الحوكمة الخوارزمية لحرية الرأي والتعبير على الإنترنت – تطوير منصات التواصل الاجتماعي لآليات الذكاء الاصطناعي لضبط السلوك البشري في المجال العام الرقمي

إعداد/ محمد فايز محمد حسين

مندوب بمجلس الدولة المصري

ومدرس مساعد بقسم القانون العام بكلية الحقوق بجامعة الإسكندرية سابقًا.

المقدمة

يشهد العصر الحالي ظهور مجتمع المنصات (Platforms Society) أو المجتمع الرقمي (Digital Society) بجانب المجتمع التقليدي، حيث أصبحت الكثير من الأحداث التي تؤثر في العالم الحقيقي تتأثر بشكل مباشر بأوجه النشاط البشري على منصات التواصل الاجتماعي (social platforms)[١]، هذا من ناحية أولى.

ومن ناحية ثانية، أصبح هناك مجالًا عامًا يوازي المجال العام التقليدي، وهو المجال العام الرقمي على المنصات، بل وتأثيره يفوق تأثير المجال العام التقليدي، حيث أصبح محلًا للحديث عن أدق المسائل المرتبطة بالنواحي الاقتصادية والسياسية في العالم الحقيقي، وكأن المنصات بمثابة ساحات عامة رقمية (Digital Public Spaces) لممارسة الحقوق والحريات الأساسية كحرية التعبير وحرية الاجتماع[٢].

رغم أن المنصات جذبت أنظار الفقه في شتى العلوم الاجتماعية، إلا إنها لم تأخذ حيزًا كبيرًا في كتابات فقه القانون الدستوري إلا مؤخرًا، وذلك عندما برزت كقوة مؤثرة على دعائم الاجتماعي الإنساني السلمي في المجتمع الحقيقي، لاسيما أنه بمرور الوقت أضحت الكثير من المؤثرات على النظم السياسية، والسياسات الاقتصادية والاجتماعية، نابعة بشكل مباشر من المظاهر المختلفة لممارسة الأفراد لحرية التعبير في الفضاء الرقمي على تلك المنصات.

ومن أبرز الأمثلة على مظاهر تأثير المنصات، والتي حتمت على الفقه الدستوري بأن يوليها الكثير من اهتمامه، ما يلي؛ (١) أصبحت قوة مؤثرة في تطور واستقرار أسس النظام الديموقراطي، لاسيما بالتأثير على آراء الناخبين، وتلك القوة برزت بالتحديد عقب حادثة تورط شركة (Cambridge Analytica) في التدخل في الانتخابات الأمريكية والتأثير على الناخبين عام ٢٠١٨، وحينها لأول مرة وقف الرئيس التنفيذي مارك زوكربيرغ (Mark Zuckerberg) لشركة تكنولوجية (Meta) – فيسبوك سابقًا (Facebook) – أمام مجلس الشيوخ الأمريكي، وكأن منصة فيسبوك مؤثرة للدرجة التي تستدعي الوقوف على تأثيرها على أسس النظام الديموقراطي في النظم الدستورية[٣].

---

[١] أنظر بشأن مفهوم الحياة المتصلة دائما بالإنترنت (Onlife) وأثر ذلك على العلاقة بين المجتمع الرقمي على الإنترنت والمجتمع التقليدي؛ لوتشيانو فلوريدي، الثورة الرابعة، ترجمة لؤي عبد المجيد السيد، المجلس الوطني للثقافة والفنون والآداب – الكويت، ٢٠١٧. ص ٨٩ وما بعدها.
يذهب البعض من علماء الاجتماع إلى أبعد من مجرد بحث نشأة المجتمع الرقمي، إلى إبراز التأثير السلبي للرأسمالية المعلوماتية (Informational Capitalism) – التي تعد منصات التواصل الاجتماعي أبرز صورها وأدواتها – على المجتمع المحلي. أنظر في ذلك؛ كيت أورتون جونسون ونيك بريور، علم الاجتماع الرقمي – منظورات نقدية، ترجمة هاني خميس، المجلس الوطني للثقافة والفنون والآداب – الكويت، ٢٠٢١. ص ١٢٧ – ١٢٨.

[٢] Thiago Dias Oliva, "Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression," Human Rights Law Review ٢٠, no. ٤ (December ٩, ٢٠٢٠). P ٦١٤.

[٣] حادثة (Facebook–Cambridge Analytica data scandal) تدور فشل فيسبوك – باعتباره منصة لتداول بيانات المستخدمين – في حماية نزاهة العملية الانتخابية، وذلك حينما تدخلت شركة (Cambridge Analytica) – وهي شركة استشارات سياسية إنجليزية تابعة لشركة استخبارات خاصة – في العملية الانتخابية بتتبع ملايين حسابات المستخدمين (ما يقارب ٨٧ مليون حساب) لأجل التأثير في آرائهم لصالح الحملة الرئاسية للرئيس الأمريكي السابق دونالد ترامب، وذلك عن طريق برنامج خاص وهو "حياتك الرقمية" (This Is Your Digital Life)، وهو عبارة عن برنامج يطرح عدة أسئلة، ولكنه في نفس الوقت يخترق بيانات المستخدمين، بل واخترق الرسائل الخاصة فيما بينهم والبعض. وقد تم تطويره من قبل العالم (Aleksandr Kogan) وتم نقلها للشركة سالفة الذكر بدون موافقة المستخدمين مما يشكل انتهاكًا لأحكام استخدام فيسبوك وكذلك لقوانين حماية البيانات الشخصية.

(٢) أصبحت نافذة لبث الأفكار المتطرفة المؤدية للعنف وزعزعة الاستقرار الاجتماعي في كافة المجتمعات، وكمثال على ذلك حادثة مذبحة كرايستشرش (Christchurch) في نيوزيلندا عام ٢٠١٩، حيث كانت على مرآى ومسمع من مستخدمي منصة فيسبوك، ولم تتدخل المنصة حينها لوقف البث لمدة استمرت حتى انتهاء الجريمة، مما دفع قادة المجتمع الدولي مع العديد من مديري الشركات التكنولوجية الكبرى إلى تبني إعلان كرايستشرش (Christchurch Call) للقضاء على المحتوى الإرهابي والعنيف (Extremist Content Online) في المجال العام الرقمي، ومطالبة المنصات بتطبيق التكنولوجيا الحديثة في الكشف التلقائي الفوري عن المحتوى الذي يحرض على الأعمال الإرهابية[٤].

(٣) أصبحت أداة لنشر المعلومات المضللة (Disinformation) بشأن السياسات العامة الاجتماعية والاقتصادية، وذلك بشكل يخرج عن سيطرة أي سلطة عامة، بل وتفاقم الأمر حتى أضحت محلًا لتحدي السياسات الوقائية الصحية التي أتبعتها النظم الاجتماعية لمواجهة فيروس كوفيد ١٩، مما أفضى إلى نتائج كارثية عكس ما اتجهت إليه سياسات الدول كالمطالبة بعدم تلقي التطعيم على سبيل المثال، وأطلق المجتمع الدولي على تلك الظاهرة مفهوم وباء المعلومات المضللة بشأن كورونا (coronavirus infodemic)[٥].

وهكذا أدت الأحداث المتعاقبة إلى تسليط فقه القانون الدستوري الضوء على الشركات التكنولوجية القائمة خلف منصات التواصل الاجتماعي، وكأنها سلطات عامة في المجتمع الرقمي، بحيث تكون ملتزمة بضبط السلوك الإنساني الرقمي، كما تلتزم السلطات العامة الوطنية بذلك في المجتمع الحقيقي. وهكذا أصبحت المنصات طرفًا في العقد الاجتماعي الرقمي الذي يربط بينها وبين الدولة والمواطنين الرقميين (Digital Citizens) في العصر الرقمي[٦].

ولكي تفي المنصات بتلك الالتزامات؛ شرعت في تبني قواعد تشريعية لضبط السلوك الرقمي، وكذلك إنشاء وحدات تقوم بمهام تنفيذية لضمان انصياع مستخدمي المنصات لتلك القواعد، وذلك عن طريق وحدات تنقيح المحتوى المنشور على المنصة (Content Moderation Units)، وأخيرًا إنشاء سلطات تشبه في جوهرها السلطة القضائية؛ لمراجعة قرارات وحدات تنقيح المحتوى.

بمرور الوقت، ومع تزايد وتيرة اتباع الدول نمط التنمية المبني على التحول الرقمي والوصول الشامل والمستدام للإنترنت، لم تعد السلطة التنفيذية الرقمية للمنصات قادرة على مواجهة الحجم الهائل من المحتوى المنشور، مما دفعها إلى تبنى أنظمة الذكاء الاصطناعي لتحقيق الضبط السريع والفعال للسلوك الرقمي (Algorithmic Content Moderation).

---

[٤] OECD, An Introduction to Online Platforms and Their Role in the Digital Transformation (OECD Publishing, Paris, ٢٠١٩). P ٥١.

ألكسندر ميلياغرو هيسشنز ونيك كادربهاي، التطرف على الإنترنت، ترجمة محمد عوض يوسف، مكتبة الإسكندرية، ٢٠٢٠.

[٥] João Marecos et al., "Health Misinformation and Freedom of Expression: Considerations for Policymakers," Health Economics, Policy and Law ١٨, no. ٢ (April ٢٠٢٣).

أنظر أيضًا؛ تقرير المقرر الخاص بتعزيز وحماية الحق في حرية الرأي والتعبير بشأن الأمراض الجائحة وحرية الرأي والتعبير، مقدم لمجلس حقوق الإنسان بالأمم المتحدة، الدورة (٤٤)، بتاريخ ٢٣ أبريل ٢٠٢٠.

[٦] Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media (Yale University Press, ٢٠٢١). P ٢٠٨.

لكن أفضى نشاط تلك الأنظمة إلى المساس بشكل سلبي بحرية الرأي والتعبير وبالحق في الخصوصية والمساواة الرقمية بين الجنسين، على نحو يحول دون الحفاظ على بيئة آمنة للاجتماع الإنساني الرقمي.

تقسيم:

في ضوء ما سبق، يسعى البحث إلى تسليط الضوء على تطور منصات التواصل الاجتماعي كسلطة عامة تتمتع بالسيادة الرقمية في المجتمع الرقمي المعاصر وذلك في المطلب الأول، ثم نتعرض في المطلب الثاني لتطوير المنصات لأنظمة التنقيح الخوارزمي لمواجهة الأشكال المتطرفة من ممارسة حرية الرأي والتعبير على الإنترنت، ثم ننتقل في المطلب الثالث لمناقشة أوجه القصور التي شابت عملية التنقيح الخوارزمي.

المطلب الأول

السيادة الرقمية لمنصات التواصل الاجتماعي

أولًا: مركز منصات التواصل الاجتماعي في النظم الدستورية المعاصرة

يعد ظهور الإنترنت من أكبر المؤثرات على تطور النظم السياسية والاجتماعية في القرن الحالي[٧]، ففي بداية الأمر كان يُنظَر للإنترنت بإعتباره قوة للتحول الإيجابي للمجتمعات، وذلك عند فريق من المبشرين بدور الإنترنت في أيامه المبكرة مثل (John Gilmore) و(John Barlow) في إعلانهم بشأن استقلالية الفضاء السيبراني (Declaration of the Independence of Cyberspace)[٨] ، حيث كان يُعتبر بمثابة مجال عام مفتوح للتعبير عن الآراء، وتلقي المعلومات، وتداولها بحرية كاملة دون امتداد سلطة الدولة لهذا المجال العام الرقمي.

وفي إطار تلك النظرة المتفائلة للإنترنت، قضت المحكمة العليا الأمريكية عام ١٩٩٧ في قضية (Reno v American Civil Liberties Union) بأن الإنترنت هو المجال العام الجديد لتبادل الأفكار بحرية (new free marketplace of ideas)[٩]، وقضت أيضًا في حكمها عام ٢٠١٧ في قضية (Packingham v. North Carolina)

---

[٧] Daniel Valchev, “Constitutional Dimensions of Information Revolution,” in The IT Revolution and Its Impact on State, Constitutionalism and Public Law (Bloomsbury Publishing, ٢٠٢١). P ٤٨ – ٥٣.

يمكننا القول بشكل عام أن تطورات تكنولوجيا المعلومات والاتصالات من الركائز الأساسية لتحقيق التنمية المستدامة في العصر الحالي، فلا غنى عنها للقضاء على الفقر باعتباره الغاية الأولى لأي نمط يُذكر من أنظام التنمية المتبعة. أنظر في ذلك؛ قرار الجمعية العامة للأمم المتحدة رقم (١٥٠/٧٧) بشأن تسخير تكنولوجيات المعلومات والاتصالات لأغراض التنمية المستدامة، الدورة (٧٧)، بتاريخ ١٤ ديسمبر ٢٠٢٢. وكذلك قرار المجلس الاقتصادي والاجتماعي في الأمم المتحدة رقم (٤/٢٠٢٣) بشأن تسخير العلم والتكنولوجيا والابتكار لأغراض التنمية، دورة عام ٢٠٢٣، بتاريخ ٧ يونية ٢٠٢٣.

كما أن الاستراتيجيات الوطنية للتنمية المستدامة تتضمن محاور خاصة بالاستفادة بتكنولوجيا المعلومات وتدعيم الاستثمارات في ذلك القطاع، أنظر في ذلك استراتيجية التنمية المستدامة: رؤية مصر ٢٠٣٠.

بالإضافة لذلك رصدت منظمة التعاون الاقتصادي والتنمية (OECD) الدور المتزايد لمنصات التواصل الاجتماعي في التواصل بين السلطة العامة والمواطنين، وهذا يعد اتجاهًا محمودًا لتدعيم الشفافية والمشاركة العامة، وهي تعد من أسس الحوكمة الرشيدة. أنظر في ذلك؛

Arthur Mickoleit, “Social Media Use by Governments: A Policy Primer to Discuss Trends, Identify Policy Opportunities and Guide Decision Makers,” OECD Working Papers on Public Governance No. ٢٦, ٢٠١٤.

وفيما يتعلق بالتجربة المصرية بالتحديد، أنظر تقرير التنمية البشرية في مصر ٢٠٢١، التنمية حق للجميع: مصر المسيرة والمسار. ص٢٢٣.

[٨] Julie E. Cohen, Between Truth and Power – The Legal Constructions of Informational Capitalism (Oxford University Press, ٢٠١٩). P ١٧٦; Jennifer Cobbe, “Algorithmic Censorship by Social Platforms: Power and Resistance,” Philosophy & Technology ٣٤, no. ٤ (December ٧, ٢٠٢١). P ٧٤٣. Valchev, “Constitutional Dimensions of Information Revolution.” P ٤٨ – ٥٣.

[٩] Reno v. ACLU, ٥٢١ U.S. ٨٤٤ (١٩٩٧).

كما أن الدول تلتزم بشكل عام بموجب القانون الدولي لحقوق الإنسان، بتدعيم الوصول للإنترنت بإعتباره وسيلة أساسية لإعمال طائفة كبيرة من الحقوق الأساسية بخلاف حرية الرأي والتعبير، مثل الحق في التعليم، والمشاركة في الحياة الاجتماعية والثقافية. أنظر في ذلك؛ تقرير مفوضية الأمم المتحدة السامية لحقوق الإنسان بشأن حجب الإنترنت: الاتجاهات والأسباب والتداعيات القانونية والتأثيرات التي تمس بمجموعة حقوق الإنسان، مُقدم لمجلس حقوق الإنسان، الدورة (٥٠)، بتاريخ ١٣ مايو ٢٠٢٢.

بأن المنصات الرقمية بمثابة الساحة العامة الحديثة (modern public square) لاجتماع الملايين من الأفراد، مما يغدو معه من غير الدستوري حرمان الأفراد من الوصول لمواقع التواصل الاجتماعي[١٠].

وعلى نفس المنوال قضت المحكمة الأوروبية لحقوق الإنسان في ديسمبر سنة ٢٠١٢ في قضية (Ahmet Yildirim v. Turkey) بأن الوصول للإنترنت أصبح حقًا دستوريًا لأنه ييسر الوصول للمعلومات بجانب حرية التعبير[١١].

وهكذا أصبح الحق في الوصول للإنترنت (Right to Internet access) من الحقوق الحديثة التي تنطلق بالنظرية الدستورية التقليدية إلى آفاق جديدة تتفق وتحولات المجتمع المعاصر؛ لذلك أعتبر المجلس الدستوري الفرنسي أن الحق في الوصول للإنترنت مستنبط من الدستور الفرنسي[١٢]، وكرست الفقرة الثانية من المادة ١٦ من دستور الإكوادور لعام ٢٠١٨ القيمة الدستورية للحق في الوصول الشامل لتكنولوجيا المعلومات والاتصالات.

كما أن الكثير من الفقه المعاصر يعتبر الحق في الوصول للإنترنت والمشاركة في مجتمع المعرفة من الدعائم الأساسية لسائر الحقوق الأساسية؛ لأن الممارسة الفعالة لسائر الحقوق تعتمد على تداول وتلقي المعلومات الصحيحة، والوصول للإنترنت ييسر الوصول لتلك المعلومات بلا أي قيود[١٣].

تلك النظرة المتفائلة للإنترنت حالت دون امتداد سلطة الدولة للمجال العام الرقمي، حتى تفاقم دور الشركات التكنولوجية والرقمية متعددة الجنسيات، وحلت بقدراتها وإمكانياتها محل الدولة، حتى تنامت سلطاتها على نحو لم يكن في الحسبان، مما مكنها من تحديد شروط الحرية في المجال العام الرقمي، وبالأخص التحكم في حرية الرأي والتعبير[١٤].

وهكذا أصبحت المنصات الرقمية تُمثل إحدى الدعامات التنظيمية لعصر الرأسمالية المعلوماتية ومجتمع المعرفة، حيث إنها تنشط في العالم الرقمي عن طريق الربط بين المستخدمين (Users) والمتلقيين (Receivers) وكلاهما يُعتبرا مواطنين رقميين، وتتحكم في حركة نقل البيانات في الفضاء السيبراني (Data traffic).

وهكذا لم يعد الإنترنت بمثابة أداة للتواصل المباشر بين الأفراد في شتى البقاع كما كان الحال خلال تسعينات القرن الماضي وأوائل القرن الحالي، وحينها كانت الدولة هي المتحكم الوحيد، وإنما أصبح مجال تنشط فيه المنصات كوسطاء جدد (New Intermediaries)، ويعتبرها الفقه بمثابة حراس الإنترنت (Internet Gatekeepers) لأنها المسئولة عن

---

[١٠] Packingham v. North Carolina, ٥٨٢ U.S. (٢٠١٧).

[١١] Yildirim v. Turkey (١٨ December ٢٠١٢), application No. ٣١١١/١٠.

[١٢] الحكم منشور كاملًا باللغة الإنجليزية على الموقع الرسمي الخاص بالمجلس الدستوري الفرنسي؛
https://www.conseil-constitutionnel.fr/en/decision/٢٠٠٩/٢٠٠٩٥٨٠DC.htm (Last accessed at: ٢٢/٨/٢٠٢٣).

[١٣] Edoardo Celeste, Digital Constitutionalism – The Role of Internet Bills of Rights (Routledge, ٢٠٢٣). p ٣٠ – ٣٢.

[١٤] Sarah T. Roberts, Behind the Screen – Content Moderation in the Shadows of Social Media (Yale University Press, ٢٠١٩). P ١٣ – ١٤; Paul B. Stephan, The World Crisis and International Law – The Knowledge Economy and the Battle for the Future (Cambridge University Press, ٢٠٢٣). P ٢٠٠ – ٢٠٤. "The technologies that first held out promise of a new dawn of freedom and prosperity, and now seems closer to the opposite, are the poisoned fruits of the knowledge economy."; Cohen, Between Truth and Power – The Legal Constructions of Informational Capitalism. P ٢٣٨ – ٢٣٩.

خلق البنية التحتية الرقمية (Digital Infrastructure) لتيسير حركة التجارة الإلكترونية، وممارسة حرية التعبير في المجال العام الرقمي، وهي بتلك الوظيفة تمارس قدرًا كبيرًا من السيادة الرقمية[١٥].

ومن هنا بالتحديد بدأت منصات التواصل الاجتماعي تلفت أنظار فقه النظرية الدستورية؛ نظرًا لأنها تُيسر ممارسة حرية التعبير، وهي الحرية الأصل في الفكر الدستوري هذا من جانب، ومن جانب آخر تلك الحرية – كسائر الحقوق والحريات بشكل عام – قد تفلت عن عقالها على نحو يؤثر بشكل جسيم على الأمن والسلم الاجتماعي في المجتمع الحقيقي والرقمي على حد سواء.

أما عن الجانب الأول؛ فقد عبرت المحكمة الدستورية العليا عن القيمة الجوهرية لحرية الرأي والتعبير حينما قضت بأن "حرية الرأي تعد من الدعامات الأساسية، التي تقوم عليها النظم الديمقراطية، فقد غدت من الأصول الدستورية الثابتة، في كل بلد ديمقراطي متحضر، وحرصت على توكيدها الدساتير المصرية المتعاقبة ... وحيث إن حرية التعبير تمثل، في ذاتها، قيمة عليا، لا تنفصل الديموقراطية عنها، وإنما تؤسس الدول، على ضوئها، مجتمعاتها، صونا لتفاعل مواطنيها معها، بما يكفل تطوير بنيانها، وتعميق حرياتها، وإن جاز القول بأن لحرية التعبير أهدافها التي يتصدرها بناء دائرة للحوار العام، لا تنحصر آفاقها ولا أدواتها ... كذلك فإن إنماءها للشخصية الفردية، وضمان تحقيقها لذاتها، إنما يدعم إسهامها في أشكال من الحياة، تتعدد ملامحها، بما يكفل حيويتها، وترابطها، فلا يكون تنظيمها مقتضيا إلا أقل القيود التي تفرضها الضرورة. وحيث إن حرية التعبير – كلما كان نبضها فاعلا، وتأثيرها عريضا – هي الطريق لبناء نظم ديموقراطية، تتعدد معها مراكز اتخاذ القرار، تتسم بتسامحها مع خصومها ومسئوليتها قبل مواطنيها وبرفضها لكل قيد يخل بمصداقيتها واستجابتها بالإقناع لإرادة التغيير، وطرحها، من خلال الحوار، لبدائل يفاضلون بينها لاختيار أصلحها، أيا كان مضمونها"[١٦].

وهكذا تستمد منصات التواصل الاجتماعي سلطاتها من كونها ساحة ممارسة الحرية التي تؤسَس عليها الدول، وهي الحرية التي تؤثر في قدر المعلومات المتاحة للأفراد والمجتمع، ومن ثم تؤثر في قدراتهم، أو بمعنى آخر في سلطاتهم، وبحيازتها لهذا القدر الهائل من البيانات والمعلومات، تكون أكثر فعالية من الدول في توجيه السلوك البشري، مما جعلها تتحدى الوظائف التقليدية للدولة[١٧].

---

[١٥] Jamila Venturini et al., Terms of Service and Human Rights (Editora Revan, ٢٠١٦). P ١٩; Cohen, Between Truth and Power – The Legal Constructions of Informational Capitalism. P ٤٨ – ٤٩; Hannah Bloch-Wehba, "Automation in Moderation," Cornell International Law Journal ٥٣ (٢٠٢٠). P ٧٥.

أنظر المرجع التالي بشأن إعادة البحث في مفهوم وخصائص المجال العام من منظور دستوري وذلك في ضوء تطورات تكنولوجيا الاتصالات والإنترنت؛ András Koltay, New Media and Freedom of Expression – Rethinking the Constitutional Foundations of the Public Sphere (Hart Publishing, ٢٠١٩).

[١٦] حكم المحكمة الدستورية العليا في الدعوى رقم ٥٤ لسنة ٣٧ قضائية "دستورية" بتاريخ ٦ يوليو ٢٠١٩م، أنظر كذلك حكمها في الدعوى رقم ١٥٣ لسنة ٢١ قضائية "دستورية" بتاريخ ٣ يونيو ٢٠٠٠م، وحكمها في الدعوى رقم ٣٧ لسنة ١١ قضائية "دستورية" بتاريخ ٦ فبراير ١٩٩٢م، وحكمها في الدعوى رقم ٦ لسنة ١٥ قضائية "دستورية" بتاريخ ١٥ أبريل ١٩٩٥م.

[١٧] Giovanni De Gregorio, "Digital Constitutionalism across the Atlantic," Global Constitutionalism, ٢٠٢٢. P ٣٠٣. "In the algorithmic society, the primary challenge for constitutional democracies no longer comes exclusively from public

أما عن الجانب الثاني؛ أفضت حرية التعبير على الإنترنت إلى نشأة أشكال جديدة من الوقائع التي أحدثت اضطرابات في النظام الاجتماعي والدستوري مثل الاستقطاب السياسي والتضليل المعلوماتي في العالم الافتراضي.

فعلى سبيل المثال كانت منصات التواصل مؤثرة بقوة في أحداث العنف التي شهدتها الانتخابات الأمريكية، مما دفع الشركات الكبرى التكنولوجية المسئولة عن منصات التواصل الاجتماعي إلى حجب الصفحة الشخصية الخاصة بالرئيس الأمريكي السابق والكثير من تابعيه، وذلك تم بدون أي تأثير حكومي آنذاك، نظرًا لأن المحتوى الذي شاركه الرئيس السابق يعد مخالفًا لقواعد وأحكام استخدام منصتي فيسبوك وتويتر.

وكذلك تلعب المنصات دورًا أساسيًا في نشر الدعاوى للإرهاب والعنف (Extremist Content Online)[١٨] . لذلك أصبحت المنصات ملتزمة أمام المجتمع المدني الوطني والدولي بضبط ممارسة حرية التعبير على الإنترنت باعتبارها – كغيرها من الحقوق – لا تستعصى على التقييد.

وقد عبرت المحكمة الدستورية العليا عن ذلك حينما قضت بأن "الحمل على اعتناق بعض الآراء، أو إقماع غيرها، سوءتان تناقضان مفهوم حوار يقوم على عرض الأفكار وتبادلها والإقناع بها. كذلك فإن موضوعية الحوار – وعلى الأخص كلما كان بناء – شرطها شفافية العناصر التي يدور الجدل حولها، بما يحول دون حجبها أو تشويهها أو تزييفها. وتعلق مفهوم الحوار بالقبول بآراء أو برفضها بعد موازنتها ببعض، وعلى ضوء حقائقها وحكم العقل بشأنها، مؤداه أن كل أقوال يكون بها الحوار منتفيًا، كتلك التي تحرض على استعمال القوة استثارة لنوازع العدوان عند من يتلقونها، وإضرارا بالآخرين،

---

authorities but stems primarily from the governance of spaces by formally private actors, which exercise functions traditionally vested in public authorities.". Kate O'Regan, "Public Law, the Digital World and Human Rights: Challenges Ahead," Judicial Review, ٢٠٢٠. P ٤; Wolfgang Schulz, "Changing the Normative Order of Social Media from Within: Supervisory Bodies," in Constitutionalising Social Media (Hart Publishing, ٢٠٢٢). P ٢٣٩.

هذا ما دفع الفقه إلى تصويرها وكأنها بمثابة أنظمة فوق الدول، مثل تعبير (Facebookistan) وهو مشتق من نموذج الشيطان في العقد الاجتماعي عند توماس هوبز (Leviathan)، أو تعبير (Googledom)، وهو مشتق من مصطلح المملكة (kingdom)، أو تعبير (Twitterverse) وهو تصور مشتق من مصطلح العالم (Universe)، وأخيرًا تعبير أرض التيكتوك (TikTokland)، أنظر؛

Edward Lee, "Virtual Governments," UCLA Journal of Law & Technology ٢٧, no. ٢ (٢٠٢٢). p ٤.

[١٨] Folkert Wilman, The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US (Edward Elgar Publishing, ٢٠٢٠). P ٤; Costica Dumbrava, Key Social Media Risks to Democracy (European Parliamentary Research Service (European Parliament), ٢٠٢١).

أنظر بشأن دور منصة فيسبوك في الجرائم التي ارتكبت ضد الروهينجا؛ الدكتورة رنا مصطفى أحمد فؤاد، المسئولية الجنائية عن خطاب الكراهية عبر الإنترنت أمام المحكمة الجنائية الدولية: دراسة تحليلية وتطبيقية على خطاب الكراهية ضد مسلمي الروهينجا في ميانمار، مجلة القانون والتكنولوجيا، المجلد الثالث، العدد الأول، أبريل ٢٠٢٣.

أنظر أيضًا بشأن مسئولية منصات التواصل الاجتماعي عما يُنشر عليها من مواد تسبب أضرارًا في المجتمع الحقيقي؛ أروى محمد تقوى، مسئولية مواقع الويب ٢,٠ عن المحتوى غير المشروع في أوروبا وفرنسا والولايات المتحدة، مجلة الحقوق، مجلد ٤٠، العدد الأول، ٢٠١٦.

لا يجوز أن تتخذ من حرية التعبير سندا، تقديرا بأن مفهوم عرض الآراء والأفكار من أجل تقييمها- على ضوء صحتها أو بهتانها- منحسر عنها، فضلا عن اقترانها بمضار لا يجوز القبول بها"[١٩].

كل هذا أفضى إلى تمتع منصات التواصل الاجتماعي، أو بمعنى أدق الشركات التكنولوجية الرقمية، بمختلف مظاهر السيادة التي تتمتع بها الدولة لأجل حوكمة السلوك الإنساني الرقمي على المنصات، والحيلولة دون زعزعة السلم والأمن الاجتماعي في المجتمع الحقيقي.

ثانيًا: سلطات منصات التواصل الاجتماعي – السيادة الرقمية

من المتعارف عليه أن ممارسة الدولة لسيادتها داخل حدود إقليمها تدور وجودًا وعدمًا مع احتكار سلطتين؛ (١) سلطة ممارسة القوة المنظمة لضبط السلوك البشري. (٢) سلطة اتخاذ القرار في كل ما يخص دعائم الاجتماع الإنساني.

لكن الدولة تعجز عن ممارسة كلا السلطتين – بشكل فردي – في مجتمع المنصات، مما حدا بها إلى الاعتماد على الشركات الرقمية القائمة خلف إدارة وتنظيم المنصات، لتقوم بوظيفة ضبط السلوك في المجتمع الرقمي، وكأن المنصة بمثابة وسيطًا رقميًا (Digital Intermediary) بين السلطة العامة التقليدية والمواطن في المجتمع الرقمي[٢٠].

وهكذا تظهر الشركة التكنولوجية في وسط العلاقة بين الفرد والدولة، مما أدى إلى تمتعها بنوعًا من السيادة الواقعية (Defacto Sovereignty) في المجتمع المعاصر، ولكن ظهورها في قلب طرفي العقد الاجتماعي، منحها سلطة صياغة

---

[١٩] حكم المحكمة الدستورية العليا في الدعوى رقم ٧٧ لسنة ١٩ قضائية "دستورية" بتاريخ ١٩٩٨/٢/٧م.

وعلى نفس المنوال قضت المحكمة الإدارية العليا بأنه "يتعين التفرقة في هذا الصدد بين التعدي على الحق الفردي للأشخاص والتعدي على المجتمع وأمنه وأمانه، وإن كان كلاهما ممقوتا ممجوجًا تلفظه الشرائع ونصوص الدستور والقانون، بيد أن المساس بالحق الشخصي كَفَلَ دَفْعَهُ ولوجُ سبيلِ التقاضي جنائيا أو مدنيا أو كليهما معا، أما حالَ المساس بأمن المجتمع وأمانه فلا يدرؤهُ إلا أن يوصَدَ منبعُ هذا الخطر. وحيث إن حرية الرأي والتعبير تنخرط في مصاف الحريات العامة، وإن تقييدها دون مقتض مشروع إنما يجرد الحرية الشخصية من بعض خصائصها ويقوض صحيح بنيانها، ولازم ذلك أن يكون الأصل هو حرية الرأي والتعبير، والاستثناء هو المنع، إلا أن ذلك لا يعني أن تكون ممارسة هذا الحق بمنأى عن أي قيد، ذلك أن شأنه شأن أي حق من الحقوق العامة يجب ممارسته في حدود القانون، وأن قيام المشرع أو السلطة المختصة بتنظيم ذلك الحق في إطار القانون دون إفراط ولا تفريط، لا يعد منعا أو صدا عن ممارسة هذا الحق؛ ذلك أنه لا يوجد تعارض بين الحرية والتنظيم، بل إن التنظيم هو الذي يعطي المناخ الملائم لممارسة الحق، وبدون التنظيم تضحى الحرية فوضى لا يمكن للفرد أن يحيا في نطاقها"

حكم المحكمة الإدارية العليا في الدعوى رقم ٣٢٣٣٩ لسنة ٥٨ ق "عليا" بتاريخ ٢٠١٥/١/٢٤م، وكذلك حكمها في الدعوى رقم ١٠١٧١ لسنة ٥٤ ق "عليا" بتاريخ ٢٠١١/٢/٢٦م.

[٢٠] أنظر بشأن الشراكة بين الدولة ومقدمي الخدمات على الإنترنت كمنصات التواصل الاجتماعي؛

Attila Menyhárd, “The Impacts of Technological Revolution on the Role of the State,” in The IT Revolution and Its Impact on State, Constitutionalism and Public Law (Bloomsbury Publishing, ٢٠٢١). P ٦٢. “Online service providers (OSPs) transmit content to the users; the transmitted content is produced either by the OSP or by other users. It follows from this that harmful content can only be avoided by involving the OSP in the process. Efficient protection requires the active control of the content by the OSP. Thus, the OSP becomes an important element in the regime of legal enforcement. This changes the role of the state as well and qualifies the relationship between the state and the OSP as a kind of public – private partnership, also called an ‘invisible handshake’ in the legal literature.”

العلاقة بينها وبين الحكومات والمستخدمين – مواطني المجتمع الرقمي – ومن هنا تحولت السيادة الواقعية إلى السيادة الرقمية (Digital Sovereignty)، وكأنها مؤسسة شبه حكومية (Quasi-Governmental Institution)، أو حكومات افتراضية (Virtual Governments)، وبالفعل بدأت الشركات التكنولوجية تعتبر نفسها بمثابة سلطات عامة تساند السلطات الحكومية الوطنية، وتتمتع بقدر من الحقوق بجانب التزاماتها في المجال العام الرقمي[٢١].

تعد السلطات التنظيمية التي تمارسها المنصات في ضبط السلوك في المجتمع الرقمي من أخص مظاهر السيادة الرقمية. وتتمثل تلك السلطات فيما يلي[٢٢]؛

(١) أحكام استخدام المنصة كمظهر من مظاهر السلطة التشريعية (Terms of Services):

نظرًا لأنه لا يمكن فض ما ينشأ من منازعات في المجال العام الرقمي بالرجوع لتشريع دولة بعينها دون أخرى؛ فإن المنصات درجت على سن مجموعة من الأحكام تُمليها الطبيعة العالمية للمنصات، وهي قواعد استخدام المنصة[٢٣].

تتبنى تلك الأحكام ذات القواعد المتعلقة بالصياغة التي يتبناها المشرع الوطني، بحيث تتسم بالعمومية والتجريد وتشبه في ظاهرها لغة التشريعات الوطنية.

تلك الأحكام تعد بمثابة الأسس التشريعية (Lex Digitalis) لسياسات وقرارات السلطة التنفيذية المتمثلة في تنقيح المحتوى. لذلك عبر البعض عن منصة فيسبوك – على سبيل المثال – كمشرع (Lex Facebook)[٢٤].

---

[٢١] Clara Iglesias Keller, "The Perks of Co-Regulation: An Institutional Arrangement for Social Media Regulation?," in Constitutionalising Social Media (Hart Publishing, ٢٠٢٢). p ٢٢٣; Luca Belli and M Horten, "Structural Power as a Critical Element of Social Media Platforms' Private Sovereignty," in Constitutionalising Social Media (Hart Publishing, ٢٠٢٢). p ٩٠; O'Regan, "Public Law, the Digital World and Human Rights: Challenges Ahead." P ٣; Lee, "Virtual Governments." p ١٩ – ٢٢.

أنظر بشأن حقوق المنصات بوصفها (Digital Gatekeepers)؛

Thomas E. Kadri, "Digital Gatekeepers," Texas Law Review ٩٩, no. ٥ (٢٠٢١). P ٩٥٨ – ٩٦٦.

[٢٢] Giovanni De Gregorio, Digital Constitutionalism in Europe – Reframing Rights and Powers in the Algorithmic Society (Cambridge University Press, ٢٠٢٢). P ١١٦ – ١٢٠.

[٢٣] Celeste, Digital Constitutionalism – The Role of Internet Bills of Rights. p ٥٠ "The law established by these private legislators, which one could generally call, lex digitalis – or, in relation to specific actors, such as social media platforms, 'platform law', or, even more specifically, 'lex Facebook', 'lex Google', etc. – shapes the way in which people can behave in the digital environment."; Dias Oliva, "Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression." P ٦١٣; Cohen, Between Truth and Power – The Legal Constructions of Informational Capitalism. P ٤٤ – ٤٦.

[٢٤] مع دخول قانون حماية البيانات الشخصية الأوروبي (European General Data Protection Regulation) حيز النفاذ، راجعت منصة فيسبوك قواعدها وغيرت مسمى أحكام الاستخدام منذ عام ٢٠١٨، لأن التنظيم الأوروبي عمل على توطيد الشفافية بين الشركات والمستخدمين وفرض التزام على جامعي البيانات بتوفير معلومات الخدمة بشكل واضح وصريح ومنسق وسهل الفهم بلغة واضحة وبسيطة، ولذلك غير الفيسبوك المسمى لآخر أكثر وضوحًا، أنظر؛

https://www.facebook.com/legal/terms/previous (Last accessed at ٢٧/٨/٢٠٢٣).

كما أن بعض الفقه يُشبهها بالدساتير؛ لأنها تنظم السلوك البشري في المجال العام الرقمي، وذلك بتحديد التزامات وحقوق المنصة بوصفها سلطة عامة رقمية، والمستخدمين بوصفهم مواطنين الرقميين، وهو ما يماثل ما تتضمنه الدساتير من نصوص بشأن الحقوق الأساسية والالتزامات التي تربط طرفي العقد الاجتماعي ببعض، والهدف في جميع الأحوال يتمثل في تشييد دعائم النظام الاجتماعي، الفارق بينهما أننا نتحدث هنا عن نظام اجتماعي رقمي.

علاوة على ذلك تطور المنصات أحكامها بتطور الظروف الاقتصادية والاجتماعية على نحو يتشابه مع الطابع التقدمي للدساتير باعتبارها وثائق حية. كما أنها تُصنَع من قبل خبراء على المستوى العالمي وبذلك لا تختلف عن الطبيعة الخاصة لصناعة الدساتير الوطنية.

(٢) تنقيح المحتوى كمظهر من مظاهر السلطة التنفيذية (Content Moderation):

تمارس المنصات سلطتها التنفيذية عن طريق تنقيح المحتوى، وتعرفه أستاذة القانون في ستانفورد (Evelyn Douek) بأنه أنظمة وقواعد المنصات لتحديد كيفية تعاملهم مع محتوى المستخدمين على المنصة في ضوء أحكامها[٢٥]، ويمكن تعريفه بشكل شامل باعتباره آليات المنصات لتفعيل قواعدها ومبادئها المنصوص عليها في عقد شروط وأحكام الاستخدام المبرم بين المنصة وبين المستخدم، بحيث يستفيد الأخير من خدمات المنصة مقابل التزامه باحترام أحكام الاستخدام، ويقوم به مجموعة المتخصصين (Moderators) المتعاقدين بعقود دائمة أو مؤقتة مع المنصة[٢٦].

ويعد التنقيح البشري الصورة التقليدية للمراجعة البشرية لمحتوى المستخدمين، وهو يعد بمثابة تنفيذ يدوي لعملية مراجعة وتتبع نشاط المستخدمين على المنصة لأجل تطبيق قواعد المنصة.

وتفتح المنصات باب المشاركة العامة في التنقيح عن طريق آليات الإبلاغ (Flagging Mechanism)، يتمثل هذا الشكل من التنقيح في الاعتماد على إحالة المستخدمين للمحتويات المنشورة للمراجعة، عن طريق الإبلاغ عنها للمنصة، وبذلك نكون أمام شكلًا من أشكال اللامركزية في السلطة الرقمية[٢٧].

Oren Perez and Nurit Wimer, “Algorithmic Constitutionalism,” Indiana Journal of Global Legal Studies, ٢٠٢٣. P ٨.

[٢٥] Evelyn Douek, “Content Moderation As Systems Thinking,” Harvard Law Review ١٣٦, no. ٢ (٢٠٢٢). P ٥٢٨.

[٢٦] Koltay, New Media and Freedom of Expression – Rethinking the Constitutional Foundations of the Public Sphere. P ١٨٧.

أنظر بشأن تعريف الدكتورة (Hannah Bloch-Wehba) لتنقيح المحتوى؛

Bloch-Wehba, “Automation in Moderation.” P ٤٨. “By “content moderation,” I mean a platform’s internal decision-making on whether user-generated content violates its rules, and if so, what the penalty might be.”

[٢٧] Douek, “Content Moderation As Systems Thinking.” P ٥٥٢; Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. P ١٢٨; Venturini et al., Terms of Service and Human Rights. P ٥٦.

من الجدير بالذكر أن أغلب المنصات تتبع التنقيح المركزي (centralized moderation)، ولكنها تسمح في الوقت ذاته بالتنقيح اللامركزي (decentralized moderation) مثل منصة (Reddit)، أنظر في ذلك؛

(٣) وسائل الترضية القضائية الرقمية كمظهر من مظاهر السلطة القضائية:

تعرضنا في السطور السابقة لمظاهر السلطتين التشريعية والتنفيذية التي تتمتع بها المنصات، ويتبقى لنا التعرض لوظيفتها شبه القضائية حتى يمكن في نهاية الأمر إطلاق القول بتمتع تلك المنصات بالسيادة الرقمية.

توفر بعض المنصات وسائل للترضية القضائية الرقمية (Digital Justice) بحيث تتيح استئناف ومراجعة القرارات التي اتخذتها وحدات تنقيح المحتوى بشأن نشاط المستخدم المتضرر، وهكذا نكون أمام سلطة قضائية (Quasi-Judicial Function) تضبط العلاقة بين المنصات كسلطة عامة والحقوق الأساسية للمستخدمين كمواطنين رقميين.

تُعبر المنصات عن الرقابة القضائية بمصطلح مجالس منصات التواصل الاجتماعي (social media councils)، وقد ظهرت تلك المجالس في بداية الأمر لتكون بمثابة آلية من آليات الحوكمة الخارجية للمنصات وذلك لضبط سياساتها وقراراتها إزاء حرية التعبير في المجال العام الرقمي لتتوافق مع قواعد القانون الدولي لحقوق الإنسان.

أبرز مثال على السلطة القضائية للمنصات؛ إنشاء منصة فيسبوك للمجلس الإشرافي (Facebook Oversight Board)، وهو النموذج الأكثر تطورًا للرقابة القضائية السيبرانية على قرارات منصة فيسبوك[٢٨].

ينتظم عمل المجلس بموجب مجموعة من الوثائق تتمثل في ميثاق إنشائها (oversight charter) والقانون المنظم للأعمال الداخلية في المحكمة (bylaws) ومدونة السلوك التي تنظم سلوك الأعضاء[٢٩]، وأخيرًا الاتفاق المنظم للأمانة العامة للمجلس (trust).

تتمثل سلطة المجلس طبقا للفقرة (٤) من المادة الأولى للميثاق فيما يلي؛

(١) توجيه طلب للمنصة لتوفير معلومات بشأن الفصل في النزاع المعروض أمامها في وقت معقول وبشفافية.

(٢) تفسير معايير المجتمع وسائر السياسات التي تنظم عمل المنصة.

(٣) توجيه الأوامر للمنصة بإزالة أو الإبقاء على المحتوى أو تأييد قرار المنصة أو التراجع عنه.

(٤) إصدار تفسيرات مكتوبة لقرارات المجلس.

(٥) إصدار إرشادات بشأن سياسة تنقيح المحتوى (policy guidance)؛ لإصلاح سياسات المنصة.

يختص المجلس – بمقتضى المادة الثانية – بالفصل في المنازعات الخاصة بنوعين من قرارات المحتوى؛ (١) قرار المنصة بشأن المحتوى الخاص بالمستخدم (Posting Person)، (٢) قرار المنصة بشأن المحتوى الخاص بمستخدمين آخرين والذي تم الإبلاغ عنه من قبل أحد المستخدمين، ولكن قُوبل إبلاغه بنتيجة لم يرتضها (Reporting Person).

Chuyi Sheng, "Automated Content Moderation," Georgetown Law Technology Review ٦, no. ١ (٢٠٢٢). P ٣٥٣.

[٢٨] سوف نختصر جملة "مجلس الإشراف على الفيسبوك" لكلمة "المجلس" فقط.

[٢٩] Kate Klonick, "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression," The Yale Law Journal ١٢٩, no. ٨ (٢٠٢٠). P ٢٤٧٧; Schulz, "Changing the Normative Order of Social Media from Within: Supervisory Bodies." P ٢٤٣ – ٢٤٥.

ثالثًا: إشكاليات التنقيح البشري ونشأة التنقيح الخوارزمي

ذكرنا سابقًا أن التنقيح يعد مظهرًا من مظاهر السلطة التنفيذية الرقمية وذلك بقيام الفرق المتخصصة على المنصات بمراجعة المحتوى المنشور. ومع تزايد أعداد مستخدمي المنصات في شتى بقاع العالم، تتزايد المواد التي تُنشر على المنصات كصورة من صور التعبير عن الرأي، وبطبيعة الحال تفلت الكثير من المواد المنشورة من الرقابة البشرية التي تستغرق وقتًا طويلًا بطبيعتها، لاسيما إذا ما تأخر إبلاغ سائر المستخدمين عن المحتوى المخالف لسياسات المنصة[٣٠].

علاوة على ذلك أدى انفلات الكثير من المواد المنشورة من المراجعة البشرية، نتيجة تلاعب المستخدمين باللغة أو تأخر إزالة المحتوى، إلى إحداث آثار واقعية في العالم الحقيقي، مما دفع العديد من الدول والمجتمع الدولي إلى الضغط السياسي على المنصات لتطوير أنظمتها الخاصة بتنقيح المحتوى.

كل هذا أدى إلى تطوير المنصات لأنظمة الذكاء الاصطناعي لكي تقوم بدور معاون لوحدات التنقيح البشري في خلق مجال عام رقمي آمن، وذلك حتى تتفادى التأثير السلبي على ربحية المنصة أو مسئوليتها القانونية[٣١].

وهكذا أصبحت المنصات تتبع أنماط مختلفة من التنقيح؛ التنقيح الخوارزمي باستخدام الذكاء الاصطناعي وهو يُستخدم للكشف المسبق للمحتوى قبل نشره أصلًا على المنصة (ex-ante moderation)، والتنقيح البشري التقليدي وهو بمثابة تنقيح لاحق على نشر المحتوى على المنصة (ex-post moderation)، والتنقيح التفاعلي وهو يتم إما بشريًا أو بواسطة الذكاء الاصطناعي وذلك إزاء المحتوى المُحال من قبل المستخدمين (Reactive moderation)[٣٢] .

بشكل عام ينشط دور الذكاء الاصطناعي في الكشف التلقائي عما يُمثل خطاب الكراهية أو الاستغلال الجنسي أو العنف، فعلى سبيل المثال رصدت إحدى الشركات التكنولوجية عام ٢٠١٦ قيام الذكاء الاصطناعي برصد الصور العنيفة أكثر من الرصد البشري[٣٣]. كما أن هناك أنواعًا من المحتوى لا يمكن أن يكتشفها سوى الذكاء الاصطناعي مثل التزييف العميق (AI generated deep-fakes).

وينشط دور الذكاء الاصطناعي في المنصات التي تعتمد على نشر أشكال من المحتويات محمية بموجب قوانين حقوق الملكية الفكرية، كما هو عليه الحال بالنسبة لمنصة يوتيوب؛ حيث تطور المنصة آليات الذكاء الاصطناعي للكشف المسبق عن كل ما يتم بثه ونشره؛ لتحديد مدى مخالفته لقوانين حماية حقوق الملكية الفكرية.

---

[٣٠] Cobbe, "Algorithmic Censorship by Social Platforms: Power and Resistance." P ٧٤٠.

[٣١] Christian Katzenbach and Lena Ulbricht, "Algorithmic Governance," Internet Policy Review ٨, no. ٤ (November ٢٩, ٢٠١٩). P ٩ – ١١; Dias Oliva, "Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression." P ٦٠٩; Cobbe, "Algorithmic Censorship by Social Platforms: Power and Resistance." P ٧٤١; Bloch-Wehba, "Automation in Moderation." P ٤٣. "This embrace of automation in moderation displays unwarranted optimism about technology's ability to solve what is fundamentally a social and political problem."

[٣٢] Cambridge Consultants, "The Use of AI in Content Moderation," ٢٠١٩. P ٣٥; Sheng, "Automated Content Moderation." P ٣٥٤.

[٣٣] Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. P ٩٧.

لكن رغم تطوير المنصات للذكاء الاصطناعي – لاسيما إن أغلب الشركات العاملة خلف المنصات الرقمية هي ذاتها الشركات التكنولوجية الكبرى الداعمة لأبحاث الذكاء الاصطناعي – إلا إن أنظمة التنقيح الخوارزمي لا تعمل بمفردها على المنصات، حيث يحيل الذكاء الاصطناعي المحتوى للفرق القائمة على التنقيح البشري، إذا ما عجز عن تحديد ما إذا كان المحتوى يُشكل انتهاكًا لسياسات المنصة من عدمه. لكن ذلك لا يقلل من فائدتها، إذ إنها تتمكن من ضبط حرية الرأي على المنصات التي يستخدمها أعداد مهولة من المستخدمين.

المطلب الثاني

آليات التنقيح الخوارزمي

أولًا: التنقيح عن طريق التصميم (Moderation By Design)

أول تطبيق من تطبيقات الذكاء الاصطناعي يتمثل في اختيارات التصميم المسبقة (ex-ante design choice)، أو التنقيح عن طريق التصميم (moderation by design) أو توجيه السلوك عن طريق هندسة العالم الرقمي[٣٤].

[٣٤] أنظر بشأن المفاهيم المختلفة التي عبر عنها الفقه بشأن دور تصميم المنصة في توجيه السلوك؛

Belli and Horten, "Structural Power as a Critical Element of Social Media Platforms' Private Sovereignty." p ٩٤. "Digital platforms can enforce their private ordering through several technical measures. In this sense, the structural power they exercise becomes evident as they enjoy the unique capacity to defi ne the software architectures of that shape the way platform's function. These software structures have a performative nature, as they do not simply defi ne what behaviours are admitted within a platform, but they operationalise such prescriptions by building a structure that only allows users to behave as prescribed."

Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. P ٢١ – ٢٣. "Platforms may not shape public discourse by themselves, but they do shape the shape of public discourse."

Celeste, Digital Constitutionalism – The Role of Internet Bills of Rights. p ٥٤. "In their virtual domains, technology companies are not only legislators, as the state in its territory, but also architects and engineers of the virtual reality where individuals act."

تدور فكرة التنقيح عن طريق التصميم حول قدرة المنصة على التأثير على السلوك البشري والتحكم في الاختيارات المتاحة أمام المستخدمين عن طريق تشكيل هيكل النشاط الاجتماعي الرقمي (software structure)، كأن تخلق بنية المحتوى (software code) بشكل يؤثر على المحتوى ذاته بدون إدراك المواطن الرقمي بالقيود الواردة على اختياراته، لذلك يعتبر الكود والنظام الهندسي بمثابة سيطرة خفية (passive form of control)[٣٥].

على سبيل المثال فالمواطن الرقمي مُستخدم المنصة لا يملك سوى الاختيارات التي تتيحها له المنصة فلا يمكنه اتباع سلوك مختلف، فيتم استبعاد بعض الاختيارات من نطاق الاختيارات المتاحة أمامه، إذا كانت الاختيارات المُستبعَدة تفضي إلى مخالفة قواعد المنصة. وهكذا يتم تضمين القاعدة القانونية الموجهة للسلوك في تصميم النظام، بحيث يكون الكود (Code) هو الأداة التنظيمية التي تحجب اختيارات معينة بحيث لا تظهر أصلًا أمام المستخدم[٣٦].

تلك القدرة على تحديد البنية التحتية للسلوك والاختيارات البشرية لا يمكن للنظام القانوني القيام بها؛ حيث إن الفرد في العالم الواقعي يملك حرية الاختيار في جميع الأحوال، ولا يملك القانون سوى ربط الثواب أو العقاب بما ينتج عن الاختيار البشري من آثار في العالم الخارجي، ولذلك فإن للكود قدرة أفضل من القانون في توجيه السلوك البشري لأنه يحجب الاختيارات المخالفة للقانون من أمامه أصلًا. كما أن السلطة التنفيذية للدولة في المجتمع الحقيقي تتقيد بالظروف والعوامل الطبيعية مثل وفاة الشخص، لكن سلطة المنصات مقيدة بالعوامل التقنية فقط للنظام[٣٧].

وتستخدم المنصة التنقيح عن طريق التصميم في حالة صناعة كود المنصة بحيث يقلل من فرص ظهور المحتوى غير المشروع على موجز الأخبار الذي يظهر للمستخدمين (News Feed)، وتعد منصفة فيسبوك ويوتيوب وتويتر ومحرك البحث الخاص بغوغل من المنصات الأكثر تطورًا فيما يتعلق بتطبيقات أنظمة الذكاء الاصطناعي المتعلقة بصناعة المنصة الرقمية.

ثانيًا: التنقيح الخوارزمي المُغذى مسبقًا (Matching Algorithm)

التطبيق الثاني من تطبيقات الذكاء الاصطناعي هو التنقيح الخوارزمي المُغذَى مسبقًا أو نظام المنصة للكشف التلقائي (platform's automatic detection software)، في هذا النظام يعمل الذكاء الاصطناعي على فحص المحتوى المنشور على المنصة، ومقارنته بالمعلومات المخزنة مسبقًا في النظام، لتحديد مدى مخالفته لقواعد المنصة.

التطبيق الفعال لذلك النمط من التنقيح يتطلب إدخال معلومات معينة وكلمات مفتاحية ترشد الذكاء الاصطناعي عما إذا كان هناك محتوى مخالف للقواعد من عدمه، أي يتم تغذيته مسبقًا بالوقائع السابقة التي تُمثل مواد غير مشروعة، ويتعرف عليها عن طريق عملية تُسمى (Matching)[٣٨].

---

[٣٥] Przemysław Pałka, "Terms of Injustice," SSRN Electronic Journal, ٢٠٢٣. P ١٤ – ١٥.

[٣٦] Amy Kapczynski, "The Law of Informational Capitalism," Yale Law Journal ١٢٩, no. ٥ (٢٠٢٠). P ١٤٧١; Douek, "Content Moderation As Systems Thinking." P ٥٤٦.

[٣٧] Eric Goldman, "Content Moderation Remedies," Michigan Technology Law Review ٢٨, no. ١ (٢٠٢١). P ٩ – ١١.

[٣٨] Sheng, "Automated Content Moderation." P ٣٥٥.

يتعرف الذكاء الاصطناعي على المحتوى المنشور عن طريق تجزئة المحتوى (hashing) إلى بصمات رقمية (digital fingerprints) والمقارنة بين بصمات المحتوى محل التنقيح والمحتويات المُغذَى بها الذكاء الاصطناعي، ولكن نادرًا ما يتحقق التطابق الكلي لسهولة التلاعب باللغة أو الصور من قبل المستخدمين، ولذلك فان هناك هامش للاختلاف بين البصمات الرقمية يأخذه الذكاء الاصطناعي في الاعتبار، كأن نكون بصدد صورة تمثل شكلًا من أشكال العنف، ولكن تختلف عن الصورة المُغذى بها الذكاء الاصطناعي من حيث الإضاءة (Perceptual hashing)[٣٩].

يجسد هذا التطبيق مبدأ الوقاية في المجتمع الرقمي؛ حيث يعمل الذكاء الاصطناعي على التطبيق التلقائي لأحكام وشروط الاستخدام، بحيث يتم الكشف على المواد المنشورة بمجرد نشرها، لذلك يطلق الفقه على ذلك التطبيق مصطلح إزالة المحتوى بشكل استباقي (Proactive content removals).

يلعب ذلك النمط من التنقيح دورًا هامًا في المحتويات الشائعة غير المشروعة، والتي تُشكل مخالفة لسياسات وأحكام استخدام المنصة، والتي بدورها تُشكل جريمة في العالم الحقيقي كالتهديدات بالعنف والتخريب أو المحتوى المنافي للآداب وما يشكل جرائم للاستغلال الجنسي للأطفال.

فعلى سبيل المثال تطبق المنصات نظام ترشيح الكلمات (Word filters)، بحيث يتم تغذية النظام بعبارات معينة تندرج ضمن المحتوى غير المشروع، أو نظام (PhotoDNA) – نظام طورته شركة مايكروسوفت عام ٢٠٠٩ وتم توفيره بشكل مجاني لمنصات التواصل الاجتماعي – حيث يتم تغذيته مسبقًا بصور كثيرة لما يشكل مظاهر للعنف الجنسي والاعتداء على الأطفال، ويتم ربطه بقواعد البيانات الخاصة بالأطفال الخاضعين للاستغلال أو المفقودين، بحيث يتم الكشف الآلي على ما يتم نشره للبحث عما إذا كان هناك تطابق بين العبارات أو الصور المنشورة وبين العبارات أو الصور المحمل بها الذكاء الاصطناعي[٤٠].

وكذلك تكونت شراكة بين يوتيوب ومايكروسوفت وتويتر وفيسبوك عام ٢٠١٦ لأجل تطوير ومشاركة قاعدة بيانات بشأن المحتوى الذي يعد إرهابيًا، وذلك لإزالته فور نشره، كما أن هناك تقنيات متطورة للكشف الآلي عن المحتويات المخالفة للآداب[٤١].

كما أن هذا الشكل من التنقيح تطبقه أغلب المنصات للكشف عن المحتويات التي تنتهك حقوق الملكية الفكرية كنظام (Content ID) على سبيل المثال. لاسيما وأن قوانين حقوق الملكية الفكرية مثل قانون حقوق المؤلف للألفية الرقمية في أمريكا (Digital Millennium Copyright Act) والتوجيه الأوروبي للتجارة الإلكترونية (Directive on electronic

[٣٩] Dias Oliva, "Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression." P ٦٠٩.

[٤٠] Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. P ٩٨ – ١٠١; Bloch-Wehba, "Automation in Moderation." P ٥٨.

[٤١] Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. P ١٠٣; Bloch-Wehba, "Automation in Moderation." P ٥٩ – ٦٢.

commerce)[٤٢]، تلزم المنصات بإزالة المحتويات المنشورة، إذا ما كان لدى مزودي الخدمات (online service providers) – أي المنصات – علم فعلي بوجود تلك المواد على المنصة وبانتهاكها لحقوق الملكية، ومن ثم تستخدم المنصات الذكاء الاصطناعي حتى لا تثور مسئوليتها القانونية إزاء انتهاك حقوق الملكية الفكرية.

ثالثًا: التنقيح الخوارزمي ذاتي التعلم (Predictive Algorithm)

طورت المنصات نظم أكثر تعقيدًا للتنقيح الخوارزمي، ونخص بالذكر التنقيح الخوارزمي التنبؤي، وهو قائم على فكرة التعلم الآلي الذاتي للذكاء الاصطناعي (Machine Learning Systems)؛ بحيث يتنبأ بما إذا كان المحتوى غير مشروع من عدمه بناء على الاحتمالات، وخاصة إذا ما كان المحتوى محل المراجعة غير وارد في قاعدة البيانات المُغذى بها الذكاء الاصطناعي مُسبقًا[٤٣].

لكن صناعة تطبيقات التنقيح الخوارزمي التنبؤي تعد أكثر تعقيدًا من صورته السابقة، حيث يجب جمع البيانات بشأن مهمة محددة ثم تدريب الذكاء الاصطناعي على التنبؤ باستخدام البيانات الموجودة، ثم اختباره عن طريق عرض معلومات وبيانات جديدة عليه، لمعرفة هل النظام يقوم بالتنبؤات الصحيحة من عدمه، ثم إعادة الاختبار للكشف عن أخطاء التنبؤ، ثم إعادة تشغيل النظام، مع المراقبة المستمرة لعمل النظام[٤٤].

يتم إعمال ذلك النظام على وجه التحديد في مواجهة التلاعب باللغات والكلمات والحروف (speech and language recognition) وذلك بتحليل المحتوى المكتوب وتصنيفه، فعلى سبيل المثال يلجأ الذكاء الاصطناعي لتقنية مجموعة الكلمات (bag of-words) لتحديد ما إذا كان المحتوى المكتوب يشكل خطابًا للكراهية من عدمه بدون الاهتمام بترتيب الكلمات أو التكوين النحوي للمكتوب، بل يهتم بكل كلمة على حدة ويمنحها قيمة رقمية لكي يحلل قيمة المحتوى المكتوب كاملًا بغض النظر عن التلاعب في الترتيب.

كما أن هناك نظام أخر يُسمى(N-grams) يركز على مجموعات الكلمات وليس كل كلمة على حدة، فالكلمة بمفردها قد لا تدل عما إذا كانت بمثابة خطاب كراهية من عدمه، ولكن مع غيرها من الكلمات تعطينا صورة أكثر وضوحًا. كما أن هناك نظام يسمح للذكاء الاصطناعي بالتعرف على سياق الكلمات (word embedding). كل تلك الآليات تحول دون هروب المستخدمين للمنصة من التلاعب بقواعد المنصة عن طريق استبدال الكلمات أو حذف الحروف.

---

[٤٢] Bloch-Wehba, “Automation in Moderation.” P ٦٢ – ٦٦.

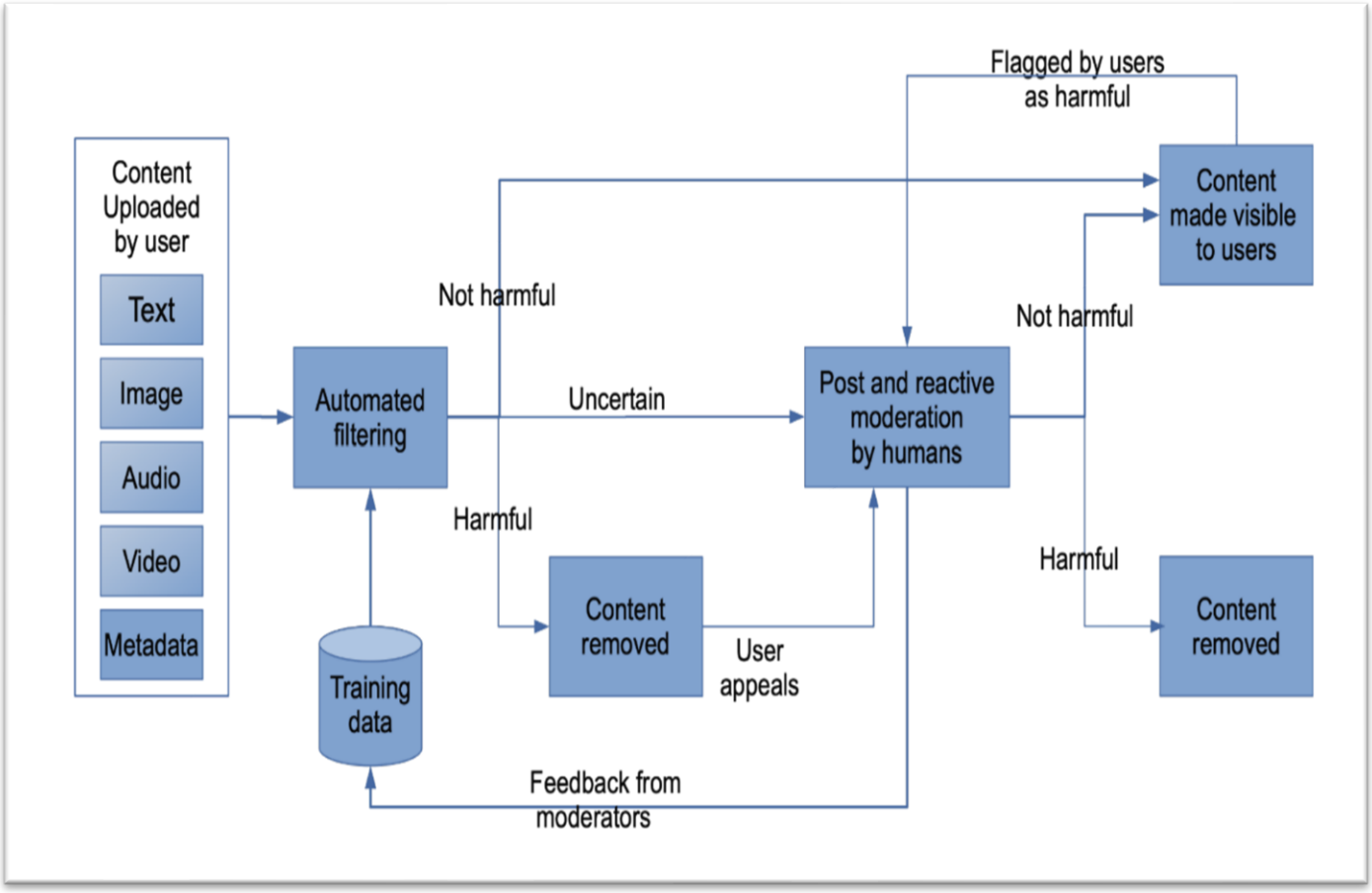
[٤٣] Sheng, “Automated Content Moderation.” P ٣٦٠; Cobbe, “Algorithmic Censorship by Social Platforms: Power and Resistance.” P ٧٤١; Nafia Chowdhury, “Automated Content Moderation: A Primer,” ٢٠٢٢. P ٢; Bloch-Wehba, “Automation in Moderation.” P ٥٦. “Modern automated techniques often use machine learning algorithms to predict the likelihood that a piece of content violates the platforms’ rules or the law.”

أنظر بشأن آليات تعلم الذكاء الاصطناعي (Machine Learning Systems)؛ د محمود حسن السحلي، أساس المسائلة المدنية للذكاء الاصطناعي المستقل: قوالب تقليدية أم رؤية جديدة؟، مجلة كلية الحقوق للبحوث القانونية والاقتصادية، العدد الثاني، ٢٠٢٢. ص٥٤ وما بعدها.

[٤٤] Chowdhury, “Automated Content Moderation: A Primer.” P ٢ – ٦.

بخلاف اللغة، يمكن تدريب النظام للتعرف على صور العنف (graphic violence) لإدراك المحتويات المصورة والمشاهد المنشورة (Scene understanding)، عن طريق البحث عن البحث عن أشكال معينة (Object detection algorithms)[٤٥].

وسواء أكنا نتحدث عن التنقيح الخوارزمي المُغذى مسبقًا أو ذاتي التعلم، فإن وجود العنصر البشري أساسي لعمل النظام؛ لأجل تزويد الذكاء الاصطناعي بالبيانات أو لإعادة ضبط قدرته على التنبؤ والكشف على المحتويات المخالفة لقواعد المنصة. ويوضح الشكل التالي تفاعل التنقيح البشري مع أنظمة الذكاء الاصطناعي على المنصات.



المصدر: مأخوذ مباشرة من تقرير تم نشره عام ٢٠٢٠، وتم إعداده من قبل الدكتور (Giovanni Sartor) والدكتور (Andrea Loreggia)، وذلك بناء على طلب لجنة البرلمان الأوروبي المعنية بحقوق المواطنين والشؤون الدستورية. أنظر؛

Giovanni Sartor and Andrea Loreggia, "The Impact of Algorithms for Online Content Filtering or Moderation," Policy Department for Citizens' Rights and Constitutional Affairs, ٢٠٢٠. P ٢٣.

[٤٥] Ibid. P ٩.

رابعًا: حوكمة المحتوى على الإنترنت: منصة فيسبوك نموذجًا[٤٦].

إذا ما انتقلنا لمنصة فيسبوك على التحديد، تُطبق المنصة نظامين للتنقيح؛

(١) النظام الأول يتمثل في حالة الإبلاغ عن المحتوى من قبل المستخدمين أو السلطات العامة أو اكتشاف الذكاء الاصطناعي للمحتوى. في تلك الحالة يفحص الذكاء الاصطناعي البلاغات لتحديد مدى توافقها مع المنصة، وإذا عجز الذكاء الاصطناعي عن التعرف على ما إذا كان المحتوى مخالف من عدمه يمكنه الإحالة لوحدات التنقيح البشري.

(٢) النظام الثاني يتمثل في حالة ما إذا كان المحتوى يندرج ضمن طائفة المحتويات الخاضعة لنظام المراجعة المزدوجة (cross-check system). في تلك الحالة تكون هناك محتويات بعينها من شخصيات أو منظمات ذات نفوذ وتأثير على مجريات الأحداث في العالم الحقيقي، بحيث تتمهل المنصة في تنقيح المواد المنشورة من قبل تلك المنظمات أو الأفراد، فتخضع لنظام تنقيح أكثر تعقيدًا عن النظام الذي تخضع له سائر المحتويات المنشورة من باقي المستخدمين. ويلعب الذكاء الاصطناعي دورًا في التنقيح المزدوج عن طريق الكشف السريع عما يُنشر من قِبل تلك المنظمات أو الأفراد المؤثرين ثم إحالة المحتوى لفرق متخصصة لاتخاذ القرار بشأنها[٤٧].

تُستخدَم التكنولوجيا في الحالتين في الحذف المسبق لأغلب المحتويات قبل نشرها على المنصة للعامة، ولكن كثيرًا ما تثور معوقات أمام الذكاء الاصطناعي في مرحلة التعرف على المحتوى، فقد تكون اللغة مختلفة أو النية الكامنة خلف المحتوى غير مشروعة أو الصورة معتمدة على سياق يؤدي إلى تغيير معناها، وهو الأمر الذي لا يستطيع الذكاء الاصطناعي الكشف عليه بسهولة، وهنا يأتي دور الإحالة للمراجعة البشرية.

يحدد الذكاء الاصطناعي أولوية العرض على التنقيح البشري، بحيث يوفر الوقت على وحدات التنقيح البشري حتى تركز على المنشورات الأكثر تأثيرًا، وبالتالي تكون وتيرة التنقيح أسرع، بجانب تدريب وتطوير الذكاء الاصطناعي.

يحدد الذكاء الاصطناعي ما إذا كان المحتوى محل المراجعة ينبغي إحالته للمراجعة البشرية قبل اتخاذ القرار، بناء على عدة عوامل تحدد أولوية العرض، وهي؛

(١) مقدار الضرر الذي يمكن أن يحدث من المحتوى سواء في العالم الافتراضي أو الحقيقي (high-severity content).

(٢) ما إذا كان المحتوى مؤثرًا أو من شأنه التأثير في مجريات الأحداث عن طريق الوصول لأكبر عدد من المستخدمين وانتشاره.

(٣) مدى احتمالية مخالفة المحتوى لسياسات المنصة.

---

[٤٦] Edoardo Celeste et al., "Digital Constitutionalism: In Search of a Content Governance Standard," in Constitutionalising social media (Hart Publishing, ٢٠٢٢). P ٢٧٨ – ٢٨٦.

https://transparency.fb.com/ar-ar/enforcement/detecting-violations/technology-detects-violations/ (Last accessed at ١/٩/٢٠٢٣).

[٤٧] Perez and Wimer, "Algorithmic Constitutionalism." P ١٠.

عقب إحالة الذكاء الاصطناعي المحتوى للتنقيح البشري، يتم إضافة المعلومات الخاصة بالمحتوى لقواعد البيانات الخاصة بالذكاء الاصطناعي بحيث يأخذ القرار بشكل آلي عند التعرض لنفس المحتوى دون الرجوع للمراجعة البشرية. وهناك فرق متخصصة لتطوير الذكاء الاصطناعي (artificial intelligence teams) تعمل على بناء قواعد البيانات، ونماذج للتعلم الآلي (machine learning models) لأداء المهام بصورة سريعة مثل التعرف على المنشورات وفهم اللغة وإدراك السياق.

يتم تدريب الذكاء الاصطناعي بالتكرار؛ في أول الأمر لا يكون لدى الذكاء الاصطناعي الثقة الكافية بشأن ما إذا كان المحتوى مخالف من عدمه، لكن بمرور الوقت يتعلم الذكاء الاصطناعي من المنقح البشري كلما تكررت إحالة الأول للثاني.

وبالتالي هناك عملية تعلم مزدوجة تشمل تعليم القائمين على التنقيح البشري، وهم بدورهم مسئولين عن تعليم الذكاء الاصطناعي وتطويره. ثم ما تلبث أن تتكرر المخالفات، ويزداد الذكاء الاصطناعي ثقة في اتخاذ القرار، وهذا مفيد على وجد التحديد في حملات التضليل المعلوماتي والإرهاب لأنها تتبع نمط متكرر.

لكن تكمن المشكلة الأساسية في أنماط المخالفات غير المتكررة أو الأنماط المتكررة التي خضعت لتغييرات جوهرية في السياق أو اللغة أو الصورة لتضليل أنظمة الذكاء الاصطناعي. هنا إما أن يحذف الذكاء الاصطناعي ملايين المنشورات غير المخالفة أو يترك مخالفات تزعزع الاستقرار الاجتماعي في المجتمع الرقمي والحقيقي على حد سواء. في كل الأحوال يجب ضمان التدريب الجيد لأنظمة الذكاء الاصطناعي. وهنا تأتي وظيفة فريق الذكاء الاصطناعي المسئول (Responsible AI team).

يضم هذا الفريق علماء من تخصصات مختلفة في ميادين العلوم الاجتماعية والسياسية وخبراء السياسة العامة والمهندسين وباحثين في مجال الذكاء الاصطناعي، وهم مختصين بتطوير إرشادات وأدوات وإجراءات لتدعيم مسئولية الذكاء الاصطناعي.

ونظرًا لأن الكشف عن خطاب الكراهية على منصة فيسبوك يثير عدة صعوبات سواء للذكاء الاصطناعي أو لفرق التنقيح البشري، من حيث اختلاف اللغة، و المدلول الثقافي، والأغراض الحقيقية، وتأثير السياق على المحتوى المنشور، تستثمر المنصة – وأغلب الشركات التكنولوجية كغوغل ومايكروسوفت بالتحديد – في أنظمة الذكاء الاصطناعي لتحسين الكشف على المحتويات المخالفة على الإنترنت سواء أكنا نتحدث عن محركات البحث أو منصات التواصل، ومن ضمن التكنولوجيا الحديثة التي طورتها المنصة؛ (١) نظام (Linformer) موجه لكشف خطاب الكراهية والمحتوى الداعي للعنف، (٢) نظام (Reinforced Integrity Optimiser) خاص بخطاب الكراهية، (٣) نظام (SimSearchNet) لمواجهة التضليل المعلوماتي، (٤) أنظمة (XLM and XLM-R) للكشف عن المحتويات المخالفة بشتى اللغات.

المطلب الثالث

إشكاليات التنقيح الخوارزمي

أولًا: الإشكاليات

ذكرنا سابقًا أن المنصات شرعت في تطوير أنظمة الذكاء الاصطناعي لتنقيح المحتوى قبل نشره على المنصة. وكانت تلك الخطوة محلًا للإشادة من المجتمع الدولي في بداية الأمر، ثم ما لبث أن ظهرت أشكال أخرى من انتهاكات الحقوق الأساسية للمستخدمين كنتيجة لأنظمة التنقيح الخوارزمي، مما أثار العديد من الشكوك بشأن مدى حيادية المنصات في تطويرها لتلك الأنظمة، وقدرتها على القيام بمهام التنقيح بفعالية، وبأقل تكلفة، ودون المساس بحقوق المواطنين الرقميين. وسنتعرض الآن لبعض تلك الإشكاليات؛

(١) الآثار الهيكلية الاجتماعية للذكاء الاصطناعي:

ذكرنا سابقًا أن التنقيح الخوارزمي القائم على تصميم المنصة يؤثر في الاختيارات المتاحة أمام المستخدمين، مما يفضي إلى قدرة الذكاء الاصطناعي على تحديد الحقيقة والخطاب المقبول في المجال العام الرقمي، والتحكم في مسار المعلومات وانتشارها[٤٨]، ومن ثم يؤثر على القوة النسبية للمستخدمين، فيستفيد البعض من اختيارات غير متوافرة لغيره من المستخدمين مما يكسب المنصات سلطة بشأن هيكلة المراكز الاجتماعية الرقمية، وهذا ما يُطلق عليه الفقه بالآثار الهيكلية للذكاء الاصطناعي على المنصات[٤٩].

تظهر تلك الآثار الهيكلية بالتحديد في قدرة الذكاء الاصطناعي على التحكم في الاقتراحات التي تُعرَض أمام المستخدم (algorithmic recommendations) والأخبار الشائعة (News Feed ranking process) أو تطورات الأحداث (Trending algorithms)[٥٠]، وهكذا يظهر لكل مستخدم ما يحدد الذكاء الاصطناعي أنه محتوى مناسب وملائم

---

[٤٨] Cobbe, "Algorithmic Censorship by Social Platforms: Power and Resistance." P ٧٥٣.

[٤٩] Ibid. P ٧٤٢ – ٧٤٣; P ٧٥٦ – ٧٥٨.

[٥٠] قد يؤدي التنقيح عن طريق التصميم إلى آثار كارثية، على سبيل المثال تأخرت منصة فيسبوك في غلق صفحة تدعو بشكل صريح لحمل السلاح للتظاهر في بلدة كينوشا في أحداث الشغب في عهد الرئيس الأمريكي السابق ترامب (Kenosha Guard's event)، وترتب على خطأ المنصة، أن استجاب العديد لنداء الصفحة مما أفضى إلى حدوث أعمال شغب ووفيات، إلى أن انتبهت المنصة للصفحة عقب ذلك، ويكمن الخلل هنا في أن الحدث الخاص بالصفحة تم الإبلاغ عنه ٤٥٥ مرة، وفي كل مرة لا يرى القائمون على التنقيح تحقق أي مخالفة لأحكام المنصة، وعقب الحادثة علق مارك على الواقعة بإنها خطأ في تصميم المنصة ذاتها وليس خطأ تنقيح فردي، وأن تصميم المنصة ساعد على انتشار الصفحة لعدد كبير من التابعين.
https://www.buzzfeednews.com/article/ryanmac/kenosha-militia-facebook-reported-٤٥٥-times-moderators. (Last accessed at ١/٩/٢٠٢٣).

له، ولكل مستخدم اقتراحات تختلف عن غيره، وتؤثر في تكوين آراءه وأفكاره واختياراته، وهذا الشكل من التحكم في المحتوى المعروض يعد نوعًا من التمييز بين المستخدمين[٥١].

وقد ظهرت تلك المفاضلة في نظام الاقتراحات على منصة يوتيوب[٥٢]، ونظام الإعانات على فيسبوك (ad delivery system)؛ حيث رصد الفقه تأثر توزيع وعرض إعلانات المنازل في الولايات المتحدة الأمريكية بالتحيزات الجنسية والعقائدية أو الحالة الاجتماعية ومحل الميلاد، مما يشكل مخالفة للقانون الأمريكي بشأن السكن العادل (Fair Housing Act)، مما حدا بالمنصة إلى إبرام اتفاق مع مكتب العدالة في الولايات المتحدة في يونيو ٢٠٢٢ بشأن المراجعة المستمرة لنظام الإعلانات، وتوقفت المنصة عن استخدام نظام الإعلانات المتخصصة الموجهة لجمهور بعينهم (Special Ad Audience)[٥٣].

(٢) التحيزات الخوارزمية وعدم المساواة الرقمية:

نظرًا لأن التنقيح يعد في حقيقة أمره قرارًا بشأن ما ينبغي أن يكون عليه سلوك المواطنين في المجال العام الرقمي، فإنه يتأثر بالقيم (Values)، والأحكام الأخلاقية (moral judgments)، والأيديولوجيات، والتحيزات (Biases) التي تؤثر في سلوك القائم على عملية اتخاذ القرار بشأن المحتوى المنشور على المنصة.

ولما كان التنقيح الخوارزمي يعتمد في ملامحه الأساسية على قواعد بيانات متوافرة بشأن ما يعد من أوجه التعبير غير المشروع، فإنه من المتصور أن تنتقل القيم والتحيزات البشرية إلى الذكاء الاصطناعي عبر عملية التغذية البشرية

---

[٥١] نظرًا لأن المنصات الرقمية تعمل على التربح عن طريق جذب أنظار مستخدمي المنصة، فإنها – وبطريقة عكسية – تغذي الجهل والتحيز والعنصرية في المجتمع، ومن ثم تخلق حالة من عدم الذكاء الاصطناعي (Artificial Unintelligence)، وهذا ما يفضي إلى تقويض دور المجال العام التقليدي والرقمي في تدعيم أسس الديموقراطية في المجتمعات النامية والمتقدمة على حد سواء.

Meredith Broussard, Artificial Unintelligence: How Computers Misunderstand the World (MIT Press, ٢٠١٨). Mathias Risse, "Artificial Intelligence and the Past, Present, and Future of Democracy," ٢٠٢١. P ٨. "Attention-grabbing algorithms of social media platforms can sow confusion, ignorance, prejudice, and chaos. AI tools then readily create artificial unintelligence."

[٥٢] Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. P ١٧٩. "Social media platforms are about not only the promise of visibility but also the threat of invisibility."

[٥٣] https://www.benton.org/headlines/groundbreaking-resolution-address-discriminatory-delivery-housing-advertisements (Last accessed at ١/٩/٢٠٢٣).

كما أن بعض الفقه يتخوف من سلطة الشركات في جمع بيانات المستخدمين بصورة فائقة عن طريق الرقابة الخوارزمية (platform surveillance)، وتتبع السلطات العامة لتلك البيانات التي تحوزها الشركات التكنولوجية الرقمية وإساءة استعمالها. أنظر في ذلك؛

Bloch-Wehba, "Automation in Moderation." P ٨١; Cobbe, "Algorithmic Censorship by Social Platforms: Power and Resistance." P ٧٤٦ – ٧٤٨.

لقواعد البيانات المُحركة لأنظمة الذكاء الاصطناعي (Biased data sets)، وتلك التحيزات بدورها تعمل على تعميق الفروق الاجتماعية الرقمية وأشكال عدم المساواة الرقمية بين مواطني المجتمع الرقمي[٥٤].

لكن تزداد جسامة التحيزات التي يُغذى بها الذكاء الاصطناعي – التحيزات الخوارزمية – عن التحيزات البشرية، نظرًا لأن المنصات تعالج معضلة التحيزات البشرية عن طريق إحالة المحتويات المُشتبَه فيها إلى فرق التنقيح البشري الموجودة في كل دولة على حدة، والتي تكون على دراية باللغة والواقع الاجتماعي والسياق الخاص بالمحتوى المنشور على المنصة، وبالتالي تقل تلك التحيزات لأقصى حد ممكن.

بينما لا يتحقق ذلك في حالة التنقيح الخوارزمي؛ حيث إنه يضع كافة السلطات بيد الذكاء الاصطناعي بداية من الضبط الإداري الرقمي وحتى المحاكمة واتخاذ القرار والعقاب الرقمي، ولا يمكن كشف تلك التحيزات بسهولة[٥٥].

فعلى سبيل المثال رصد المجتمع المدني تحيز الذكاء الاصطناعي تجاه الأقليات ووضع خطاب المستخدمين ذوي البشرة غير البيضاء وكأنه خطاب كراهية بشكل آلي، مما حدا بمنصة فيسبوك إلى إصلاح نظام الذكاء الاصطناعي المتعلق بخطاب الكراهية[٥٦].

وهكذا فبدلًا من أن يكون الذكاء الاصطناعي بمثابة محكم حيادي (neutral arbiter) أضحى عاملًا مؤثرًا في تعميق الفروق والتقسيمات الاجتماعية، عن طريق إضفاء أولوية على أنواع معينة من الخطاب تعكس التحيزات الاجتماعية (social bias) الكامنة في ذهن القائمين على صناعة قواعد البيانات المُغذية للذكاء الاصطناعي، مما يؤثر سلبًا على حماية الأقليات والقيم الديموقراطية في النظم الدستورية المختلفة[٥٧].

ويمكننا أن نذهب لأبعد من ذلك إلى الكود الخاص بصناعة المنصة ذاتها، حيث إن الكود عبارة عن التعبير التقني للقيم التي تعتنقها الشركات القائمة على تشغيل المنصة، ومن ثم يتأثر هيكل المنصة نفسه بالتحيزات والقيم والأولويات التي يؤمن بها القائمين على المنصة، ومن هنا بالتحديد نُدرك كيف تساهم تلك المنصات في زعزعة السلم والأمن الاجتماعي في الكثير من الدول التي تعتنق ثقافات مختلفة عن الثقافة الأمريكية لحرية التعبير[٥٨].

---

[٥٤] Katzenbach and Ulbricht, “Algorithmic Governance.” P ٧.

أنظر بشأن عدم المساواة الرقمية بين الجنسين؛ تقرير مفوض الأمم المتحدة السامي لحقوق الإنسان بشأن تعزيز وحماية حقوق الإنسان على الإنترنت والتمتع بها – السبل الكفيلة بسد الفجوة الرقمية بين الجنسين من منظور حقوق الإنسان، مقدم لمجلس حقوق الإنسان التابع للأمم المتحدة في الدورة (٣٥) بتاريخ ٥ مايو ٢٠١٧.

[٥٥] Dias Oliva, “Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression.” P ٦١٢ – ٦١٣ “Consequently, intermediaries, the ones behind the development of algorithms governing social media, exercise great power over users’ speech and access to information in the digital environment”.

[٥٦] Douek, “Content Moderation As Systems Thinking.” P ٥٦٩ – ٥٧٢; Sarah T. Roberts, Behind the Screen – Content Moderation in the Shadows of Social Media. P ٢٠٦ – ٢٠٨.

[٥٧] Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. P ١٨٢.

[٥٨] Ibid. P ١٧٩.

(٣) سلطة الذكاء الاصطناعي التنبؤي في إعادة خلق قيم المنصة:

ذكرنا سابقًا أن تطوير المنصات لأنظمة الذكاء الاصطناعي أفضى إلى ابتداع نظم للتنقيح قائمة على التنبؤ بمدى مشروعية المحتوى محل التنقيح، لاسيما إذا لم يكن متطابقًا مع أي بيانات مُغذَى بها النظام، ويعمل حينئذ على إعادة تغذية نفسه بالمعطيات الجدية. لكن تثور هنا معضلة بشأن قدرة الذكاء الاصطناعي على اتخاذ قرارات مستقلة وتطوير القيم ذاتها المُغذى بها، على نحو يفضي إلى ضبطه للسلوك الرقمي بما يتناسب مع القيم التي طورها، دون قيم المنصة ذاتها![٥٩] .

وتتضح مدى جسامة هذه المعضلة إذا ما أدركنا كيف يسيطر الذكاء الاصطناعي بشكل آلي مسبق على كافة المحتويات قبل نشرها على المنصات، وبالتالي يمكنه إساءة تفسير بيانات ومعلومات المستخدمين على نحو يفضي إلى إظهار بعض المحتويات غير المرغوب فيها أو التي تتماشي مع ميولهم وتحيزاتهم، والتي تفضي إلى اضطرابات جسيمة للسلم والأمن الاجتماعي، كأن تقترح أنظمة الذكاء الاصطناعي التنبؤية بعض الحسابات لأشخاص وجماعات متطرفة للمستخدم بمجرد ولوجه لإحداها، وهكذا تعمل على تعميق الفكر المتطرف للمستخدم عن طريق العرض المستمر لتلك الحسابات أمامه مما يؤثر على فكره، على نحو لم يكن في حسبان المستخدم، وهذا ما عليه الحال على منصة تويتر. ولذلك من الضرورة بمكان إحكام السيطرة على إمكانية تطوير الذكاء الاصطناعي لنفسه[٦٠].

(٤) عجز الذكاء الاصطناعي عن الإحاطة بسياق ممارسات حرية التعبير:

تبذل المنصات جهدًا كبيرًا في تطوير أنظمة الذكاء الاصطناعي حتى تكون قادرة على ضبط نشاط المستخدمين بفعالية، ولكن تظل معضلة تدريب تلك الأنظمة على مواجهة التضليل الذي يقوم به المستخدمين في طريقة عرض وكتابة المحتوى من أهم عوائق التنقيح الخوارزمي الفعال.

فكثيرًا ما يتم التلاعب بالسياق حتى يفلت المحتوى من التنقيح. لكن المعضلة لا تقف عند هذا الحد، فعدم إدراك نظم الذكاء الاصطناعي للسياق يؤدي بها في كثير من الأحوال إلى اتخاذ القرار الخاطئ إزاء المحتوى المنشور، وهنا إما أن يصنف الذكاء الاصطناعي المحتوى بأنه مشروع رغم أنه خضع للتضليل المعلوماتي من قبل المستخدمين (false negatives)، أو يُصنف المعلومات الصحيحة وكأنها غير مشروعة (false positives)[٦١].

---

[٥٩] Perez and Wimer, “Algorithmic Constitutionalism.” P ١٩ – ٢٢. “A shift to full AI governance raises two types of risks: potential increase in cases of over and under enforcement (false positives and false negatives) due to loss of human oversight and algorithmic bias, and potential misaligned, AI-driven revision of the system’s fundamental normative structure.”

[٦٠] Ibid. P ٤.

[٦١] Dumbrava, Key Social Media Risks to Democracy. P ٢٠. “Automated tools and algorithms promise to address the problems of scale and resources linked to moderation.”

ثانيًا: مقاربات الحوكمة الخوارزمية الرشيدة للمجال العام الرقمي

يتضح مما سبق أن المنصات أصبحت طرفًا في العقد الاجتماعي الرابط بين الدولة والأفراد؛ بحيث أصبحت تحدد هيكل وشروط ممارسة الحقوق الأساسية في العالم الرقمي بشكل يؤثر على مجريات الأحداث في العالم الحقيقي. واعتبارها طرفًا في العقد الاجتماعي، أفضى إلى ضرورة تلبيتها واجبات السلطة التنفيذية في المجتمع الرقمي بتنقيح المحتوى بشكل مسبق عن طريق الذكاء الاصطناعي أو بشكل لاحق عن طريق التنقيح البشري.

ولكن – كما ذكرنا سابقًا – تطوير المنصات لأنظمة الذكاء الاصطناعي لم يكن حلًا كافيًا لضبط آثار إساءة ممارسة حرية التعبير في المجتمع الرقمي، مما حدا بالكثير من الفقه إلى إعادة تصور العلاقة التي تربط الدولة بالمنصات في العقد الاجتماعي الرقمي، وذلك عن طريق التأثير على سياسات بناء أنظمة التنقيح الخوارزمي، ومن هنا نشأ ما يُطلق عليه الفقه الحوكمة المشتركة بين الدولة والمنصات للسلوك البشري[٦٢].

سنتعرض لعدد من المقاربات التي تمثل اتجاهات مختلفة للحوكمة الرشيدة لاستخدام المنصات للذكاء الاصطناعي، أو بصورة أدق، مقاربات الحوكمة الخوارزمية الرشيدة للمجال العام الرقمي.

الاتجاه الأول: أخلاقيات الذكاء الاصطناعي (Ethics of AI)

في بداية الأمر نادى البعض بتطبيق قيم ومبادئ الهندسة الأخلاقية (ethical engineering) لضبط عملية تطبيق الذكاء الاصطناعي في ضبط السلوك البشري الرقمي، ولكن أخلاقيات الذكاء الاصطناعي في حد ذاتها غير كافية لأنها لن تحل معضلة انتقال التحيزات الفردية من الأفراد إلى قواعد البيانات الخاصة بالذكاء الاصطناعي، ومن ثم لن يأخذ الذكاء الاصطناعي في اعتباره تعددية القيم على منصات التواصل الاجتماعي، وسوف يُغلب إحداها على سائر القيم التي تعتنقها ثقافات أخرى مغايرة للثقافة الأمريكية[٦٣].

ورغم أن قيم الهندسة الأخلاقية تعد من قبيل القوانين المرنة التي تلزم بها الشركات التكنولوجية نفسها، أو يلزمها بها المجتمع المدني، إلا إنه لا ينبغي استبدال القانون بمدونات أخلاقية لضبط صناعة واستخدام الذكاء الاصطناعي، فيجب تحديد أي التحديات الخاصة بالذكاء الاصطناعي يمكن تركها للأخلاقيات وأيها يلزم تنظيمه قانونًا.

الاتجاه الثاني: الضبط الدستوري الخوارزمي (Algorithmic Constitutionalism)

[٦٢] Niva Elkin-Koren, "Government-Platform Synergy and Its Perils," in Constitutionalising Social Media (Hart Publishing, ٢٠٢٢). P ١٨١.

[٦٣] Perez and Wimer, "Algorithmic Constitutionalism." P ٢.

ذهب جانب آخر من الفقه إلى المناداة بتطبيق الضبط الدستوري الخوارزمي على صناعة أنظمة الذكاء الاصطناعي القائمة على التنقيح[٦٤].

تعتمد تلك الفكرة على التفرقة بين مستويين من مستويات نظام الذكاء الاصطناعي (layered architecture) وقدرة النظام إزاء كل مستوى؛

(١) المستوى الأول هو الكود الأساسي (metalevel code)؛ يتضمن القيم الأساسية التي يستهدي بها نظام التنقيح الخوارزمي، تلك القيم عبارة عن تعبير حسابي عن أحكام وشروط استخدام المنصة بشأن ما يشكل تعبيرًا مشروعًا عن الرأي. ينبغي أن يتم الحيلولة دون استطاعة الذكاء الاصطناعي التأثير في الكود الأساسي للنظام الخاص به عن طريق إعادة تفسير القيم الأساسية أو إضافة قيم أخرى لها، وبذلك يتم ضبط تأثير الذكاء الاصطناعي فيما يتعلق بالكود الأساسي (algorithmic meta-reasoning). هذا المستوى بمثابة المستوى الدستوري لنظام الذكاء الاصطناعي، ولذلك يُسمى بالدستور الخوارزمي (algorithmic constitution).

(٢) المستوى الثاني العملي من الكود (operative level)؛ وهو الذي يوجه النظام على اتخاذ قرارات مستقلة بشأن المحتوى محل التنقيح في ضوء القيم الأساسية الواردة في الكود الرئيسي، وذلك في العلميات اليومية. في ذلك المستوى يملك الذكاء الاصطناعي القدرة على مراجعة قراراته في ضوء ما يتم تغذيته به من بيانات وعلى أساس القيم الموجودة في المستوى الأساسي، ومن ثم يتمتع بقدر كبير من السلطة التقديرية واسعة في الرقابة على المحتوى (object level algorithm).

ولكن ما إن يصطدم النظام بمحتويات جديدة يُستشكَل أمرها عليه، فإنه يجب صناعة النظام بحيث يعود في كل مرة إلى القائم على التنقيح البشري، وحينها إما أن تُضاف معلومات جديدة لقواعد البيانات أو يتم إعادة ضبط القيم الواردة في الكود الأساسي. ومن ثم لا يمكن للذكاء الاصطناعي أن يبدأ إصلاحات في الكود الأساسي من تلقاء نفسه كأن يلغي القيود الواردة على حرية الراي والتعبير بشأن خطاب العنف والدعاوى الإرهابية.

الاتجاه الثالث: الحقوق الدستورية الخوارزمية

بالإضافة للاتجاهات التي توجه بصفة رئيسية للقائمين على صناعة أنظمة الذكاء الاصطناعي، هناك مقاربات أخرى تعمل على ابتداع طائفة من الحقوق الأساسية للأفراد في مواجهة الذكاء الاصطناعي، ويستوي هنا أن يكون الحديث موجه لمطوري الذكاء الاصطناعي أو لصناع السياسات والمجتمع المدني.

الحقوق الدستورية الخوارزمية مستمدة في جوهرها من الحقوق الأساسية الراسخة في النظرية الدستورية منذ القرن الثامن عشر، وهي تشكل طائفة واحدة فقط من مجموعة كبيرة من الحقوق الدستورية للعصر الرقمي والتي درج الفقه على تسميتها إعلانات الحقوق لأجل الإنترنت (Internet Bills of Rights)[٦٥].

---

[٦٤] Ibid. P ٢٢ – ٣١.

[٦٥] درج الفقه الدستوري المقارن على تسليط الضوء على الطابع الدستوري لوثائق الحقوق لأجل الإنترنت؛ أنظر في ذلك المراجع الآتية؛

تعد إعلانات الحقوق عند بعض الفقه مثل (Gill) و(Gasser) و(Redeker) بمثابة مبادئ حماية الحقوق الأساسية في المجتمع الرقمي، وتطورها يعكس التطور الحادث في نظرية السلطة العامة، ففي بداية الأمر كانت الدولة محل الخوف على الإنترنت، وقد عبرت أولى وثائق الحقوق عن تلك النظرة بالتشديد على حرية الراي والخصوصية وعدم المراقبة في مواجهة السلطات الوطنية، ولكن بمرور الوقت لم تصبح الدولة المصدر الوحيد للتهديد على الإنترنت، ولم يعد التساؤل يدور حول الحوكمة العامة أو الخاصة للإنترنت بقدر ما أصبح الحديث يتمحور حول حماية الحقوق الأساسية للأفراد في كل الأحوال، و بناء مستقبل رقمي شامل للجميع[٦٦].

هذا التطور لنظرية الحرية في الفكر الدستوري أمر يقتضيه طبيعة التحولات التي تطرأ على السلطة وممارستها، وقد اعترفت المحكمة الدستورية العليا المصرية بتطور فهم النصوص الدستورية حتى تتفق مع حقائق العصر[٦٧]، وذهب الرئيس الأسبق للمحكمة ذاتها، إلى أنه لا يمكن حصر الحقوق الأساسية في قائمة بعينها، بل تتطور الحقوق والحريات بتطور التجربة البشرية، وتقدم المجتمع[٦٨]، وهكذا فإن الحقوق الأساسية تتطور لتلائم طبيعة العلاقة بين السلطة والحرية في مجتمع المنصات، لاسيما إزاء أنظمة الذكاء الاصطناعي، بل والأمر يتعدى ذلك لتطور المؤسسات والآليات الدستورية ذاتها مع تطور العقل البشري، وما أفرزه من إبداعات تؤثر في تطور المجتمع مما يحتم تطور المؤسسات استجابة لذلك.

من أمثلة تلك الحقوق؛

أولًا: الحق في الحياة في المجتمع الرقمي:

من المستقر عليه أن المادة السادسة من العهد الدولي للحقوق المدنية والسياسية بشأن الحق في الحياة تفرض على الدولة التزامين؛ التزام سلبي وهو أن تمتنع الدولة عن الاعتداء على حياة الفرد، والتزام إيجابي يتمثل في تبني الدولة للسياسات

Celeste, Digital Constitutionalism – The Role of Internet Bills of Rights. p ١٥٧ – ١٥٩. “Internet bills of rights aim to craft – one could say – a new, enlarged social contract”. Lex Gill, Dennis Redeker, and Urs Gasser, “Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights,” Berkman Center Research Publication No. ٢٠١٥–١٥, ٢٠١٥. P ٣, P ١٠ – ١١; Edoardo Celeste, “Digital Constitutionalism: A New Systematic Theorisation,” International Review of Law, Computers & Technology ٣٣, no. ١ (٢٠١٩). P ١٠.

[٦٦] Gill, Redeker, and Gasser, “Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights.” P ١٨; Celeste, Digital Constitutionalism – The Role of Internet Bills of Rights. p ٤٦ “These documents manifestly present constitutional features.”; Ibid. p ١٢١. “They deal with rights and the balancing of power: the two cardinal missions of constitutional law.”

[٦٧] حكم المحكمة الدستورية العليا في الدعوى رقم ٢٣ لسنة ١٦ ق "دستورية" بتاريخ ١٨ مارس سنة ١٩٩٥م.

[٦٨] المستشار عوض المر، الرقابة على دستورية القوانين في ملامحها الرئيسية، مركز رينيه – جان ديوي للقانون والتنمية، بدون تاريخ نشر. ص ١٣٤٧، تقرير مجلس حقوق الإنسان التابع للأمم المتحدة بشأن تعزيز وحماية حقوق الإنسان على الإنترنت والتمتع بها، الدورة الثانية والثلاثون، بتاريخ ٢٧ يونيو ٢٠١٦.

والإجراءات اللازمة لكفالة مقتضيات ومتطلبات الحق في الحياة (survival requirements)، وهو ما عبر عنه المستشار عوض المر بأن الحق في الحياة يشمل كافة الشروط الضرورية لكفالة الحياة اللائقة[٦٩].

وفي نطاق المجتمع الرقمي؛ فإن حماية الحق في الحياة تقتضي؛ (١) أن تلتزم المنصة – بوصفها سلطة عامة رقمية – بإعمال قواعد تنقيح المحتوى بشكل يحول دون المساس بحق الأفراد في الحياة في العالم الحقيقي، نتيجة سلوكيات نابعة من المجتمع الرقمي. (٢) تلتزم المنصة بأن تتدخل بشكل إيجابي في تهيئة بيئة آمنة للاجتماع الرقمي في المجال العام الرقمي، وذلك يقتضي تطوير الذكاء الاصطناعي لكي يؤدي مهامه بفعالية لكن مع وضع الضوابط التي تحول دون حدوث أوجه القصور التي تعرضنا لها سابقًا.

ثانيًا: حرية التعبير في المجتمع الرقمي:

عادة ما تلجأ إعلانات الحقوق لأجل الإنترنت للعهد الدولي للحقوق المدنية والسياسية للحديث عن ضمانات حرية التعبير وقيودها، لاسيما أن العهد الدولي لا يفرق بين أي وسيلة من وسائل التعبير عن الرأي[٧٠]، وبالتالي تمتد حماية العهد الدولي لمظاهر التعبير على منصات التواصل الاجتماعي.

حماية حرية التعبير على المنصات تقتضي إتاحة الفرصة للأفراد في فهم كيفية اتخاذ الذكاء الاصطناعي للقرارات بشأن منشوراتهم على المنصات، باعتبارها من مظاهر ممارسة حرية التعبير التي لا يجب المساس بها إلا إذا كانت تصطدم بمصالح أخرى أجدر بالرعاية، وهذا ما يُطلق عليه الفقه الحق في تفسير تقنيات الذكاء الاصطناعي[٧١].

كما أن بعض إعلانات الحقوق لأجل الإنترنت تنص على الحق في المراجعة البشرية للقرارات التي يتخذها الذكاء الاصطناعي، خاصة إذا ما كان من شأنها تعريض حياة الفرد لخطر حقيقي في العالم الواقعي، والبعض الأخر منها ينص على عدم خضوع الأفراد لقرارات مبنية على المعالجة الآلية بتقنيات الذكاء الاصطناعي، وهو الحق في عدم المعالجة الآلية

---

[٦٩] المستشار عوض المر، الرقابة على دستورية القوانين في ملامحها الرئيسية. ص ١٣٢٣، أنظر كذلك حكم المحكمة الدستورية العليا في الدعوى رقم ٣٤ لسنة ١٥ق "دستورية"، جلسة ٢ / ٣ / ١٩٩٦.

[٧٠] Sarah Joseph and Melissa Castan, The International Covenant on Civil and Political Rights – Cases, Materials, and Commentary (Oxford University Press, ٢٠١٣). P ٥٩٠ – ٥٩٣. P ٥٩٦ – ٥٩٨.

تعد حرية الرأي والتعبير على الإنترنت المدخل الطبيعي لسائر الحقوق والحريات الرقمية كحرية التجمع، أنظر في ذلك؛ تقرير المقرر الخاص المعني بتعزيز وحماية الحق في حرية الرأي والتعبير، مُقدم لمجلس حقوق الإنسان التابع للأمم المتحدة، الدورة (٣٨)، يوليو ٢٠١٨. وكذلك تقرير المقرر الخاص المعني بالحق في حرية التجمع السلمي وتكوين الجمعيات – بشأن الحق في حرية التجمع السلمي وتكوين الجمعيات في العصر الرقمي، مقدم لمجلس حقوق الإنسان التابع للأمم المتحدة في الدورة (٤١) بتاريخ ١٢ يوليو ٢٠١٩م.

[٧١] Oreste Pollicino and Giovanni De Gregorio, “Constitutional Law in the Algorithmic Society,” in Constitutional Challenges in the Algorithmic Society (Cambridge University Press, ٢٠٢١). p ٢١; Paul Nemitz, “Constitutional Democracy and Technology in the Age of Artificial Intelligence,” ٢٠١٨. P ٧ – ٨.

للقرارات التي تنتج آثارًا قانونية تؤثر على الفرد (right against processing)، وهو حق منصوص عليه في الفقرة الأولى من المادة ٢٢ من قانون حماية البيانات الأوروبي (GDPR)[٧٢].

وبعض الفقه الأمريكي يعبر عن هذا الحق بمصطلح الحق في القرار البشري (Right to a Human Decision)[٧٣]، بإعتبار أنه يجب اتباع إجراءات عادلة في إصدار القرارات الماسة بالحقوق الأساسية للأفراد، وهذا يقتضي أن يكون الفرد على بينة بالقرار الصادر في حقه، والأسس التي أعتمد عليها الذكاء الاصطناعي في إصدار القرار[٧٤].

الخاتمة

تعرضنا في البحث الماثل لتطور منصات التواصل الاجتماعي والمجتمع الرقمي في النظرية الدستورية المعاصرة، حتى أضحت تُعتبر طرفًا في العقد الاجتماعي الذي يربط بين الدولة والفرد. ورأينا كيف طورت المنصات سياسات وآليات لا تختلف كثيرًا عن تلك المتعارف عليها في النظرية الدستورية التقليدية بشأن دور الدولة في توجيه السلوك الاجتماعي والضبط القانوني للنظام الاجتماعي.

ونظرًا لأن حجم البيانات على المنصات هائل لدرجة تجعل تلك المنصات أيسر الطرق لزعزعة السلم والنظام الاجتماعي، دأبت المنصات على تطوير تقنيات الذكاء الاصطناعي لضبط حرية التعبير في المجال العام الرقمي، وهو ما أطلقنا عليه الحوكمة الخوارزمية[٧٥]، ولكن لم تخل تلك الحوكمة من انتقادات نتيجة الثقة الزائدة عن الحد بشأن قدرة الآلة والتكنولوجيا في العصر الحديث[٧٦]. لكن رغم الانتقادات التي وجهت لتطبيقات التنقيح الخوارزمي على المنصات، فإنها خطوة جوهرية لخلق مجتمعات رقمية آمنة، في عصر بدأت تندثر فيه الحدود الفاصلة بين المجال العام الرقمي على الإنترنت وبين المجال العام الحقيقي في المجتمع.

كل هذا يُحتم أن يتوجه الفقه الدستوري بآلياتهم التحليلية لدراسة منصات التواصل الاجتماعي وتحليل أدوارها في المجتمع المعاصر، حتى يمكن فهم كيفية ضبط تأثيرها، ويعد هذا البحث مجرد محاولة أولية في هذا الطريق.

---

[٧٢] Aziz Z. Huq, “A Right to a Human Decision,” Virginia Law Review ١٠٦, no. ٣ (٢٠٢٠). P ٦٢٢.

يمكن للقضاء الوطني بسط رقابته على قرارات التنقيح الخوارزمي إذا ما تسببت في انتهاك حرية التعبير لُمستخدم منصة التواصل الاجتماعي، وذلك قياسًا على الإتجاه المتنامي في الفقه المعاصر بشأن مسئولية الذكاء الاصطناعي. أنظر بشأن الرقابة على القرارات الإدارية الصادرة استنادًا للذكاء الاصطناعي في المجتمع الحقيقي؛ د محمد محمد عبد اللطيف، المسئولية عن الذكاء الاصطناعي بين القانون الخاص والقانون العام، بحث مقدم إلى مؤتمر الجوانب القانونية والاقتصادية للذكاء الاصطناعي وتكنولوجيا المعلومات، كلية الحقوق – جامعة المنصورة، ٢٣ – ٢٤ مايو ٢٠٢١م. ص ٣٢ وما بعدها.

[٧٣] Ibid. P ٦٢٤ – ٦٢٧.

[٧٤] لا يمكن القول بأن هناك حقًا جديدًا ولد في النظام القانوني الأوروبي أو الأمريكي بشأن المراجعة البشرية على عمل الآلة، ولكنه يتبادر لأذهان المشرعين والفقه القانوني. أنظر في ذلك؛

Ibid. P ٦٢٨; P ٦٨٦.

[٧٥] Consultants, “The Use of AI in Content Moderation.” P ١٣.

[٧٦] Sheng, “Automated Content Moderation.” P ٣٦١; Bloch-Wehba, “Automation in Moderation.” P ٨١.

قائمة المراجع

المراجع العربية:

١. المستشار عوض المر، الرقابة على دستورية القوانين في ملامحها الرئيسية، مركز رينيه – جان ديوي للقانون والتنمية، بدون تاريخ نشر.
٢. د محمد محمد عبد اللطيف، المسئولية عن الذكاء الاصطناعي بين القانون الخاص والقانون العام، بحث مقدم إلى مؤتمر الجوانب القانونية والاقتصادية للذكاء الاصطناعي وتكنولوجيا المعلومات، كلية الحقوق – جامعة المنصورة، ٢٣ – ٢٤ مايو ٢٠٢١م.
٣. د رنا مصطفى أحمد فؤاد، المسئولية الجنائية عن خطاب الكراهية عبر الإنترنت أمام المحكمة الجنائية الدولية: دراسة تحليلية وتطبيقية على خطاب الكراهية ضد مسلمي الروهينجا في ميانمار، مجلة القانون والتكنولوجيا، المجلد الثالث، العدد الأول، أبريل ٢٠٢٣.
٤. د محمود حسن السحلي، أساس المسائلة المدنية للذكاء الاصطناعي المستقل: قوالب تقليدية أم رؤية جديدة؟، مجلة كلية الحقوق للبحوث القانونية والاقتصادية، العدد الثاني، ٢٠٢٢.
٥. أروى محمد تقوى، مسئولية مواقع الويب ٢,٠ عن المحتوى غير المشروع في أوروبا وفرنسا والولايات المتحدة، مجلة الحقوق، مجلد ٤٠، العدد الأول، ٢٠١٦.
٦. كيت أورتون جونسون ونيك بريور، علم الاجتماع الرقمي – منظورات نقدية، ترجمة هاني خميس، المجلس الوطني للثقافة والفنون والآداب – الكويت، ٢٠٢١.
٧. ألكسندر ميلياغرو هيسشنز ونيك كادربهاي، التطرف على الإنترنت، ترجمة محمد عوض يوسف، مكتبة الإسكندرية، ٢٠٢٠.
٨. لوتشيانو فلوريدي، الثورة الرابعة، ترجمة لؤي عبد المجيد السيد، المجلس الوطني للثقافة والفنون والآداب – الكويت، ٢٠١٧.
٩. تقرير التنمية البشرية في مصر ٢٠٢١، التنمية حق للجميع: مصر المسيرة والمسار.
١٠. قرار المجلس الاقتصادي والاجتماعي في الأمم المتحدة رقم (٤/٢٠٢٣) بشأن تسخير العلم والتكنولوجيا والابتكار لأغراض التنمية، دورة عام ٢٠٢٣، بتاريخ ٧ يونية ٢٠٢٣.
١١. تقرير مفوضية الأمم المتحدة السامية لحقوق الإنسان بشأن حجب الإنترنت: الاتجاهات والأسباب والتداعيات القانونية والتأثيرات التي تمس بمجموعة حقوق الإنسان، مُقدم لمجلس حقوق الإنسان، الدورة (٥٠)، بتاريخ ١٣ مايو ٢٠٢٢.
١٢. قرار الجمعية العامة للأمم المتحدة رقم (١٥٠/٧٧) بشأن تسخير تكنولوجيات المعلومات والاتصالات لأغراض التنمية المستدامة، الدورة (٧٧)، بتاريخ ١٤ ديسمبر ٢٠٢٢.
١٣. تقرير المقرر الخاص بتعزيز وحماية الحق في حرية الرأي والتعبير بشأن الأمراض الجائحة وحرية الرأي والتعبير، مقدم لمجلس حقوق الإنسان بالأمم المتحدة، الدورة (٤٤)، بتاريخ ٢٣ أبريل ٢٠٢٠.

١٤. تقرير المقرر الخاص المعني بالحق في حرية التجمع السلمي وتكوين الجمعيات بشأن الحق في حرية التجمع السلمي وتكوين الجمعيات في العصر الرقمي، مقدم لمجلس حقوق الإنسان التابع للأمم المتحدة في الدورة (٤١) بتاريخ ١٢ يوليو ٢٠١٩م.

١٥. تقرير المقرر الخاص المعني بتعزيز وحماية الحق في حرية الرأي والتعبير، مُقدم لمجلس حقوق الإنسان التابع للأمم المتحدة، الدورة الثامنة والثلاثون، يوليو ٢٠١٨.

١٦. تقرير مفوض الأمم المتحدة السامي لحقوق الإنسان بشأن تعزيز وحماية حقوق الإنسان على الإنترنت والتمتع بها – السبل الكفيلة بسد الفجوة الرقمية بين الجنسين من منظور حقوق الإنسان، مقدم لمجلس حقوق الإنسان التابع للأمم المتحدة في الدورة (٣٥) بتاريخ ٥ مايو ٢٠١٧.

١٧. تقرير مجلس حقوق الإنسان التابع للأمم المتحدة بشأن تعزيز وحماية حقوق الإنسان على الإنترنت والتمتع بها، الدورة الثانية والثلاثون، بتاريخ ٢٧ يونيو ٢٠١٦.

المراجع الإنجليزية:

١٨. Belli, Luca, and M Horten. “Structural Power as a Critical Element of Social Media Platforms’ Private Sovereignty.” In *Constitutionalising Social Media*. Hart Publishing, ٢٠٢٢.

١٩. Bloch-Wehba, Hannah. “Automation in Moderation.” *Cornell International Law Journal* ٥٣ (٢٠٢٠).

٢٠. Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press, ٢٠١٨.

٢١. Celeste, Edoardo. “Digital Constitutionalism: A New Systematic Theorisation.” *International Review of Law, Computers & Technology* ٣٣, no. ١ (٢٠١٩).

٢٢. ———. *Digital Constitutionalism – The Role of Internet Bills of Rights*. Routledge, ٢٠٢٣.

٢٣. Chowdhury, Nafia. “Automated Content Moderation: A Primer,” ٢٠٢٢.

٢٤. Cobbe, Jennifer. “Algorithmic Censorship by Social Platforms: Power and Resistance.” *Philosophy & Technology* ٣٤, no. ٤ (December ٧, ٢٠٢١).

٢٥. Cohen, Julie E. *Between Truth and Power – The Legal Constructions of Informational Capitalism*. Oxford University Press, ٢٠١٩.

٢٦. Consultants, Cambridge. “The Use of AI in Content Moderation,” ٢٠١٩.

٢٧. Dias Oliva, Thiago. “Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression.” *Human Rights Law Review* ٢٠, no. ٤ (December ٩, ٢٠٢٠).

٢٨. Douek, Evelyn. “Content Moderation As Systems Thinking.” *Harvard Law Review* ١٣٦, no. ٢ (٢٠٢٢).

٢٩. Dumbrava, Costica. *Key Social Media Risks to Democracy*. European Parliamentary Research Service (European Parliament), ٢٠٢١.

٣٠. Elkin-Koren, Niva. “Government–Platform Synergy and Its Perils.” In *Constitutionalising Social Media*. Hart Publishing, ٢٠٢٢.

٣١. Gill, Lex, Dennis Redeker, and Urs Gasser. “Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights.” Berkman Center Research Publication No. ٢٠١٥-١٥, ٢٠١٥.

٣٢. Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, ٢٠٢١.

٣٣. Goldman, Eric. “Content Moderation Remedies.” *Michigan Technology Law Review* ٢٨, no. ١ (٢٠٢١).

٣٤. Gregorio, Giovanni De. “Digital Constitutionalism across the Atlantic.” *Global Constitutionalism*, ٢٠٢٢.

٣٥. ———. *Digital Constitutionalism in Europe - Reframing Rights and Powers in the Algorithmic Society*. Cambridge University Press, ٢٠٢٢.

٣٦. Giovanni Sartor and Andrea Loreggia, “The Impact of Algorithms for Online Content Filtering or Moderation,” Policy Department for Citizens’ Rights and Constitutional Affairs, ٢٠٢٠.

٣٧. Huq, Aziz Z. “A Right to a Human Decision.” *Virginia Law Review* ١٠٦, no. ٣ (٢٠٢٠).

٣٨. Joseph, Sarah, and Melissa Castan. *The International Covenant on Civil and Political Rights - Cases, Materials, and Commentary*. Oxford University Press, ٢٠١٣.

٣٩. Kadri, Thomas E. “Digital Gatekeepers.” *Texas Law Review* ٩٩, no. ٥ (٢٠٢١).

٤٠. Kapczynski, Amy. “The Law of Informational Capitalism.” *Yale Law Journal* ١٢٩, no. ٥ (٢٠٢٠).

٤١. Katzenbach, Christian, and Lena Ulbricht. “Algorithmic Governance.” *Internet Policy Review* ٨, no. ٤ (November ٢٩, ٢٠١٩).

٤٢. Keller, Clara Iglesias. “The Perks of Co-Regulation: An Institutional Arrangement for Social Media Regulation?” In *Constitutionalising Social Media*. Hart Publishing, ٢٠٢٢.

٤٣. Klonick, Kate. “The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression.” *The Yale Law Journal* ١٢٩, no. ٨ (٢٠٢٠).

٤٤. Koltay, András. *New Media and Freedom of Expression – Rethinking the Constitutional Foundations of the Public Sphere*. Hart Publishing, ٢٠١٩.

٤٥. Lee, Edward. “Virtual Governments.” *UCLA Journal of Law & Technology* ٢٧, no. ٢ (٢٠٢٢).

٤٦. Marecos, João, Ethan Shattock, Oliver Bartlett, Francisco Goiana-da-Silva, Hendramoorthy Maheswaran, Hutan Ashrafian, and Ara Darzi. “Health Misinformation and Freedom of Expression: Considerations for Policymakers.” *Health Economics, Policy and Law* ١٨, no. ٢ (April ٢٠٢٣).

٤٧. Menyhárd, Attila. “The Impacts of Technological Revolution on the Role of the State.” In *The IT Revolution and Its Impact on State, Constitutionalism and Public Law*. Bloomsbury Publishing, ٢٠٢١.

٤٨. Mickoleit, Arthur. “Social Media Use by Governments: A Policy Primer to Discuss Trends, Identify Policy Opportunities and Guide Decision Makers.” OECD Working Papers on Public Governance No. ٢٦, ٢٠١٤.

٤٩. Nemitz, Paul. “Constitutional Democracy and Technology in the Age of Artificial Intelligence,” ٢٠١٨.

٥٠. O’Regan, Kate. “Public Law, the Digital World and Human Rights: Challenges Ahead.” *Judicial Review*, ٢٠٢٠.

٥١. OECD. *An Introduction to Online Platforms and Their Role in the Digital Transformation*. OECD Publishing, Paris, ٢٠١٩.

٥٢. Pałka, Przemysław. “Terms of Injustice.” *SSRN Electronic Journal*, ٢٠٢٣.

٥٣. Perez, Oren, and Nurit Wimer. “Algorithmic Constitutionalism.” *Indiana Journal of Global Legal Studies*, ٢٠٢٣.

٥٤. Pollicino, Oreste, and Giovanni De Gregorio. “Constitutional Law in the Algorithmic Society.” In *Constitutional Challenges in the Algorithmic Society*. Cambridge University Press, ٢٠٢١.

٥٥. Risse, Mathias. “Artificial Intelligence and the Past, Present, and Future of Democracy,” ٢٠٢١.

٥٦. Sarah T. Roberts. *Behind the Screen – Content Moderation in the Shadows of Social Media*. Yale University Press, ٢٠١٩.

٥٧. Schulz, Wolfgang. “Changing the Normative Order of Social Media from Within: Supervisory Bodies.” In *Constitutionalising Social Media*. Hart Publishing, ٢٠٢٢.

٥٨. Sheng, Chuyi. “Automated Content Moderation.” *Georgetown Law Technology Review* ٦, no. ١ (٢٠٢٢).

٥٩. Stephan, Paul B. *The World Crisis and International Law – The Knowledge Economy and the Battle for the Future*. Cambridge University Press, ٢٠٢٣.

٦٠. Valchev, Daniel. “Constitutional Dimensions of Information Revolution.” In *The IT Revolution and Its Impact on State, Constitutionalism and Public Law*. Bloomsbury Publishing, ٢٠٢١.

٦١. Venturini, Jamila, Luiza Louzada, Marília Ferreira Maciel, Nicolo Zingales, Konstantinos Stylianou, and Luca Belli. *Terms of Service and Human Rights*. Editora Revan, ٢٠١٦.

٦٢. Wilman, Folkert. *The Responsibility of Online Intermediaries for Illegal User Content in the EU and the US*. Edward Elgar Publishing, ٢٠٢٠.