

الأسرار المظلمة لقواعد فيسبوك حول خطاب الكراهية والمحتوى المسيء



ترجمة وتحرير: نون بوست

يتعرض موقع فيسبوك للنقد بسبب افتقاره للشفافية، وتوخيهِ الاعتدال على حد السواء. ففي شهر آذار/مارس، ألقى مرزوقي داروسمان، الذي يعمل مع الأمم المتحدة في بعثة لتقصي الحقائق في مجال حقوق الإنسان، باللوم على موقع التواصل الاجتماعي للعبة "دورا قياديا" في نشر رسائل الكراهية ضد الروهينغا، وهي جماعة عرقية مسلمة في بورما. ويستخدم الموقع على نطاق واسع في هذا البلد (أكثر من 30 مليون مستخدم في بلد يقطنه 53 مليون نسمة) ليصبح بمثابة قناة صرف حقيقية للإسلاموفوبيا.

قبل بضعة أشهر، أثار بول لوغان، وهو من الأشخاص المشهورة على موقع يوتيوب ضجة عند تصويره فيديو وهو يقف إلى جانب جثة في "غابة الانتحار" اليابانية المعروفة. وقد شوهد الفيديو عدة ملايين من المرات قبل أن يُمحى من قبل صاحبه، دون أن يتم حذفه من قبل يوتيوب. فبالنسبة للذكاء الاصطناعي، لم يكن وجود جثة في مثل هذا المكان أمراً شاذاً. وكان من الأخلاق، وهي مفهوم لا يمكن إدراكه إلا من قبل الإنسان، الإقرار بعكس ذلك.

لكن، أين تبدأ الرقابة وأين ينتهي الاعتدال؟ لدى الفنانة الأمريكية إيلا غور فكرة حول هذا الموضوع، وهي التي رأت أن حسابها على فيسبوك قد تم قمعه بوحشية، بسبب صورة شاركتها تجسد دونالد ترامب وهو بقضيب ذكري صغير، التي لاقت رواجاً منذ ذلك الوقت. أما في فرنسا، فتعود بنا الذاكرة إلى الاضطراب الذي أثير في سنة 2011 بسبب الرقابة التي فرضت على لوحة "أصل العالم" لغوستاف كوربيه على موقع التواصل الاجتماعي.

يتم التعامل مع رسائل الكراهية والتعليقات المسيئة وفقاً لمعادلة أساسية وهي: "الفئة المحمية + هجوم = خطاب كراهية"

في أغلب المنشورات التي وقع تجريمها، يعد الهدف هو تقييم السياق، وهو ما كشفت عنه كل من

صحيفة "زود دويتشي تسایتونج" الألمانية و"الغارديان" البريطانية في سنة 2016 و2017 من خلال نشر إرشادات الإشراف على "المحتوى المعتدل" على موقع فيسبوك للجمهور. ف فيما يتعلق بالتعري على سبيل المثال، هناك قاعدة أساسية وهي أنه في حال كان المحتوى قد "صنع يدويًا"، فيسمح بمشاركة، أما إذا كان رقميا، فهناك فرصة كبيرة لأن يصنف كمحتوى إباحي.

من جانب آخر، يتم التعامل مع رسائل الكراهية والتعليقات المسيئة وفقًا لمعادلة أساسية وهي: "الفئة المحمية + هجوم = خطاب كراهية". فإذا كان الشخص المهاجم ينتمي إلى فئة محمية، يتم تصنيف المنشور على أنه خطاب يحرض على الكراهية. وتشمل هذه الفئات جميع التوجهات الجنسية، والجماعات الدينية، والأصول العرقية، أو الجنسيات. وعلى سبيل المثال، بالإمكان انتقاد أي دين، ولكن ليس لدينا الحق في مهاجمة أصحاب المعتقد. فأن تصف "الكاثوليكية أو الإسلام أو البوذية" بالهراء، هو أمر مسموح به، أما أن يتم وصف الكاثوليكين أو المسلمين أو البوذيين بالأغبياء فهذا غير مسموح.

في الوثيقة التي نشرتها صحيفة "زود دويتشي تسایتونج" الألمانية، لا تخلو بعض الأمثلة من السخرية. فإذا كنت ممنوعًا من كتابة "جميع الفرنسيين مدمنون على الكحول"، يمكنك دون صعوبة كتابة: "لا تثق في الأولاد الذين يقولون لك إنهم يحبونك" أو "السيقان الكبيرة غريبة جدًا"، لأن الفتيان والفتيات لا ينتمون إلى أي فئة محمية. كما يخضع المهاجرون واللاجئون إلى ما يمكن اعتباره "شبه حماية"، إذ يسمح بوصفهم "بطلبي لجوء في الخارج"، ولا يسمح بوصفهم "بالمهاجرين المقرفين". وينطبق الأمر نفسه على اللعب بالمفردات على غرار "اغتصاب - اللاجئ".

نحو بداية الشفافية

يمكننا أن نرى كيف يكون الاعتدال فنا دقيقا عندما يتعلق الأمر بالحكم على سياق أو قصد محدد. وعلى الرغم من أن إهانة المبولات الجنسية أمر محظور، إلا أن قول "يعتبرني الناس مغنية مشهورة، ولكنني شاب مخنت" لا يمثل أية مشكلة، لأن المحتوى المراد إبلاغه يعتبر ساخرًا. وبالمثل، لن يتم إزالة التهديد "العام أو غير المقنع" مثل "اللجنة عليك، أمل أن تقتل".

من بين المواضيع الحساسة الأخرى أنه لا يتم السماح بالمنشورات المتعلقة بالعنف ضد الحيوانات إلا إذا تم استخدامها للشجب والاستنكار

أما الأمر الأكثر إثارة للدهشة فيتمثل في أن عبارة "ضرب أشقر" مقبولة بينما لا يُقبل عبارة "أقتل ترامب"، بصفته رئيس دولة وينتمي للفئة المحمية. وفيما يتعلق بالتحرش، تخضع جميع المنشورات التي تضم برازا أو قيئا أو بولا للرقابة، إلا إذا كنت من المشاهير. وسُعتبر شخصية عامة، ولن تكون جزءًا من فئة الأشخاص العاديين، إن كان لديك أكثر من 100 ألف متابع.

علاوة على ذلك، ومن بين المواضيع الحساسة الأخرى أنه لا يتم السماح بالمنشورات المتعلقة بالعنف ضد الحيوانات إلا إذا تم استخدامها للشجب والاستنكار. وبالمثل، قد يتم السماح بعرض مقاطع الفيديو التي تتضمن تشويهات وتهديدات بالانتحار أو الوحشية إذا كان الهدف منها التنبيه أو التحذير. ففي سنة 2011، رفعت الأمريكية هيلاري آدامز دعوة قضائية ضد والدها بسبب سوء المعاملة بعد تصويره قبل سبع سنوات، وهو بصدد القيام "بتأديبها" بطريقة عنيفة. وقد نُشر الفيديو على يوتيوب، وتم استخدامه في المحاكمة.

بعد التكتّم على مسألة سياسة الإشراف على المحتوى عن عمد، بدأ يخرج عمالقة الويب تدريجيًا عن صمتهم. وفي مقال نُشر على أعمدة صحيفة الغارديان البريطانية، سلطت مونيكا بيكرت، مديرة سياسة المحتوى في شركة فيسبوك، الضوء، خلال شهر آيار/ مايو، على "ضوابط المجتمع" التي تم الإعلان عنها منذ شهر نيسان/ أبريل، والتي تشاور الجميع حولها على حسابها على فيسبوك.

في الواقع، وبعد أن أشارت بيكرت إلى الكم الهائل من البيانات التي تتم معالجتها يوميًا، أكدت أن أصعب ما تبقى فعله هو وضعها في سياقها قائلة إنه ”من الصعب الحكم على النية الكامنة وراء المنشور أو المخاطر الناجمة عن منشور آخر. فعندما يقوم شخص ما بنشر فيديو عنيف خاص بهجوم إرهابي، نحن لا نعرف إن كان أناس آخرون سيحذون حذوه أو أنه سيشجع الناس على التنديد بمثل هذه الممارسات؟“

أدركت بيكرت أوجه قصور نظام الضبط الحالي، أفادت أن ”الوضع لم يسبق له مثيل في تاريخ الويب، ويجب علينا وضع حد للفوضى بطريقة ما. ولكن لا يملك فيسبوك، في الوقت الحالي، التكنولوجيا اللازمة للقيام بتحليل سياق“.

أما فيما يتعلق بمسألة الشفافية، فقد قدمت بيكرت إجابة مذهشة إلى حد ما، وهي أنه ”إذا كانت شركة فيسبوك لا تقوم سوى بعدد قليل من الدعايات لإرشادات ضبط المحتوى، فإن ذلك يهدف إلى عدم تشجيع الأفراد على العثور على الثغرات“.” وتصر بيكرت على أن معايير المجتمع لا تميل إلى أن تكون ثابتة. وفي هذا الصدد، أوضحت أن ”معاييرنا تتطور بنمو مجتمعنا وتطور القضايا الاجتماعية حول العالم. ونحن في حوار دائم مع الخبراء والمنظمات المحلية للحديث عن كل شيء، بدءًا من المسائل المتعلقة بسلامة الأطفال ووصولًا إلى قضايا الإرهاب وحقوق الإنسان“.

عندما أدركت بيكرت أوجه قصور نظام الضبط الحالي، أفادت أن ”الوضع لم يسبق له مثيل في تاريخ الويب، ويجب علينا وضع حد للفوضى بطريقة ما. ولكن لا يملك فيسبوك، في الوقت الحالي، التكنولوجيا اللازمة للقيام بتحليل سياق“.” وللحد من هذه المشكلة، أعلنت الشركة عن توظيف 3 آلاف ”مشرف محتوى“ جديد في السنوات القادمة.

حراس الويب أو أنظمة مراقبة جديدة هشة؟

اليوم، يضطلع الذكاء الاصطناعي بمهام معينة في هذه الضوابط وخاصة بالنسبة للرسائل غير المرغوب فيها، التي يقوم فيسبوك بالإشراف عليها بصفة تلقائية. لكن في انتظار أن تكون هذه التقنية قادرة على فهم الكوميديا التي تنم عن التهكم، أو المزاح السخيف، أو المحاكاة الساخرة، أو فن التصوير العاري، فسيكون لدور المشرفين البشريين مستقبل زاهر.

على الرغم من أنه لا يتم إيلاء أهمية كبيرة للدور الشاق الذي يضطلعون به، ناهيك عن أنهم يتقاضون أجورًا غير منصفة، إلا أن المسؤولية الموكلة لهم تظل مهمة. في هذا السياق، كتبت الباحثة سارة تي. روبرتس في مقال بعنوان ”مراقبة المحتوى التجاري: العمل القذر للعمال الرقميين“ أن ”الإشراف على المحتوى ليس نشاطًا مستقلًا بحد ذاته نظرًا لأن هؤلاء العمال منتشرون في جميع أنحاء العالم“.

يتجاوز عدد المشرفين على المحتوى في العالم 100 ألف شخص. وعلى الرغم من أنه يتم تسليط الضوء تدريجيًا على طبيعة الدور الذي يضطلعون به، إلا أن عملهم يظل مخفيًا.

في هذا الصدد، لعب أولئك الذين تطلق عليهم روبرتس اسم ”حراس البوابات الرقمية“ دورًا مهمًا في المحافظة على هوية الموقع. فهم يختارون بعض المنشورات أو يحتفظون بها، ويلعبون دور الحاكم في مواجهة الأشخاص الذين ينتهكون الذوق الرفيع أو القواعد أو المعايير الاجتماعية الثقافية، وما يمكن الاحتفاظ به على الموقع. وبالنسبة لها، فقد حان الوقت أن نطرح أسئلة أخلاقية تتعلق بالأساس حول ما هو مقبول عالميًا وما هو غير مسموح به على الشبكة العالمية.

في الوقت الراهن، يتجاوز عدد المشرفين على المحتوى في العالم 100 ألف شخص. وعلى الرغم من أنه يتم تسليط الضوء تدريجيًا على طبيعة الدور الذي يضطلعون به، إلا أن عملهم يظل مخفيًا. وفي الواقع، تؤكد هذه الحقيقة بشكل غريب، تحليل سارة ت. روبرتس، ومفاده أن ”المشرف الجيد يتسم

بالقدرة على أن يكون عاملاً غير مرئي لا يترك وراءه أي أثر.

المصدر: سلايت

رابط المقال: <https://www.noonpost.com/24498/>