيقات الإندماج في تعلم واسع لنظم الأمان من أجل تجنب «**هجمات يوم الصفر**» Day Zero،

27 .www.clearview.ai (

ويشير هذا المصطلح إلى العديد من الثغرات الأمنية المكتشفة مؤخرًا في التطبيقات والمواقع ونظم المحادثة والتي يمكن للمتسللين استخدامها لمهاجمة الأنظمة، كما يشير إلى أن المبرمجين قد علموا للتو بوجود ثغرة أو خلل يمكن اختراق النظام من خلاله، وأن أمامهم «**صفر يوم**» للقيام بإصلاح هذه الثغرة، ويحدث هجوم يوم الصفر عندما يستغل المتسللون الخلل قبل أن تتاح فرصة لمعالجته [٢٨]

العديد من المبادرات أيضاً ارتبطت بضرورة تفعيل نظم التأكد من مصادر الصور ومقاطع الفيديو، حيث أطلق عدة جهات متخصصة مثل Adobe و Qualcom و Trupic مبادرة توثيق المحتوى con-tent initiative authenticity مع التركيز على الصور، وذلك من خلال نظام مفعل بالهواتف الذكية وأجهزة الحاسب الآلي يتمكن من الكشف عن الصور المزيفة وتقديم معلومات وصفية دقيقة عنها، ويسمح هذه النظام بتجزئة الأصول المشفرة للصور لكي يتم التحقق منها ومن مصدرها، وفي حالة الكشف عن وجود صورة ملفقة أو مزيفة ورغبة المستخدمين في

28) Wittenberg ,C .al et .(2021) "The Per- (Minimal) Advantage of Political Video over Text ", Proceedings of the National Academy of Sciences ,Vol .118 ,No .47 ,p.20.

> **رغم تطور التكنولوجيا والحاجة إلى مثل هذه المبادرات يظل الاستخدام العام لمثل هذه التقنيات مكلف وغير واسع الانتشار على الأقل في الوقت الحالي**

مشاركتها ستكون مصحوبة بتحذير رقمي بأن هذه الصورة ملفقة أو مزيفة أو تم التلاعب بجزء منها.[٢٩] رغم تطور التكنولوجيا والحاجة إلى مثل هذه المبادرات يظل الاستخدام العام لمثل هذه التقنيات مكلف وغير واسع الانتشار على الأقل في الوقت الحالي.

وفي ذات السياق، وإذا كانت منصات التواصل الاجتماعي هي الوسيلة الأكثر استخداما في نشر وبث ومشاركة المحتوى الزائف، فعلى مسؤولي هذه المنصات التوسع في تفعيل وتدعيم نظم الحماية من خلال وضع علامات لتمييز المحتوى الزائف والعنيف والمضلل، وتتعدد الطرق التي يقترحها الخبراء للقيام بذلك مثل وضع الملصقات على المحتوى العنيف أو الزائف، أو العلامات المائية التي تميزه، أو تفعيل نظم الإبلاغ عن هذا المحتوى وكذلك تفعيل آلية الإزالة بعد التأكد من مضمونه وخطورته، لا شك أن جميع هذه الجهود مرتبطة أيضاً بتعلم الآلة وتدريبها على الفحص والتصنيف.

إدماج الشباب في عملية واسعة من التثقيف والتوعية والتفكير الناقد

في ظل الحديث عن استراتيجية وطنية متكاملة لمواجهة مخاطر توظيف الذكاء الاصطناعي وتطبيقاته في بث خطاب الكراهية والعنف والتحريض، على الدول والوزارات المعنية بالتعليم والتثقيف العمل على تضمين المناهج التعليمية مواد تتعلق بدراسة « التربية الإعلامية» Literacy Media والتي تعنى بدعم قدرات الجمهور والمواطنين في التعامل الناقد مع المواد الإعلامية في الوسائل التقليدية والحديثة، ومن ثم بناء وعي بكيفية التعامل مع الشائعات والدعاوى الهدامة والتخريبية وعدم مشاركتها

29) Helmus, T. C. op.cit, p.11.

مناهج التعليم في كافة المراحل الابتدائية والثانوية والجامعة يجب أن تشمل برامج التربية الإعلامية وتطوير مهارات التفكير الناقد، هذا الأمر لم يعد رفاهية وإنما ضرورة حتمية في ظل تغير أساليب التلقي والتعرض للمواد الإعلامية عبر منصات التواصل الاجتماعي

تعمل برامج التربية الإعلامية على تحفيز قدرات الأفراد على البحث حول مصادر المعلومات وتقييم مصداقياتها وعدم التسليم بكل ما يتم نشره دون تحقق، التدريب على مهارات التفكير الناقد هو حائط الصد الأهم أمام سهولة مشاركة الصور ومقاطع الفيديو عبر منصات التواصل الاجتماعي، وعلى الحكومات والهيئات المعنية تمويل مثل هذه البرامج وتكرار القيام بها على نطاق واسع[30]

وفي دراسة تجريبية حديثة، وجد الباحثون أن الأفراد الذين تلقوا تدريب على طرق التمييز بين المحتوى الحقيقي والزائف (الصور ومقاطع الفيديو التي تم التلاعب بها من خلال الذكاء الاصطناعي) كانوا أكثر ميلا إلى التأكد من مصادر المعلومات مقارنة بالأفراد ممن لم يتلقوا مثل هذا التدريب، كما كشفت نتائج الدراسة أن هؤلاء الأفراد الذين تلقوا برنامج تدريبي للتربية الإعلامية ظهرت لديهم آلية دفاعية تسمى « آلية تعزيز الدفاعات الموقفية» atti- fortifying defenses tudinal ضد كل من الأشكال التقليدية والتزييف العميق للمحتوى المقدم.[٣١]

لا بد من حث الشباب على التحلي بقيم المصداقية والموضوعية واحترام الآخر، ليس هذا فقط، وإنما تحري الدقة في النشر والمشاركة، على سبيل المثال أشارت دراسة حديثة إلى الخطورة الشديدة لما يمكن أن نطلق عليها مقاطع الفيديو المزيفة بسهولة shal- fakes low [٣٢]* ويُطلق عليها أيضا «مقاطع الفيديو الضحلة» ، وهي بمثابة مقاطع فيديو تم تعديلها يدوياً من خلال تقنيات التزييف العميق، وتتسم بمعدل مشاركة مرتفع وتساعد في تضليل وتعمل كآلة لتضليل الجمهور وخاصة الشباب بشكل كبير[٣٣]. يجب أن ينخرط الشباب في عملية واسعة من التعليم وتعلم مبادئ احترام الآخر وعدم مشاركة أي مواد دعائية دون التدقيق في مصدرها وحقيقتها

جهود حكومية لمواجهة الفكر المتطرف ومخاطر التوسع في توظيف الذكاء الاصطناعي

بالإضافة إلى الجهود السابقة على المسؤولين في الدولة والحكومة الإيمان بضرورة البدء في عملية تثقيف تكنولوجي واسع النطاق على مستوى كافة المؤسسات الرسمية ومنظمات المجتمع المدني والعاملين بالمؤسسات الحكومية، بتعلم طرق كشف المحتوى الزائف غير الحقيقي، هذه العملية يجب أن يتم تمويلها وتحديد أطر تطبيقية وعقد ورش عمل متكررة على مستوى كافة القطاعات حتى يمكن بناء وعي عام مجتمعي يمثل حائط الصد الأساسي لأي

30 (Stamos, Alex, et al. (2019). "Combatting Organized Disinformation Campaigns from State Aligned Actors,", Stanford, Calif.: Freeman Spogli Institute for International Studies, Stanford University, p.44.

31 (Hwang, Y., Ji, Y., R. & and Se-Hoon, J., (2021). "Effects of Disinformation Using Deepfake: The Protective Effect of Media Literacy Education," Cyberpsychology, Behavior, and Social Networking, Vol. 24, No. 3, pp. 188–193.

32 *) على الرغم من تزايد التركيز على فهم آليات التزييف العميق deepfakes وتطبيقاته ومخاطرة، فإن العديد من الباحثين المعنيين بتكنولوجيا الذكاء الاصطناعي قد أشاروا إلى مخاطر التوسع في التزييف الضحل أو ما يطلق عليه shallow fakes ، إذ انتشرت العديد من التطبيقات السهلة التي يمكن من خلالها تعديل الصور ومقاطع الفيديو والصوت لإعادة إنتاج محتوى زائف وملفق يمكن مشاركته بسهولة، هذا يحدث بشكل كبير عبر منصات التواصل الاجتماعي، ولذلك يجب التركيز على هذا النوع من التزييف والتوعية بمخاطره وتبعاته

33 (Stoll, Ashley, "Shallow Fakes and Their Potential for Fake News," Washington Journal of Law, Technology, and Arts, January 13, 2020, p.3.

دعاوى تخريبية هدامه يمكنها أن تنال من وحدة وتماسك المجتمع وبناؤه الفكري.

يجب أيضا الاستثمار في دعم التكنولوجيا المتقدمة والمستخدمة في الكشف عن المحتوى المزيف detec-tion technology وذلك من خلال عقد الشراكات مع الجهات المعنية والقطاع الخاص لتعزيز مثل هذه التطبيقات المتطورة والتوسع في استخدامها في الوزارات والهيئات المختلفة للدولة

هذا مع ضرورة تقييد الوصول إلى مصادر المعلومات التي تقدم محتوى مزيف عبر الإنترنت، إذ أن تتبع المواقع الإلكترونية والمنصات المعروفة بنشر وبث خطاب الكراهية بشكل متكرر يجب أن يكون ممارسة مستمرة، ليس هذا فقط بل يجب تقييد وصول المستخدمين لمثل هذه المواقع حفاظًا على السلام المجتمعي وعدم الإضرار بالبناء الفكري السليم للنشء والشباب

ويؤمن الباحث أنه من خلال هذه الاستراتيجية متعددة الأبعاد يمكن الحد من التأثيرات السلبية لتكنولوجيا الذكاء الاصطناعي على المستخدمين والأفراد، مع عدم الجزم بإمكانية منع هذه التأثيرات بشكل كلي، يمكن على الأقل الإسهام في تفعيل وبناء وعي مجتمعي قادر على التمييز بين الحقيقي والمصطنع، ومع تكرار عمليات التثقيف والتوعية لا شك سيتم تفعيل آليات التعرض الانتقائي لدى الأفراد والقائمة على التعرض الواعي لكافة المضامين الحقيقية والابتعاد عن كل محتوى مغرض يحرض على العنف والكراهية وهي البذور الأولية التي تنمو من خلالها وتتغذى الجماعات المتطرفة والإرهابية.

يجب أيضا الاستثمار في دعم التكنولوجيا المتقدمة والمستخدمة في الكشف عن المحتوى المزيف detection technology وذلك من خلال عقد الشراكات مع الجهات المعنية والقطاع الخاص

Gulf Research Center
Knowledge for All

مركز الخليج للأبحاث
المعرفة للجميع

www.grc.net