



إدارة المحتوى العربي المغاربي على وسائل التواصل الاجتماعي

منى السواح
أكتوبر ٢٠٢٤



يعد مركز الديمقراطية والتكنولوجيا منظمة غير حزبية وغير ربحية رائدة في مجال الحقوق المدنية و الحريات المدنية في العصر الرقمي. نحن نساهم في تشكيل سياسة التكنولوجيا، والحوكمة، والتصميم مع التركيز على المساواة والقيم الديمقراطية. بدأ المركز نشاطه لأول مرة في عام ١٩٩٤، أصبح CDT مدافعاً موثقاً عن الحقوق الرقمية منذ الأيام الأولى للإنترنت. يقع المقر الرئيسي للمنظمة في واشنطن العاصمة، وله مكتب في أوروبا في بروكسل، بلجيكا.

منى السواح
المؤلف

إدارة المحتوى العربي المغاربي على وسائل التواصل الاجتماعي

تقرير بحثي لمركز الديمقراطية والتكنولوجيا
المؤلف: منى السواح

بمساهمة من

سمير جاين، وداناراج تاكور، وعاليا بهاتيا، وديفان ل. هانكرسون.

الرسوم التوضيحية والتصميم الإضافي بواسطة أوشين سيف.

شكر وتقدير:

نود أن نشكر فريق المواطنة الرقمية، بمن فيهم أمانة ميزوني، ونام الورداني، وعلى مهني، على دعمهم الذي لا يُقدر بثمن. ونشكر أعضاء لجنتنا الاستشارية، عفاف أبروجي، وخوان كارلوس لارا، ونانجيرا سامبولي وجيليان يورك. كما نشكر المشاركين في الدراسة الذين تفضلوا علينا بمشاركتنا وقتهم وأفكارهم. نشكر ناتالي ماريشال على ترجمة الإستبيان. كما نود أن نعرب عن امتناننا للملاحظات والاقتراحات التي قدمها لنا كيت روبين، وميخال لوريا، وغابرييل نيكولاس، وفرحانة شهيد. وأخيرًا، لم يكن من الممكن إكمال هذا العمل دون دعم تيموثي هوغلاند وجورج سلوفر وسارة زولاد. جميع الآراء الواردة في هذا التقرير هي آراء مركز الديمقراطية والتكنولوجيا.

خرج هذا العمل إلى النور بفضل منحة مقدمة من مؤسسة مجتمع الإنترنت
(Internet Society Foundation)

الاقتباس المقترح:

السواح، م. (٢٠٢٤). إدارة المحتوى العربي المغاربي على وسائل التواصل الاجتماعي. مركز الديمقراطية والتكنولوجيا.

[<https://cdt.org/insights/moderating-maghrebi-arabic-content-on-social-media>]

تتضمن المراجع في هذا التقرير الروابط الأصلية بالإضافة إلى الروابط المؤرشفة والمختصرة بواسطة خدمة Perma.cc. تحتوي روابط Perma.cc أيضًا على معلومات حول تاريخ الاسترجاع والأرشفة.



هذا التقرير مُرخص بموجب ترخيص المشاع الإبداعي المنسوب للمؤلف - الترخيص المماثل ٤.٠ الدولي

محتويات

٥	مقدمة
٧	اللهجات العربية المغاربية: إرث ما بعد الاستعمار
٩	النتائج الرئيسية
٩	١. انعدام الثقة في منصات التواصل الاجتماعي في منطقة المغرب العربي
١٢	٢. عمليات إزالة المحتوى غير المبررة والحظر الخفي
١٥	٣. حالة تقييد المحتوى العربي المغاربي
٢١	التوصيات
٢١	١. الاستعانة بمفسرين من المتحدثين الأصليين باللغة
٢١	٢. إعطاء الأولوية لتوظيف فريق برمجة لغوية عصبية متنوع
٢٢	٣. تعزيز قنوات الاتصال والدعم
٢٢	٤. تعزيز الشفافية حول إزالة المحتوى والحظر الخفي
٢٣	٥. التعامل مع لهجات المشرفين وخبراتهم
٢٤	المرفقات
٢٤	المنهج البحثي وجمع البيانات
٢٦	المراجع

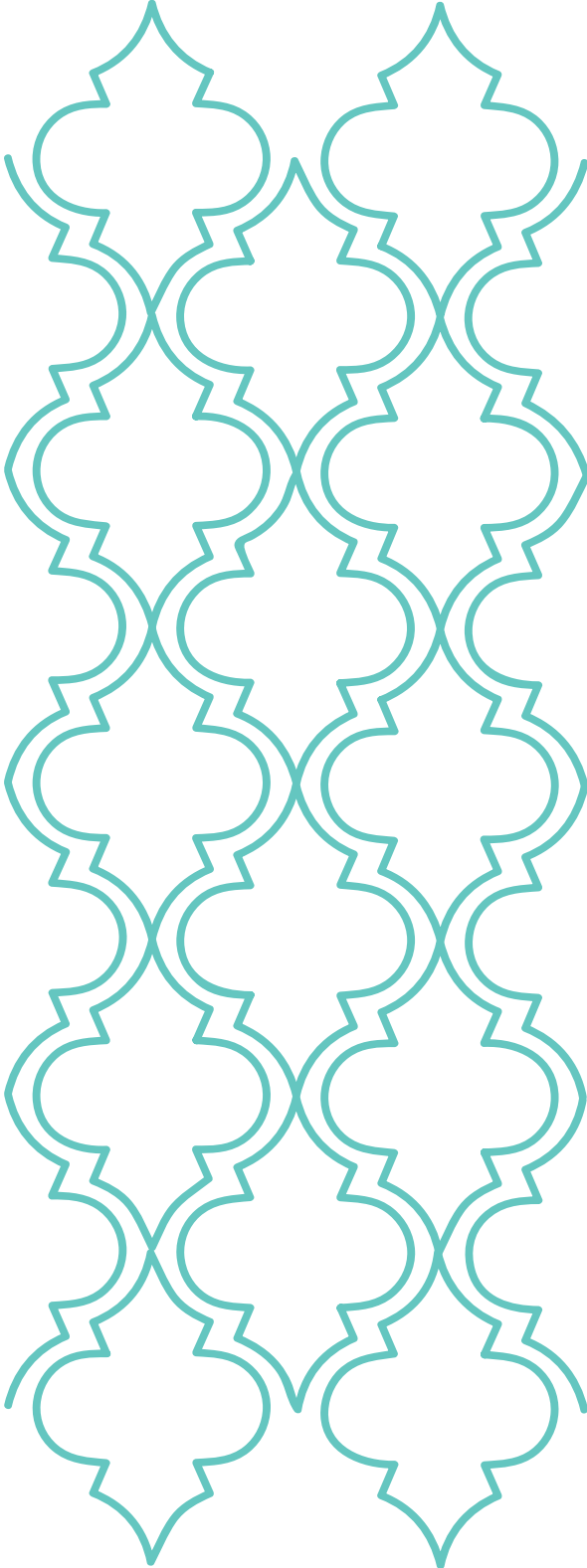
مقدمة

يعبر الناس عن أعمق رغباتهم وعواطفهم وأفكارهم الخيالية من خلال اللغة. ومع ذلك، فإن لغات شمال الكرة الأرضية، وخاصة الإنجليزية، تهيمن على تبادل المعرفة والتكنولوجيا. وقد أدى ذلك إلى تهميش لغات الدول التي تقع في الجنوب وعدم دعمها بشكل كافٍ في المجال الرقمي، لا سيما في مجال الإشراف على المحتوى. يتضمن الإشراف على المحتوى آليات حوكمة تحدد المشاركة في مجتمع معين وتتحكم في ما يتم مشاهدته وما يتم تقييمه عبر الإنترنت (جرميلمان، ٢٠١٥). من خلال حذف المحتوى الضار أو تقليله، يساعد الإشراف على المحتوى على ضمان تجارب إيجابية للمستخدم، ويحمي سمعة العلامة التجارية لمنصات التواصل الاجتماعي، ويعمل على تمكين الامتثال للمتطلبات القانونية، ويزيد من إيرادات الإعلانات (روبرتس، ٢٠١٨). في البداية، كان الاعتماد فقط على أشخاص لإدارة المحتوى، إلا أن النطاق الواسع للمحتوى الإلكتروني المنشور عبر الإنترنت استلزم دمج أنظمة الإشراف الآلية. ومع ذلك، غالبًا ما تواجه هذه الأنظمة صعوبة في التعامل مع لغات البلاد التي تقع في جنوب الكرة الأرضية بسبب عدم كفاية البيانات لتدريب أنظمة الإشراف الآلية، بالإضافة إلى الافتقار إلى فهم الفروق الثقافية الدقيقة التي تحدد معنى اللغة.

في هذا التقرير، الذي هو جزء من سلسلة مركز الديمقراطية والتكنولوجيا التي تبحث في تحيزات وفوارق الإشراف على المحتوى في الدول غير الغربية (السواح، ٢٠٢٤)، نتناول على وجه التحديد التحديات والآثار المترتبة على تقييد المحتوى في اللهجات العربية المغاربية في شمال أفريقيا. بالإضافة إلى ذلك، لم يتم التطرق إلى اللهجات العربية المغاربية في الأبحاث السابقة التي تناولت موضوع حوكمة المحتوى. باستخدام عدة مناهج بحثية تجمع بين المقابلات ومجموعات البحث والاستبيان عبر الإنترنت، وجدنا ما يلي:

(أ) تستخدم معظم شركات التواصل الاجتماعي التي يوجد مقرها في الولايات المتحدة استراتيجية عالمية للإشراف على المحتوى تطبق نفس السياسات في جميع أنحاء العالم، في حين تتبنى منصة تيك توك نهجًا محليًا بسياسات مختلفة مصممة خصيصًا لكل منطقة، خاصة فيما يتعلق بالمسائل الثقافية.

(ب) يستخدم المستخدمون المغاربة العرب أساليب عدة مثل "اللغة المشفرة" للتهرب من خوارزميات وسائل التواصل الاجتماعي خوفاً من تقييد محتواهم؛ ذلك لأنهم يعتقدون أن هذه المنصات تخضع للرقابة لأسباب سياسية. بالإضافة إلى ذلك، يميل المستخدمون إلى التعويض عن آليات الإبلاغ غير الفعالة باستخدام الإبلاغ الجماعي لإزالة المحتوى المزعج.



ج) إن عدم التنوع في تعيين فرق معالجة اللغات الطبيعية التي تطور أنظمة آلية للإشراف على المحتوى في شركات التواصل الاجتماعي، بالإضافة إلى عدم كفاية البيانات التدريبية للهجاء العربية المغاربية والاعتماد على المفسرين من غير المتحدثين الأصليين للغة، يؤثر سلباً على دقة عملية الإشراف الآلي على المحتوى.

د) يتم تعيين محتوى لمشرفي المحتوى من أي بلد في العالم العربي. مشرفي المحتوى يعملون في ظل ظروف قاسية. رغم ذلك يُتوقع من مشرفي المحتوى اتخاذ القرارات على الرغم من الفروق الثقافية واللغوية الدقيقة التي تختلف عبر المنطقة. يمكن أن يؤدي هذا إلى أخطاء في الإشراف على المحتوى في بعض الحالات.



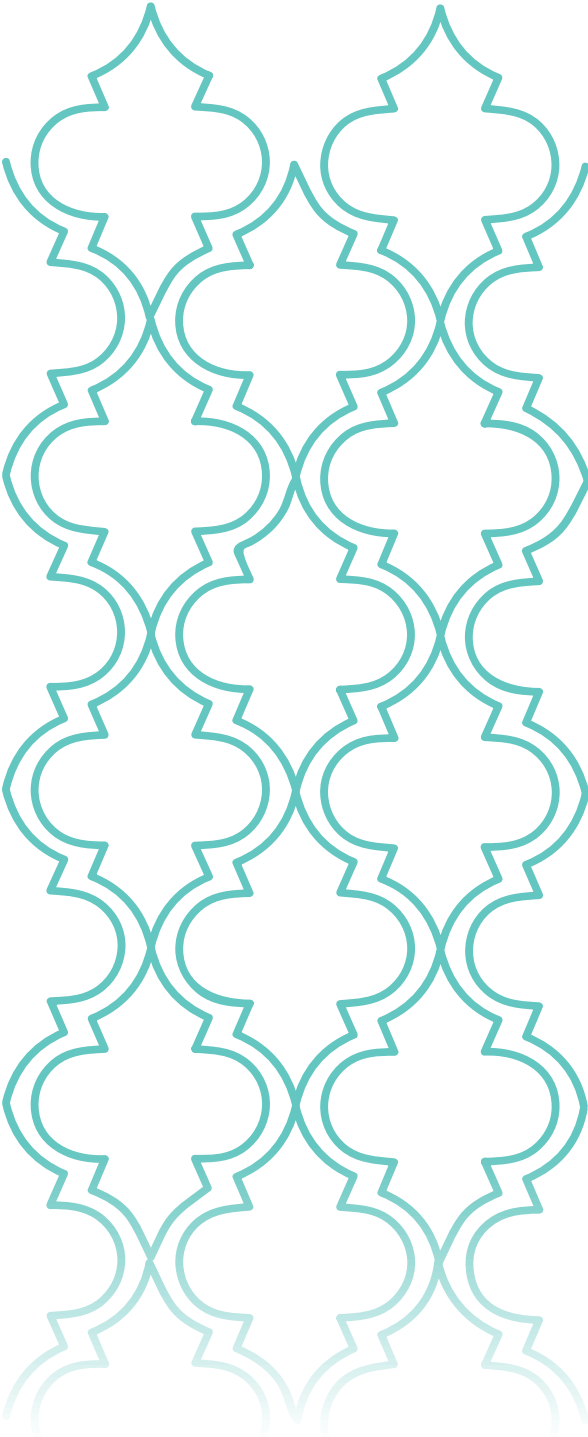
اللهجات العربية المغربية: إرث ما بعد الاستعمار

اللغة العربية هي واحدة من أكثر اللغات انتشاراً في العالم. كما أنها لغة دلالية وتعبيرية للغاية، حيث تحتوي على العديد من اللهجات والتنوعات التي يمكن للكلمات فيها أن تحمل معانٍ وعواطف مختلفة. في جميع أنحاء العالم العربي، يتحدث أكثر من ٤٠٠ مليون شخص في ٢٢ دولة هذه اللغة ولهجات مختلفة (جيليل وآخرون، ٢٠٢١). وقد تطورت اللغة العربية على مر القرون من خلال التفاعلات مع الحكام العثمانيين والبريطانيين والفرنسيين والإيطاليين ومختلف الثقافات والحضارات الأخرى. من الناحية الجغرافية، تنقسم اللهجات العربية إلى قسمين رئيسيين: لهجات الشرق الأوسط (المشرق) وشمال أفريقيا (المغرب العربي). (حرات وآخرون، ٢٠١٨). يركز هذا التقرير على اللهجات العربية المغربية (أو الدارجة) في المغرب وتونس والجزائر، والتي تشترك في العديد من أوجه التشابه الاجتماعية والتاريخية والثقافية ولها خصوصيات لغوية مشتركة (حرات وآخرون، ٢٠١٨).

تُعد اللهجات العربية المغربية إحدى اللهجات المتعددة للغة العربية العامية، ولكل لهجة مجموعة فريدة من المفردات والفروق الدقيقة. قد تحمل بعض الكلمات والعبارات معانٍ أو دلالات مختلفة عبر منطقة المغرب العربي. على سبيل المثال، كلمة "قَرعة" تعني زجاجة ماء في الجزائر ويقطين (قَرع) في تونس. وتتأثر اللهجات العربية المغربية بالعديد من اللغات الأخرى، بما في ذلك الأمازيغية واللغات الأمازيغية الأخرى، والفرنسية، والإسبانية، والإنجليزية.

أثر الاستعمار، إلى جانب أوجه القصور في الأجهزة والبرمجيات المتاحة، على الطريقة التي يكتب بها الناس اللغة العربية، مما أدى إلى ظهور شكل جديد من الكتابة باللغة العربية يعرف باسم "العربيبي" (الإنجليزي المُعرب أو الفرانكو). يتضمن "العربيبي" كتابة الكلمات العربية باستخدام الأرقام والأحرف اللاتينية في شكل مكتوب صوتياً (درويش، ٢٠١٤؛ الدويري وآخرون، ٢٠١٦). ويمكن استخدامه لتمثيل أي لهجة عربية ويُستخدم بشكل أساسي على منصات التواصل الاجتماعي وفي الرسائل النصية غير الرسمية (درويش، ٢٠١٤؛ حاجبي وآخرون، ٢٠٢٢). ظهرت طريقة "العربيبي" كاستجابة لضرورة الكتابة باللغة العربية على منصات تفتقر إلى الدعم الأصلي للنص العربي، فضلاً عن حقيقة أن العديد من المتحدثين باللغة العربية يتحدثون لغتين. (درويش، ٢٠١٤). ومن الآثار ما بعد الاستعمارية الأخرى على اللهجات العربية المغربية ما يعرف بـ "التبديل اللغوي"، وهو استخدام لغتين أو أكثر في نفس الجملة (حاجبي وآخرون، ٢٠٢٢). وي طرح هذا النوع من التبديل اللغوي تحدياً إضافياً، حيث يتطلب القراءة من اليسار إلى اليمين للجزء المكتوب بالأحرف اللاتينية وفي الاتجاه المعاكس للجزء العربي (يونس وآخرون، ٢٠٢٠).

وعلى الرغم من أن أكثر من ١٠٠ مليون شخص يتحدثون اللهجات العربية المغربية، إلا أنها تعتبر "لغات منخفضة الموارد". (يونس وآخرون،



٢٠٢٠). يشير هذا المصطلح إلى اللغات التي لا يتوفر لها سوى عدد محدود من النصوص عالية الجودة، مما يُعقّد عملية تدريب أنظمة الإشراف الآلي على المحتوى (نيكولاس وباتيا، ٢٠٢٣). ويؤدي انتشار العريبزي والتبديل اللغوي إلى تفاقم هذه التحديات، مما يعيق الإشراف الآلي الفعال وتطوير نماذج قوية للذكاء الاصطناعي. وسوف نناقش هذه التحديات وغيرها بمزيد من التفصيل في هذا التقرير.



^١ هناك طرق أخرى لتصنيف اللهجات العربية. على سبيل المثال، يقسمها البعض إلى ٥ فئات: لهجات الجزيرة العربية، والمصرية، والشامية، ولهجة بلاد الرافدين، واللهجات العربية المغاربية (يونس وآخرون، ٢٠٢٠).

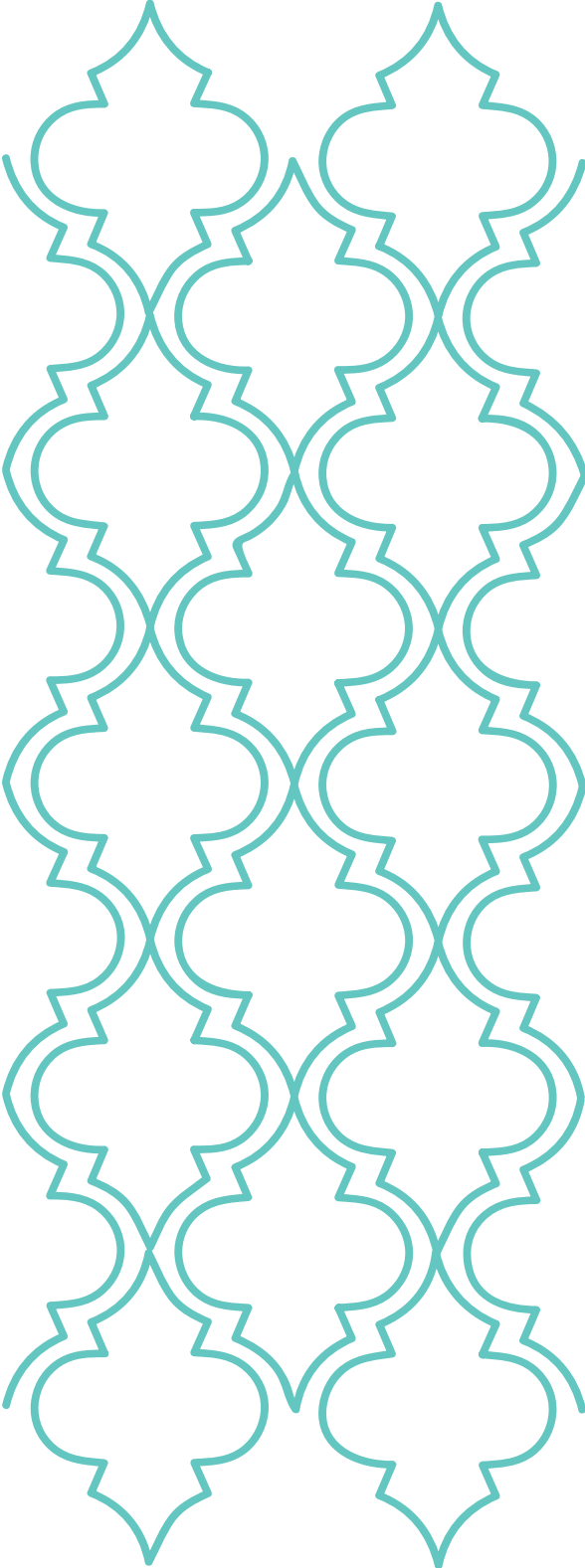
النتائج الرئيسية

في هذا التقرير، ندرس كيف تعمل الخدمات عبر الإنترنت على تعديل اللهجات العربية المغربية. أجرينا ١٤ مقابلة عبر الإنترنت مع مشرفي المحتوى والموظفين الحاليين والسابقين من أقسام الثقة والسلامة (T&S) والسياسة في ميتا وتيك توك ويوتيوب وإكس. كما أجرينا مقابلات مع المدافعين عن الحقوق الرقمية وأجرينا استطلاعاً عبر الإنترنت مع ١١١ مستخدماً للخدمات عبر الإنترنت في تونس والجزائر والمغرب. بالإضافة إلى ذلك، أجرينا خمس جلسات جماعية مركزة عبر الإنترنت مع مستخدمي الإنترنت البارزين وصانعي المحتوى في منطقة المغرب العربي (انظر الملحق لمزيد من التفاصيل حول طرق البحث المستخدمة في هذا التقرير). علاوة على ذلك، قمنا بتنظيم مائدة مستديرة مع باحثين في معالجة اللغات الطبيعية (NLP) من منطقة المغرب العربي لمناقشة التحديات والفرص لتطوير نماذج ذكاء اصطناعي موثوقة وعادلة للهجات العربية المغربية. وبناء على هذا البحث، توصلنا إلى النتائج التالية:

١. انعدام الثقة في منصات التواصل الاجتماعي في منطقة المغرب العربي

إن المجال العام على الإنترنت في منطقة المغرب العربي نابض بالحياة. ويعتمد المستخدمون على منصات التواصل الاجتماعي للتواصل والعمل والمشاركة في الأنشطة والمناقشات السياسية. من خلال استطلاعنا عبر الإنترنت، تعد منصات مثل فيسبوك وإنستغرام ويوتيوب وتيك توك هي الخدمات الأربع الأكثر شعبية عبر الإنترنت بين المستخدمين المغاربة الناطقين باللغة العربية. وعلى الرغم من هذا التفاعل، فإن مستخدمي وسائل التواصل الاجتماعي في منطقة المغرب العربي يُظهرون مستوى من عدم الثقة تجاه منصات وسائل التواصل الاجتماعي. أفاد حوالي ٥٧٪ من المستخدمين الذين شاركوا في استطلاعنا أنهم لا يعتقدون أن شركات وسائل التواصل الاجتماعي تبذل قصارى جهدها لإدارة المحتوى من منطقة المغرب العربي. وأعرب حوالي ٥٨٪ عن عدم ثقتهم بشركات التواصل الاجتماعي فيما تقدمه من محتوى. بالإضافة إلى ذلك، أعرب أكثر من ٦٢٪ من المشاركين في الاستطلاع عن قلقهم بشأن حذف منشوراتهم. وخلال جلسات مجموعة التركيز، ذكر المشاركون مراراً وتكراراً أنهم "يفكرون مرتين قبل إنشاء أي محتوى" لتجنب أي حذف أو قيود لمحتواهم.

وقد ينبع انعدام الثقة هذا من منظور ما بعد الاستعمار لشعوب منطقة المغرب العربي، وهو ما يعكس الحساسيات التاريخية وديناميكيات السلطة. كما أعرب المشاركون في مجموعة التركيز عن أن الأصول الأمريكية للعديد من شركات وسائل التواصل الاجتماعي تساهم في



تصورها ككيانات بعيدة منفصلة عن الواقع المحلي للمستخدمين في منطقة المغرب العربي. وقد يفسر هذا أيضًا تزايد شعبية تيك توك في المنطقة، حيث أعرب المشاركون في مجموعات التركيز والمقابلات عن مستوى أعلى من الثقة تجاه منصة تيك توك المملوكة لشركة ByteDance الصينية، مقارنة بشركة ميتا المملوكة للولايات المتحدة عندما يتعلق الأمر بالتعبير عن آرائهم بحرية. قال أحد صانعي المحتوى من المغرب:

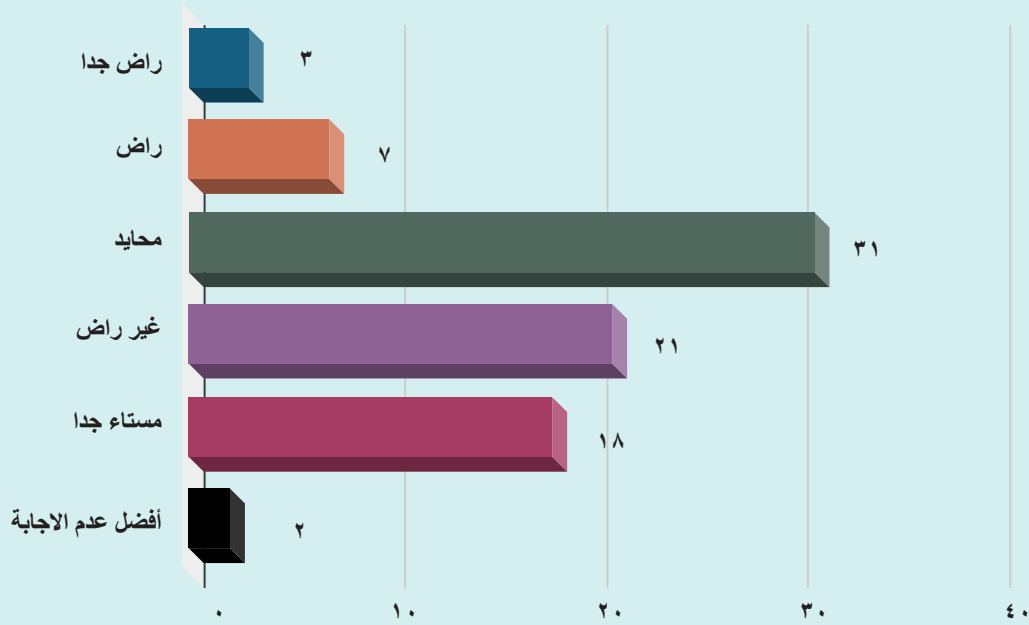
”أعتقد أن أي شيء تجده على تيك توك، يمكنك العثور عليه أيضًا على إنستغرام. تكمن جاذبية منصة تيك توك في الحرية التي توفرها. لقد لاحظنا الكثير من الحظر الخفي على إنستغرام وفيسبوك، لكن تيك توك أكثر حرية، وهو ما يجذب العديد من الأشخاص الذين يمكنهم مشاهدة المحتوى بشغف لساعات عديدة.“ (صانع محتوى، المغرب، أبريل ٢٠٢٤)

لا يرتبط انعدام الثقة هذا فقط بالحفاظ على المحتوى، ولكنه يرتبط أيضًا بكيفية استجابة الشركات للتقارير للبلاغات المتعلقة بالمحتوى الذي يعتبره المستخدمون غير مناسب. وقال أكثر من ٧٤٪ من المشاركين في الاستطلاع إنهم أبلغوا في السابق عن محتوى ينتهك سياسات بعض المنصات التي استخدموها. إلا أن حوالي ٥٠٪ كانوا غير راضين عن مدى استجابة منصة التواصل الاجتماعي لبلاغاتهم، مشيرين إلى أن المحتوى المخالف بقي على المنصة (انظر الشكل التوضيحي ١). صرح أحد صانعي المحتوى التونسيين قائلًا: ”لا توجد تقنية شفافة للإبلاغ في الإبلاغ، ولا فهم للسياق التونسي والعربي والمحلي.“

وقد دفع انعدام الثقة هذا، إلى جانب غياب قنوات الاتصال الفعالة، بعض المشاركين إلى أخذ الأمور على عاتقهم، خاصة عندما يتعلق الأمر بالإبلاغ عن المحتوى. ذكر المشاركون أنهم لجأوا إلى ”حملات الإبلاغ الجماعية“ لإزالة المحتوى أو الحسابات التي فشلت شركات التواصل الاجتماعي في إزالتها باستخدام قنوات الإبلاغ العادية. الإبلاغ الجماعي يعني ”إساءة استخدام خاصية الإبلاغ في منصات التواصل الاجتماعي لحذف الحسابات والصفحات“ (السواح، ٢٠٢٣). تم اللجوء إلى هذه الطريقة من قبل بعض المستخدمين في منطقة المغرب العربي الذين رأوا أن شركات وسائل التواصل الاجتماعي تقلل بشكل غير عادل من المحتوى الذي يجدونه مزعجًا أو مسيئًا.

ونتيجة لذلك، تحملت منظمات المجتمع المدني في المنطقة على عاتقها مسؤولية العمل كوسيط حيث تربط المستخدمين المعنيين بشركات التكنولوجيا، خاصة فيما يتعلق بعمليات الإزالة وتقييد المحتوى الضار.

ما مدى رضاك عن إستجابة منصات التواصل الاجتماعي لبلاغك؟



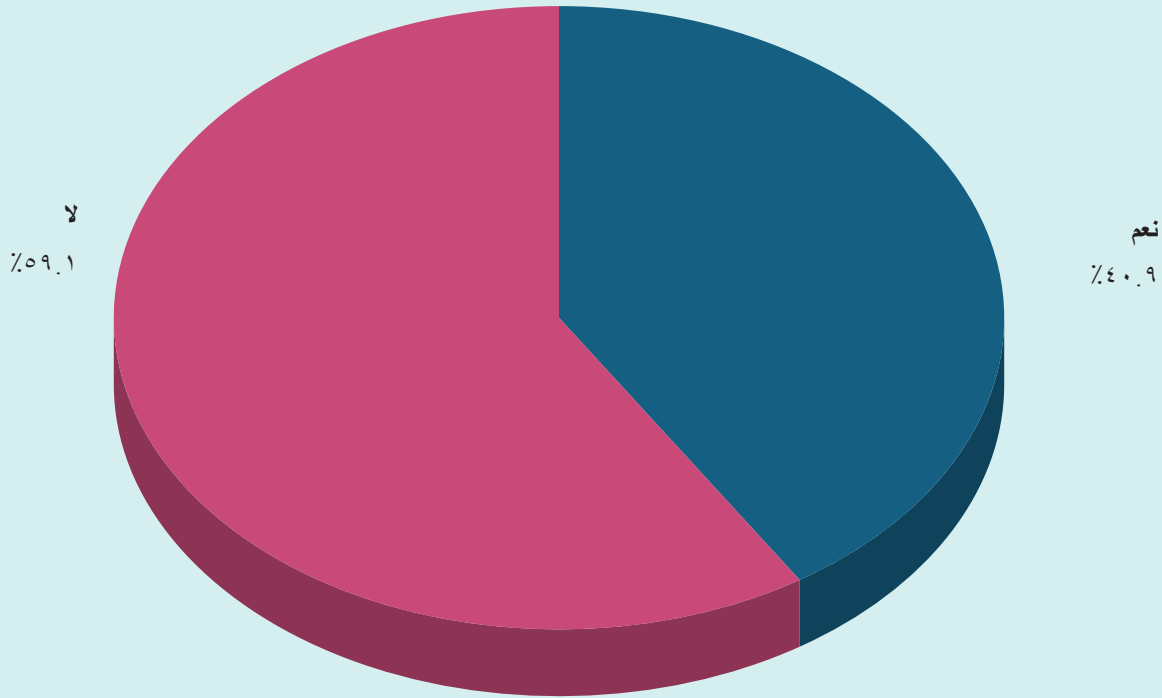
الشكل ٠١. مدى رضا المستخدمين عن استجابة منصات التواصل الاجتماعي لعملية الإبلاغ (% من الذين قالوا إنهم أبلغوا عن محتوى إلى إحدى منصات التواصل الاجتماعي، العدد = ٨٢). المصدر استطلاع عبر الإنترنت قام به مركز الديمقراطية والتكنولوجيا (أبريل-مايو ٢٠٢٤).

وقد أصبحت هذه الممارسة الآن موحدة، وأصبح الرد على هذه البلاغات يتم بشكل مؤسسي داخل شركات التكنولوجيا. ومع ذلك، أشار المدافعون عن الحقوق الرقمية في مقابلاتنا إلى أن شركات التكنولوجيا قد تستغرق أشهرًا للرد حتى على تدخلات المجتمع المدني. وتحدث أحد أعضاء فريق السياسات في إحدى شركات وسائل التواصل الاجتماعي التي مقرها الولايات المتحدة عن هذا التصعيد من طرف ثالث، قائلاً:

”هناك حالات تصلني من خلال منظمات المجتمع المدني مثلاً، أو من خلال بعض الأكاديميين وأنا أقوم بتصعيد الأمر داخلياً للمراجعة. لذا، فإن هذا في بعض الأحيان جزء من العملية أيضاً. إنه مثل نوع من تصعيد عبر طرف ثالث.“
(موظف فريق السياسات، أبريل ٢٠٢٤)

هذا التصعيد من جانب المجتمع المدني بسبب عدم الاستجابة الكافية لبلاغات المستخدمين يؤكد الحاجة إلى قنوات اتصال أكثر شفافية وسهولة في الوصول إلى شركات التكنولوجيا للإبلاغ والمناشدات.

هل سبق أن تمت إزالة المحتوى الخاص بك أو تقييده بواسطة إحدى منصات التواصل الاجتماعي؟



٢. عمليات إزالة المحتوى غير المبررة والخطر الخفي

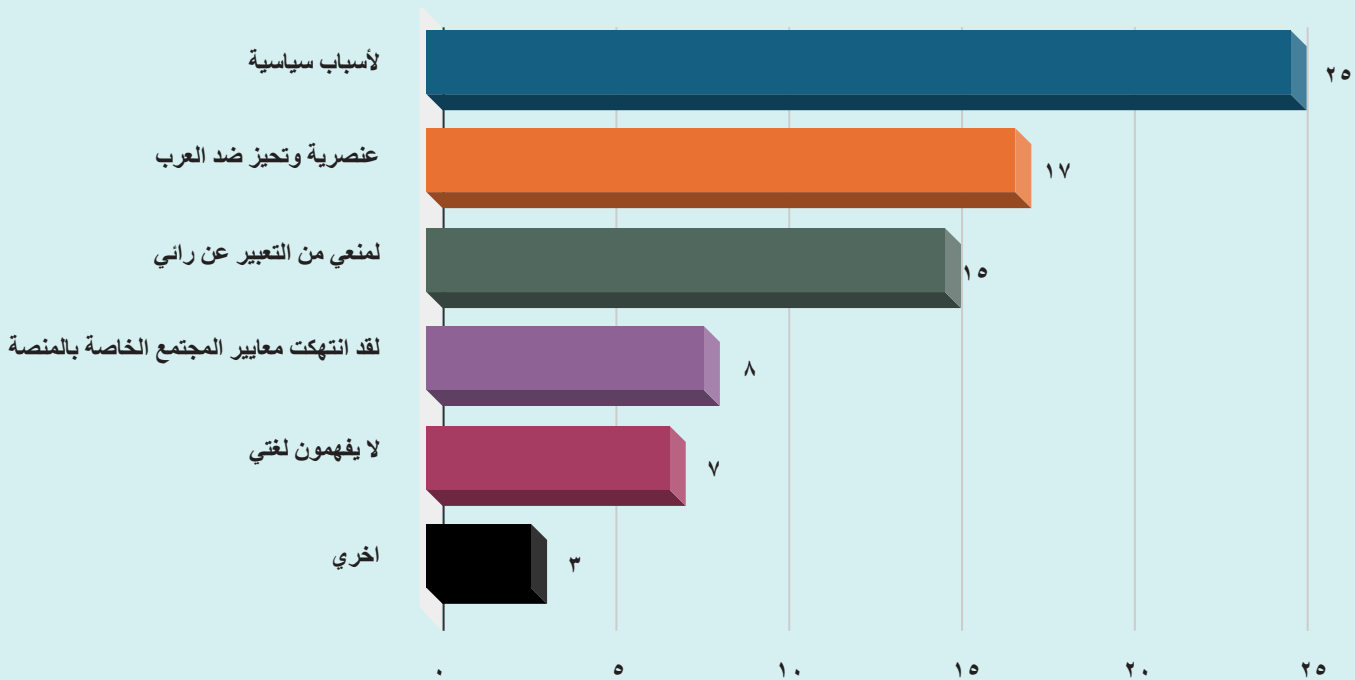
واجه المشاركون في هذه الدراسة إزالة لمحتواهم أو الحظر الخفي أثناء نشر المحتوى على منصات مختلفة عبر الإنترنت. والخطر الخفي هو شكل من أشكال عملية الإشراف على المحتوى غير المكشوف يتم من خلاله إخفاء محتوى المستخدم أو تقليل ظهوره بشكل كبير دون إبلاغ المستخدم (نيكولاس، ٢٠٢٢). يمكن أن يؤدي الحظر الخفي إلى جعل تعليقات المستخدم أو التعامل معه أو المحتوى يبدو طبيعيًا بالنسبة له ولكنه يكون غير مرئي أو يصعب على الآخرين العثور عليه (نيكولاس، ٢٠٢٢).

في الاستطلاع، أشار حوالي ٤٠٪ من المشاركين إلى أنهم تعرضوا لعملية إزالة المحتوى في مرحلة ما، ويعتقد الأغلبية أن ذلك كان لأسباب سياسية (انظر الشكلين ٣ و ٢). بالإضافة إلى ذلك، أشار أكثر من ٦٢٪ من المتحدثين باللغة العربية المغاربية الذين شاركوا في الاستطلاع إلى أنهم يشعرون بالقلق إزاء إسكاتهم بطريقة أو بأخرى من قبل شركات وسائل التواصل الاجتماعي.

غالبًا ما تكون عملية إزالة المحتوى مُربكة وغير متوقعة. أبلغ المشاركون عن عدم اليقين بشأن توقيت ومعايير إزالة المحتوى، مما يُعقد قدرتهم على

الشكل ٢. تكرار عمليات إزالة المحتوى أو القيود التي يواجهها المشاركون على منصات التواصل الاجتماعي المختلفة (٪ من المشاركين في الاستطلاع، العدد = ١١٠) المصدر: استطلاع عبر الإنترنت قام به مركز الديمقراطية والتكنولوجيا (أبريل-مايو ٢٠٢٤).

لماذا تعتقد أن منصة التواصل الاجتماعي قامت بإزالة المحتوى الخاص بك؟ (يمكنك اختيار أكثر من إجابة)



الشكل ٣. الأسباب التي يراها المستخدمون لإزالة المحتوى. (٨٢٪ ممن قالوا إنه تمت إزالة المحتوى الخاص بهم أو تقييده بواسطة إحدى منصات التواصل الاجتماعي، العدد = ٨٢). أتيح للمشاركين في هذا الاستبيان اختيار أكثر من إجابة. المصدر: استطلاع عبر الإنترنت قام به مركز الديمقراطية والتكنولوجيا (أبريل-مايو ٢٠٢٢)

توقع مثل هذه الإجراءات. وخلال جلسات مجموعة التركيز، لاحظ بعض المشاركين أنه تتم أحياناً إزالة المحتوى الذي يبدو غير ضار، في حين قد يظل المحتوى الإشكالي كما هو دون معالجة. ووجد المشاركون أن الحظر الخفي - خاصة المحتوى المتعلق بفلسطين - أكثر إرباكاً.

أشار معظم المشاركون إلى أن الحرب في غزة هي التي عجلت بفرض الرقابة الشديدة التي واجهوها على منصات التواصل الاجتماعي، وخاصة على فيسبوك وإنستغرام، وكلاهما مملوك لشركة ميتا. وبحسب المشاركين، فإن هذا الشعور بالحظر الخفي هو السائد في المحتوى المتعلق بفلسطين بشكل عام. وأشار صانعو المحتوى إلى أن المحتوى الخاص بهم عن فلسطين لم يصل إلى نفس مستوى المشاهدة الذي حققته أنواع المحتوى الأخرى. يقول أحد المؤثرين على إنستغرام:

“كلما ذكرت فلسطين في قصصي، لا يتم مشاهدة قصصي. عادةً ما أحصل على ٢٠ ألف شخص يشاهدون قصصي، لكن عندما أتحدث عن فلسطين، بالكاد أتمكن من الوصول إلى ١٠٠٠ مشاهدة.”
(صانع محتوى، المغرب، مارس ٢٠٢٤)

الحظر الخفي هو مفهوم واسع؛ فجميع المنصات تصمم خوارزمياتها لتحديد أولويات بعض المحتوى وتقليل أولوية المحتوى الآخر. في بعض الحالات، يمكن لمشرفي المحتوى اختيار خفض محتوى معين بشكل انتقائي

بدلاً من إزالته بالكامل. كشف أحد مشرفي المحتوى التابع لمنصة تيك توك الذين قابلناهم أن إحدى الأدوات المتاحة لهم هي تصنيف مقطع فيديو على أنه "يصعب العثور عليه". قد يتسبب هذا في قيام الخوارزميات بإخفاء المحتوى على المنصة، مما يجعل من المستحيل على المستخدمين العثور على الفيديو.

وفي المقابلات التي أجريناها، أوضح موظف سابق في ميتا أن هذه الظاهرة قد تظهر من خلال آليتين أساسيتين. أولاً، يمكن للحظر الخفي أن يستهدف "حسابات" معينة تنتهك سياسات المنصة بشكل متكرر. ثانياً، يمكن للحظر الخفي أن يستهدف "موضوعات أو مجالات محتوى" محددة. يُعرف هذا بإلغاء الأولويات الخوارزمية، خاصة في أنظمة البحث أو التوصية (نيكولاس، ٢٠٢٢). على سبيل المثال، أعلنت شركة ميتا مؤخراً أنها ستتوقف عن التوصية بشكل استباقي بالمحتوى السياسي على إنستغرام وثريندز (لورنز ونيكس، ٢٠٢٤).

لتجنب الحظر الخفي وإزالة المحتوى، اعتمد بعض المستخدمين ومنشئي المحتوى في منطقة المغرب العربي ما يعرف بتكتيكات "اللغة المشفرة". اللغة المشفرة هي اعتماد كلمات أو عبارات رمزية إبداعية لتجنب إزالة المحتوى أو خفض التصنيف بواسطة أنظمة الإشراف على المحتوى (لورنز، ٢٠٢٢). سلط المشاركون في جلسات مجموعة التركيز والمقابلات الضوء على أربعة أساليب للغة المشفرة. أولاً، يقوم المستخدمون بشكل استراتيجي بتحميل محتوى عشوائي مُصاحب للمواد التي يعتقدون أنها ستكون عرضة للتقييد. ثانياً، يتلاعب المستخدمون بلغتهم العربية عن طريق استبدال الحروف بالأرقام أو استخدام تقنيات تُهجن غير تقليدية أخرى مثل كتابة اللغة العربية بدون نقاط. على سبيل المثال، بدلاً من كتابة "جمعة مباركة"، يمكن للمستخدم كتابتها ~~جمعه مباركه~~، مع إزالة النقاط وإدراج مسافات إضافية دون تغيير المعنى. ثالثاً، يعد استخدام الرموز أو الرموز التعبيرية كبديل لمفاهيم أخرى أمراً شائعاً أيضاً؛ أحد الأمثلة الرئيسية هو استخدام رمز البطيخ كإشارة إلى العلم الفلسطيني. وأخيراً، يختار بعض الأفراد تجنب كلمات معينة تماماً لتجنب الحذف.

لقد كشفت المقابلات مع باحثي معالجة اللغات الطبيعية وموظفي شركات التكنولوجيا أن أساليب المحاكاة هذه قد تكون ناجحة في التحايل على اكتشاف وسائل التواصل الاجتماعي. ومع ذلك، تقوم المنصات بتحديث خوارزمياتها باستمرار من خلال التدريب المتكرر. لذلك، أكد أحد الباحثين في معالجة اللغات الطبيعية والذي عمل في إحدى شركات التكنولوجيا على أن "المستخدمين بحاجة إلى مواصلة الإبداع في اللغة المشفرة الخاصة بهم حتى لا تصبح هذه اللغة المشفرة شيئاً يمكن للنماذج أن تعناد عليه ومن ثم تكتشفه."

٣. حالة تقييد المحتوى العربي المغاربي

أ. الصدمة النفسية المرتبطة بالإشراف على المحتوى

كما هو الحال مع اللغات الأخرى، يتم الإشراف على محتوى اللهجات العربية المغاربية باستخدام مزيج من الإشراف البشري والآلي على المحتوى. كان الإشراف البشري (اليدوي) على المحتوى هو النوع الأول من الإشراف المستخدم لإدارة المحتوى قبل دمج الأدوات الآلية ونماذج التعلم الآلي. وهذا ما قد صفه موظف سابق عن السنوات الأولى من الإشراف على المحتوى في شركة ميتا:

”كانت الخطة في الأساس هي توظيف طلاب جامعيين لإدارة طلاب جامعيين آخرين ... لكن [فيسبوك] استمر في النمو، وتوسع أولاً خارج الكليات ثم بسرعة كبيرة خارج الولايات المتحدة. ... وهذا يعني أنه كان علينا إضافة المزيد من الأشخاص إلى القوى العاملة المسؤولة عن إدارة المحتوى.“
(موظف سابق في شركة ميتا، أبريل ٢٠٢٤)

يتم الاستعانة بمصادر خارجية للإشراف على المحتوى لشركات مختلفة في العالم العربي. يُدفع للجهات التي تقدم خدماته لشركات التكنولوجيا حول العالم ملايين الدولارات سنوياً لتقديم خدمات الإشراف على المحتوى لشركات التكنولوجيا الأجنبية (ساتاريانو وإسحاق، ٢٠٢١). وقد ذكر المشاركون في الدراسة ستة مقدمي خدمات مختلفين موجودين في أربع دول عربية في المنطقة. ومقدمو الخدمات هؤلاء، بما فيهم Teleperformance و Concentrix في تونس و Teleperformance في المغرب، مسؤولون إلى حد كبير عن إدارة المحتوى العربي لشركات التكنولوجيا المختلفة. للحصول على وظيفة مشرف محتوى، يحتاج المتقدم فقط إلى إثبات كفاءة متوسطة في مهارات اللغة الإنجليزية واجتياز مقابلة مع ممثل الموارد البشرية وطبيب نفسي. وفي حالة التوظيف، فإنهم يخضعون لتدريب يمكن أن يستمر لمدة شهر تقريباً. في المقابلات التي أجريناها، لاحظ المشاركون أن مشرفي المحتوى يتلقون تعليمات حول سياسات مختلفة وغالباً ما يُطلب منهم حفظها أثناء التدريب. ثم يتعين عليهم اجتياز اختبارات تثبت مدى إلمامهم بسياسات الشركة وفهمهم لمختلف اللهجات في العالم العربي. وإذا اجتازوا كلا الاختبارين، فإنهم يبدأون مهمة الإشراف على المحتوى من خلال تقييم المحتوى الذي يتلقونه.

وكما هو الحال في المناطق الأخرى، تؤثر هذه الوظيفة بشدة على مشرفي المحتوى (دوسكين وآخرون، ٢٠١٩). يصف العديد من مشرفي المحتوى الذين أجرينا مقابلات معهم بأنهم ما زالوا يتعافون من الصدمة التي تعرضوا لها خلال عملهم. وذكر المشاركون أنهم عانوا من القلق والاكتئاب والأرق، حيث احتاج البعض إلى مضادات الاكتئاب للتعامل مع هذه الأعراض، حتى بعد الاستقالة من مناصبهم. وفي مقابل هذا العمل الشاق، يحصلون على راتب شهري يعادل ٤٠٠ - ٦٠٠ دولار أمريكي، مع مكافأة محتملة قدرها ١٠٠ دولار أمريكي بناءً على أدائهم. واعتبر المشاركون أن هذا التعويض منخفض وغير كافٍ نظير الصدمة التي تعرضوا لها. يوجد لدى جميع مقدمي الخدمات قسم خاص بعلم النفس لتقديم الدعم في مجال الصحة العقلية والنفسية لمشرفي المحتوى، على الرغم من أن المشاركين اعتبروا هذا القسم غير فعال إلى حد كبير. وأشار أحد مشرفي المحتوى:

”لقد أثر الأمر عليّ وعلى صحتي العقلية النفسية بشدة. حتى الأخصائيين النفسيين لم يساعدوا. لقد وصلت إلى مرحلة اضطررت فيها إلى تناول الأدوية حتى أتمكن من النوم. كان الأمر بهذا السوء. صدقني، رؤية هذا المحتوى مرة أو مرتين أمر طبيعي نوعاً ما. لكن عندما ترى ما يقارب ١٥٠ من هذا المحتوى كل يوم لمدة عام، فهذا ليس طبيعياً. ليس هذا فقط. إنه يؤثر على طريقة تفكيرك أيضاً. أنت ترى الناس بطريقة أخرى لأن الأشخاص الذين تراهم على المنصة يختلفون كثيراً عن الأشخاص الذين تعرفهم في الحياة الواقعية.“
(مشرف المحتوى، مارس ٢٠٢٤)

لا يقتصر الأمر على أن الصدمة النفسية تجعل من المستحيل الاستمتاع بهذه الوظيفة، بل يتعين على مشرفي المحتوى أيضاً التعامل مع نظام تقييم صارم يعتمد على عدد مقاطع الفيديو التي يراجعونها يومياً، والوقت الذي يستغرقه لاتخاذ القرارات، ومدى دقة القرار. ذكر المشاركون أنه يتعين عليهم التعامل مع مئات الحالات من المحتوى المزعج بشكل يومي. وذكر أحد مشرفي المحتوى الذي يعمل في تيك توك أنه يتعين عليهم مراجعة حوالي ١٧٠ مقطع فيديو يومياً، واتخاذ قرار بشأن كل مقطع فيديو في غضون ٢٠ ثانية إلى دقيقة. وكلما أكملوا المهمة بشكل أسرع، كانت مقاييس أدائهم تبدو أفضل في نهاية الأسبوع. وفي بعض الحالات، كان عليهم تسريع مقاطع الفيديو لاتخاذ القرار بشكل أسرع. على الرغم من ضغوط العمل السريع، فإن أي خطأ قد يؤثر على راتبهم الشهري وقد يؤدي إلى إنهاء خدمتهم.

ب. نهج إدارة المحتوى العالمي مقارنة بالمحلي

من خلال المحادثات مع العديد من مشرفي المحتوى وممثلي شركات التكنولوجيا وفرق السياسات في شركات التكنولوجيا المختلفة، وجدنا أن

هناك طريقتين للإشراف على المحتوى: أ) نهج عالمي، والذي يستلزم استخدام نفس السياسات على جميع المستخدمين، و ب) نهج محلي يصمم ويغير بعض السياسات لتلائم كل منطقة. لقد وجدنا أن تيك توك تتبنى عمومًا استراتيجية أكثر محلية للإشراف على المحتوى، مع الاعتراف بتميز كل منطقة وتنفيذ سياسات مخصصة لاستيعاب الاختلافات الإقليمية والفروق الثقافية الدقيقة. وفقًا للمشاركين، فإن هذه الخطوة تأتي في محاولة لتبني نهج أقل تقييدًا في إدارة المحتوى من أجل اكتساب قاعدة جماهيرية قوية في بلدان الجنوب العالمي.

وعلى العكس من ذلك، تختار ميتا وإكس ويوتيوب نهجًا عالميًا للإشراف على المحتوى، حيث تستخدم سياسات موحدة على جميع المستخدمين في جميع أنحاء العالم. ويعتبر هذا النهج أسهل في التطبيق ويهدف إلى تحقيق معاملة عادلة لكافة المحتويات. صرّح موظف سابق في فريق السياسات بإحدى هذه الشركات بما يلي:

”بصراحة تامة، من الأسهل بالنسبة للشركة أن يكون لديها نسخة واحدة من السياسة لتطبيقها عالميًا، وبالتالي فإن جعلها النسخة الأكثر صرامة التي تنطبق على سوق رئيسية مثل أوروبا، فهو أسهل بكثير من وجود نسخة سياسة منفصلة تمامًا لأوروبا ولا تنطبق على أي مكان آخر.“

(عضو سابق في فريق السياسات، أبريل ٢٠٢٤)

وأشار مشاركون آخرون إلى أن الإشراف على المحتوى العالمي هو الطريقة التي تضمن بها شركات التكنولوجيا عدم قيام الحكومات بتحديد قواعد الإشراف على المحتوى، وبالتالي المساعدة في تجنب انتهاكات حقوق الإنسان. في حين يسعى نهج الإشراف على المحتوى العالمي إلى تطبيق جميع السياسات على جميع البلدان دون تمييز، فإن هذه السياسات تلتزم في المقام الأول بقوانين الشمال العالمي، حيث تخضع شركات التكنولوجيا للمساءلة. على سبيل المثال، تتبع قائمة الأفراد والمنظمات الخطرة المثيرة للجدل التي أعدتها شركة ميتا، والتي اتُهمت بالتحيز ضد المسلمين والعرب، السياسات الخارجية للولايات المتحدة والمملكة المتحدة والاتحاد الأوروبي. (بيدل، ٢٠٢١). وقد صرّح موظف سابق في شركة ميتا:

”الطريقة التي قمنا بها بتجميع هذه القائمة كانت في الأساس من خلال أخذ قائمة المواطنين الأمريكيين المعيّنين خصيصًا والمنظمات التي يعينونها، بالإضافة إلى المنظمات التي يديرها أي مواطن محدد خصيصًا ودمجها مع قائمة الاتحاد الأوروبي وقائمة المملكة المتحدة. وبالتالي فإن التحيز الذي تراه هناك هو بصراحة نتيجة للتحيز في تلك القوائم.“

(موظف سابق، ميتا، أبريل ٢٠٢٤)

كما ذكرنا سابقاً، يطبق تيك توك نهجاً للإشراف على المحتوى المحلي. تقسم إدارة محتوى تيك توك العربي منطقة الشرق الأوسط وشمال أفريقيا إلى منطقة الشرق الأوسط وشمال أفريقيا ١ (لبنان ومصر وفلسطين وسوريا)، ومنطقة الشرق الأوسط وشمال أفريقيا ٢ (السودان والخليج واليمن)، ومنطقة الشرق الأوسط وشمال أفريقيا ٣ (تونس والمغرب والجزائر وموريتانيا ليبيا). في الآونة الأخيرة، تم دمج منطقة الشرق الأوسط وشمال أفريقيا ٢ و ١. ويتم تعيين مشرف محتوى في تيك توك لمنطقة فرعية واحدة أو أكثر، تحتوي كل منها على قائمة بالألفاظ البذيئة والسياسات والقواعد الثقافية التي تتم مشاركتها مع مشرفي المحتوى. على سبيل المثال، من غير المقبول أن ترتدي المرأة قمصاناً بأكمام قصيرة في منطقة الشرق الأوسط وشمال أفريقيا ٢، ولكن ليس في منطقتي الشرق الأوسط وشمال أفريقيا ٣ و ١. إلا أن المشرفين عبروا عن ارتباكهم إزاء هذه السياسات وعدم قدرتهم على فهم كل هذه السياسات الإقليمية المميزة في وقت واحد. بالإضافة إلى ذلك، شعروا أن هذه السياسات تفتقر إلى تفسيرات واضحة لدرجة أن بعض مشرفي المحتوى ذوي الخبرة اضطروا إلى كتابة قائمة تفسيرية لشرح السياسات للموظفين الجدد. وأضاف أحد مشرفي المحتوى:

”لكننا قمنا بتحديث جدول البيانات [الألفاظ البذيئة]. كنا نقوم بتحديث بعض أسماء المنظمات وأعلامها وسياساتها بشكل عام، مثل فهم السياسات. وقد كان لديهم دائماً تفسيرات مختلفة. وكنا نحاول مواكبة تفسيرهم الخاص للأشياء.“
(مشرف المحتوى، أبريل ٢٠٢٤)

في حين أن منصة تيك توك تطبق نهجاً أكثر محلية، فإن الشركات الأخرى التي قمنا بدراساتها في هذا التقرير تستخدم نهجاً أكثر اتساقاً للإشراف على المحتوى، وإن كان مع بعض الاختلاف حسب المنطقة/البلد. وفقاً للذين تمت مقابلتهم، لدى شركة ميتا قائمة بالكلمات البذيئة لكل بلد، والتي يتم تحديثها بشكل متكرر. ومع ذلك، لا يوجد تمييز بين المناطق فيما يتعلق بتوظيف سياساتها.

على الرغم من ذلك، أفاد مشرفي المحتوى على كل من تيك توك وميتا أن الاختلافات في اللهجات عبر العالم العربي غالباً ما تسبب ارتباكاً. وقد تم إسناد محتوى للمشرفين الناطقين باللغة العربية من أي بلد في المنطقة، بغض النظر عن لهجتهم الأصلية، الأمر الذي أدى في كثير من الأحيان إلى حدوث أخطاء. ويفترض هذا النهج أن جميع العرب يفهمون جميع اللهجات والسياقات العربية. وأشار أحد مشرفي المحتوى:

”[المعرفة باللهجات العربية] لم تكن معياراً في البداية، ولا يمكنك تغييره الآن. لذا، من الأفضل أن تستمر في غض الطرف عن ذلك، كما تعلم، وتنتظر بأنك على ما يرام. لكن، كما تعلمون، ...معظم مقاطع الفيديو، ومعظم الأخطاء تم ارتكابها، تم ذلك، بسبب عدم فهم اللغة.“ (مشرف المحتوى، أبريل ٢٠٢٤)

ج اللهجات العربية المغاربية والإشراف الآلي على المحتوى

(ACM)

تُعد الأدوات الآلية مكملاً ضرورياً للإشراف اليدوي على المحتوى، نظراً لحجم المحتوى الذي ينشئه المستخدمون حول العالم على وسائل التواصل الاجتماعي (شينكمان وآخرون، ٢٠٢١). يعد الإشراف الآلي على المحتوى (ACM) هو المستوى الأول من تحليل المحتوى.

على سبيل المثال، يستخدم فيسبوك أداة لتعيين درجات تنبؤية للمنشورات، وتقدير احتمالية انتهاكها لسياسات الإرهاب التي حددتها الشركة. ويتم إعطاء الأولوية للمشاركة ذات الدرجات الأعلى ليقوم بمراجعتها مشرفون بشريون متخصصون (جوروا وآخرون، ٢٠٢٠). على الرغم من أن المراجعين البشريين عادة ما يتخذون القرار النهائي، إلا أن أنظمة الإشراف الآلي على المحتوى يمكنها تلقائياً إزالة المنشورات أو تقليل ظهورها إذا كانوا واثقين تماماً من أن المحتوى يروج للإرهاب (جوروا وآخرون، ٢٠٢٠).

إذا لم يتمكن الإشراف الآلي على المحتوى من إصدار حكم موثوق، فسوف يقوم بتمرير العناصر إلى مشرف محتوى، الذي سيقوم بإجراء المزيد من التحقيق في الأمر واتخاذ القرارات. ومع ذلك، وفقاً لمشرفي المحتوى في تيك توك، هناك استثناء للمستخدمين المهمين، عادةً أولئك الذين لديهم مئات الآلاف من المتابعين، والذين تتم مراجعة محتوهم حصرياً بواسطة مشرف محتوى.

لاحظ مشرفو المحتوى أن الإشراف الآلي على المحتوى يرتكب أخطاء، خاصة عند تفسير اللغة العربية غير الواضحة. اللهجات العربية المغربية هي لهجات قليلة الموارد. أدت فجوة الموارد هذه إلى ظهور نماذج لا تحقق أداءً جيداً مقارنة باللغة العربية الفصحى أو حتى اللهجات الأخرى الأكثر شعبية مثل اللهجة المصرية. بالإضافة إلى ذلك، تفتقر مجموعات البيانات إلى أمثلة كافية عن العريزي والتبديل اللغوي لتدريب النماذج. يقول أحد باحثي معالجة اللغات الطبيعية الذي يعمل في إحدى شركات التواصل الاجتماعي:

”نظراً لنقص البيانات، فإن الكثير من النماذج التي تم تدريبها على اللغة العربية المغربية أو حتى اللهجة الجزائرية لا تقدم أداءً جيداً مقارنة بالنماذج التي تم تدريبها على اللغة العربية الفصحى الحديثة.“ (باحث بمعالجة اللغات الطبيعية، شركة وسائل التواصل الاجتماعي، أبريل ٢٠٢٤)

بالإضافة إلى ذلك، فإن التعليقات التوضيحية المعيبة للبيانات، وهي عملية تصنيف البيانات لتدريب نماذج الذكاء الاصطناعي، يمكن أن تؤدي أيضاً إلى أخطاء في الإشراف الآلي على المحتوى (بينز وآخرون، ٢٠١٧). وقد لاحظ العديد من الباحثين في معالجة اللغات الطبيعية المتخصصين في العربية المغربية أن الاستعانة بالمفسرين الذين لا يتحدثون اللهجات المغربية ونقص التمويل والموارد أدى إلى تناقضات وتحيزات وعدم دقة في مجموعات البيانات تؤثر هذه الأخطاء في النهاية على جودة المصنفين والإشراف على المحتوى بشكل عام، مما يؤدي إلى اتخاذ قرارات غير صحيحة وعمليات إزالة المحتوى غير المبررة.

إن عدم التنوع في فرق البحث في معالجة اللغات الطبيعية يضر أيضًا بجودة النماذج التي تستخدم في الإشراف الآلي. ذكر موظفون سابقون في شركات تقنية مختلفة أنه لا يوجد تنوع مقصود في الخلفيات اللغوية في فرق معالجة اللغات الطبيعية وأن اللغة أو المعرفة الثقافية ليست شرطًا للحصول على هذه الوظيفة. في حين أن هذه ممارسة شائعة في مجتمع معالجة اللغات الطبيعية، إلا أنها قد تؤدي إلى تصنيفات أقل صرامة للغات غير الإنجليزية منخفضة الموارد، بما في ذلك اللهجات العربية المغربية. على سبيل المثال، لاحظ العديد من صانعي المحتوى في منطقة المغرب العربي أن تطبيق إنستغرام يخفي على الفور أي تعليق يتضمن عبارة "الله أكبر"، وهي جملة شائعة الاستخدام في العالم العربي للصلاة والتعبير عن السعادة والفرح، ولكنها قد تم ربطها في بعض الأحيان أيضًا بالسياقات الإرهابية. وعندما سُئل أحد الباحثين في معالجة اللغات الطبيعية عن هذه الظاهرة، أوضح أن هذا قد يكون بسبب "عندما لا يكون لديك شخص في الغرفة يمكنه إيقاف هذه الأشياء أثناء تدريب النماذج... فهذه هي النتائج التي تحصل عليها." (باحث في معالجة اللغات الطبيعية، شركة تواصل اجتماعي، إبريل ٢٠٢٤).

التوصيات

من المهم أن تعمل شركات التكنولوجيا على تحسين إدارتها للمحتوى في اللهجات العربية المغاربية لتجنب إدامة إرث الاستعمار الذي شوّه بالفعل اللغة والثقافة والاقتصاد في هذه المنطقة. إن الأخطاء في الخوارزميات وسوء الفهم من قبل مشرفي المحتوى هي مشكلات يمكن حلها بالنوايا والجهود الصحيحة، مثل ما يلي:

١. الاستعانة بمفسرين من المتحدثين الأصليين باللغة

في الوقت الحالي، تعتمد شركات التكنولوجيا على شركات خارجية للحصول على البيانات لتدريب نماذج الذكاء الاصطناعي الخاصة بها. غالبًا ما تقوم هذه الشركات الخارجية بتعيين مفسرين ناطقين باللغة العربية من أي بلد في المنطقة لتفسير أي لهجة عربية حتي وإن لم تكن لهجتهم الأصلية. تؤدي هذه الممارسة إلى تفسير خاطئ بشكل كبير للبيانات وتؤدي إلى اتخاذ قرارات غير صحيحة تؤثر على دقة بيانات تدريب الذكاء الاصطناعي. عادةً ما يكون المفسرون أفضل في تحديد لهجاتهم الخاصة ومن المرجح أن يخلطوا بين اللهجات التي لا يعرفونها (أبو فرحة ومجدي، ٢٠٢٢). سيكون لدى المفسرين المتحدثين الأصليين باللغة فهم أعمق للسياقات الثقافية، والألفاظ البذيئة، والتعابير، وغيرها من العناصر المعقدة لغويًا، مما سيؤدي إلى تحسين جودة البيانات المستخدمة لتدريب نماذج الذكاء الاصطناعي. يتضمن ذلك فهم ظواهر مثل العريزي والتبديل اللغوي. وبالتالي، يجب على مقدمي الخدمات ضمان توظيف مفسرين من المتحدثين الأصليين باللغة ولديهم خلفيات ديموغرافية متنوعة لتقليل التحيزات والأخطاء في البيانات.

٢. إعطاء الأولوية لتوظيف فريق برمجة لغوية عصبية متنوع

يجب على شركات التكنولوجيا أيضًا إعطاء الأولوية لتوظيف فريق متنوع من علماء معالجة اللغات الطبيعية الذين يعكسون المناطق والثقافات المختلفة التي يخدمونها، بما في ذلك منطقة المغرب العربي. ومن خلال القيام بذلك، يمكن لشركات التكنولوجيا ضمان دقة أفضل للنماذج المستمدة من الفهم الشامل للفروق الثقافية الدقيقة. بالإضافة إلى ذلك، تحتاج شركات التكنولوجيا إلى الاستفادة من خبرات الباحثين المحليين في منطقة المغرب العربي الذين درسوا الخصائص الفريدة للهجات العربية المغاربية، مثل العريزي والتبديل اللغوي. ويمكن القيام بذلك من خلال تزويد مجموعات البحث المحلية بالدعم الفني والمالي.

٣. تعزيز قنوات الاتصال والدعم

أحد أكبر مخاوف المستخدمين في هذه المنطقة هو عدم قدرتهم على التواصل بشكل فعال مع شركات التكنولوجيا عند الإبلاغ عن الانتهاكات أو استئناف القرارات أو معالجة مشكلات مثل القرصنة. هذا النقص في التواصل المباشر والفعال يجعل المستخدمين يشعرون بالضعف، خاصة عند التعامل مع الأمور العاجلة أو الخطيرة. ويجب على شركات التكنولوجيا إعطاء الأولوية لإنشاء قنوات اتصال قوية من خلال مكاتبها الإقليمية لمعالجة هذه المخاوف. ويجب أن تكون قنوات الاتصال هذه سريعة الاستجابة ومراعية للاختلافات الثقافية والاحتياجات الفريدة للمستخدمين في المنطقة.

بالإضافة إلى ذلك، ينبغي أن تكون عمليات المناشدات والإبلاغ أكثر تنظيماً ووضوحاً للمستخدمين. يجب أن يكون المستخدمون قادرين على تتبع حالة بلاغاتهم ومناشداتهم وفهم الإجراءات التي يتم اتخاذها وما إذا كانت قد تمت مراجعتها بواسطة الخوارزمية أو بواسطة وسيط.

كما ينبغي لشركات التكنولوجيا أن تكون أكثر استجابة لطلبات منظمات المجتمع المدني التي تساعد المستخدمين على أرض الواقع وتدافع عنهم. وتلعب هذه المنظمات دوراً حاسماً في سد الفجوة بين المستخدمين وشركات التكنولوجيا، وتعزيز بيئة معلومات أكثر أماناً وشمولاً في منطقة المغرب العربي.

٤. تعزيز الشفافية حول إزالة المحتوى والحظر الخفي

أعرب المستخدمون في منطقة المغرب العربي في كثير من الأحيان عن إحباطهم إزاء عمليات إزالة المحتوى غير المبررة والحظر الخفي، خاصة فيما يتعلق بالمشاركات السياسية. وقد أدى عدم اليقين بشأن مصير المحتوى الخاص بهم إلى خلق جو من الرقابة الذاتية، مما دفع المستخدمين إلى الإفراط في التفكير والتدقيق في المحتوى الخاص بهم قبل نشره. وهذا يعيق التبادل الحر للأفكار والمشاعر والعواطف ويساهم في خلق مناخ من الرقابة على الإنترنت.

وبالتالي، يجب على شركات التكنولوجيا إعطاء الأولوية للشفافية فيما يتعلق بعمليات إزالة المحتوى. يجب أن تتضمن هذه الشفافية إعلام المستخدمين عند اتخاذ قرار الإشراف وسبب الإجراء. كما يجب أن

تتضمن الشفافية حالات إخفاء محتوى المستخدم. ويجب على الشركات أيضاً توضيح ما إذا كانت هذه القيود قد تم تنفيذها من قبل البشر أو الخوارزميات. ومن شأن مثل هذه الجهود أن تعمل على إعادة بناء ثقة المستخدمين في هذه الخدمات.

بالإضافة إلى ذلك، يجب على الشركات استخدام الحظر الخفي فقط في حالات نادرة. ويمكن تبريره عند الضرورة لمنع الجهات الفاعلة السيئة من إساءة استخدام الخدمة، وهو مفيد بشكل خاص في حماية المستخدمين من الرسائل العشوائية والمعلومات المضللة المنسقة والمضايقات من خلال الحسابات المزيفة. عند استخدام نظام الحظر الخفي، يجب على الشركات الكشف عن المواقف المحددة التي دفعت إلى استخدامه.

٥. التعامل مع لهجات المشرفين وخبراتهم

يعمل مشرفو المحتوى فوق طاقتهم، ويتقاضون أجوراً زهيدة، ويتعرضون لمحتوى مزعج يصيبهم بالصدمة. ونتيجة لذلك، يضطر العديد من المشرفين إلى ترك العمل لمعالجة الصدمة التي يعانون منها. يؤدي فقدان المشرفين المدربين واستبدالهم بموظفين جدد إلى زيادة تحديات الإشراف على المحتوى نظراً لأنهم يستغرقون الكثير من الوقت لفهم السياسات والتعود على أنظمة الإشراف. وقد أفاد مشرفو المحتوى أنهم كانوا سيقعون لو عُرض عليهم رواتب أعلى، وتوازن أفضل بين العمل والحياة، ودعم نفسي أفضل. يمكن لشركات التكنولوجيا تقليل معدل تغيير الموظفين من خلال تزويد المشرفين بالدعم المناسب. إن وجود مشرفين ذوي خبرة يمكنهم أداء وظائفهم بأمان سيؤدي حتماً إلى تحسين نظام الإشراف.

بالإضافة إلى ذلك، على الرغم من وجود العديد من مقدمي الخدمات في جميع أنحاء العالم العربي، غالباً ما يتم إسناد محتوى إلى المشرفين بغير لهجتهم الأصلية. يؤدي هذا إلى أخطاء عديدة في قرارات مشرفي المحتوى بسبب سوء الفهم. وبدلاً من ذلك، يجب على مشرفي المحتوى اتخاذ القرارات بشأن المحتوى بلغتهم الخاصة فقط.

وأخيراً، يجب على الشركات تقديم تفسير أفضل للسياسات حتى يتمكن مشرفي المحتوى من تنفيذها بدقة. ويجب أن يتضمن التدريب المزيد من الأمثلة الإقليمية ذات السياقات الثقافية لمساعدة مشرفي المحتوى على فهم المحتوى الذي يقومون بمراجعته واتخاذ قرارات دقيقة.



المنهج البحثي وجمع البيانات

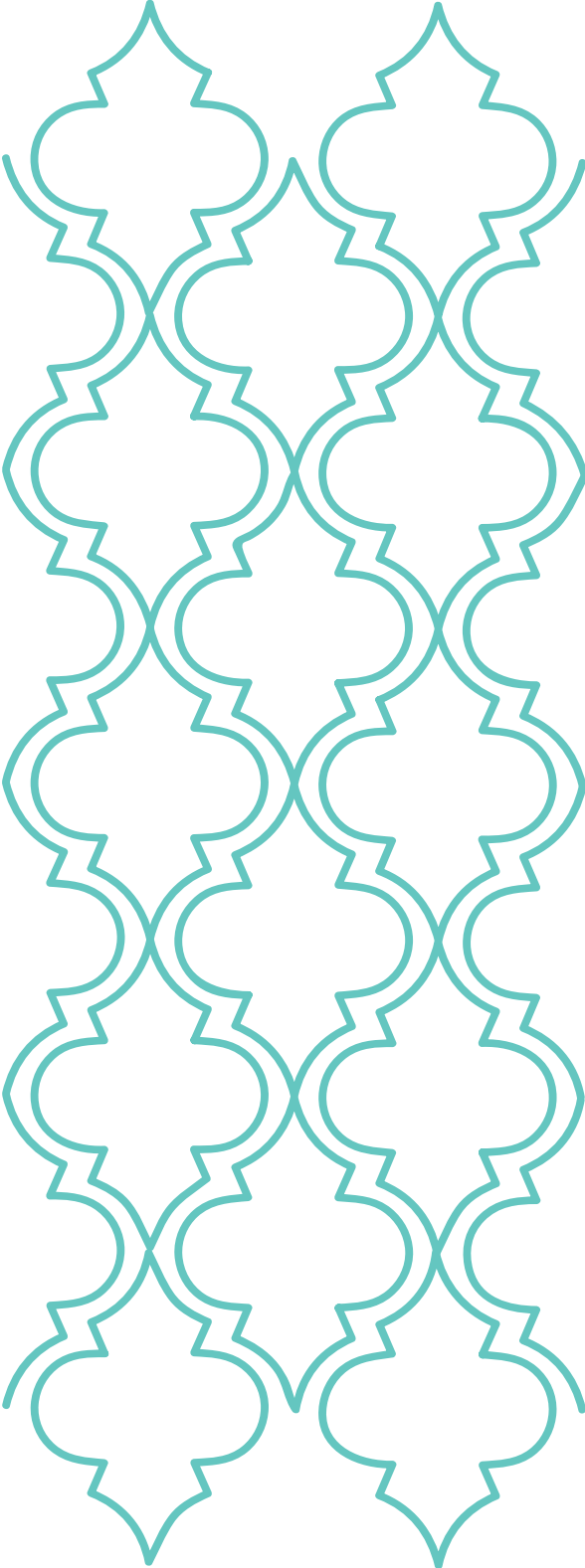
المرفقات

لدراسة أنظمة إدارة المحتوى للغة العربية المغاربية، استخدمنا منهجًا مختلطًا يجمع بين الأساليب النوعية والكمية. أولاً، أجرينا مقابلات شبه منظمة ومتعمقة لتقييم نظام المعلومات في منطقة المغرب العربي، وحوكمة شركات التكنولوجيا للغة المغاربية العربية، والتحديات التي يواجهها نشطاء الحقوق الرقمية في الدعوة إلى الاعتدال العادل للهجات العربية المغاربية. تمكنا من إجراء مقابلات مع ١٤ مشاركًا من خلفيات متنوعة. لقد أجرينا مقابلات مع ستة مشرفي محتوى حاليين وسابقين في منطقة المغرب العربي. أجرينا أيضًا مقابلات مع ستة مشاركين عملوا أو ما زالوا يعملون في فرق السياسة أو الثقة والسلامة أو فرق البحث في ميتا وتويتر ويوتيوب وتيك توك. وأخيرًا، أجرينا مقابلات مع اثنين من أبرز المدافعين عن الحقوق الرقمية من منطقة المغرب العربي.

تم إجراء جميع هذه المقابلات عبر الإنترنت. وأشرنا إلى جميع المشاركين في الدراسة بأسماء مستعارة. وقد أجريت المقابلات في الفترة ما بين مارس ومايو ٢٠٢٤. أجريت المقابلات باللغة الإنجليزية، بينما أجريت مقابلة واحدة باللغة العربية. تم تدوين ملاحظات ميدانية أثناء المقابلات، وتم نسخ تسجيلات المقابلات وتحليلها لاحقًا للعثور على مواضيع مشتركة بين المشاركين.

بالإضافة إلى هذه المقابلات، أجرينا خمس جلسات جماعية مركزة ضمت مستخدمي الإنترنت من تونس والمغرب والجزائر. وبمساعدة شريكنا الإقليمي "المواطنة الرقمية" في تونس، تمكنا من الجلوس والتحدث مع أكثر من ٢٥ مشاركًا من منطقة المغرب العربي لمناقشة استخداماتهم اليومية للخدمات عبر الإنترنت، والعقبات التي يواجهونها، ومعرفة معايير المجتمع، وكيفية التهرب من الخوارزميات. عُقدت مجموعات التركيز هذه عبر الإنترنت، واستغرقت كل جلسة ساعتين تقريبًا.

أخيرًا، أجرينا استطلاعًا عبر الإنترنت لفهم كيفية استخدام مستخدمي الإنترنت في منطقة المغرب العربي للخدمات عبر الإنترنت، والإبلاغ عن المحتوى غير المناسب، والتعامل مع عمليات إزالة المحتوى. طرح الاستطلاع أسئلة حول ثقة المستخدمين في شركات التواصل الاجتماعي وتصوراتهم حول مدى خطورة منصات الإنترنت في منطقتهم. تم استخدام منصة Alchemer لتوزيع الاستطلاع عبر الإنترنت في الفترة من ١ أبريل إلى ٥ مايو ٢٠٢٤. وتم تقديم مكافأة متواضعة قدرها ١٠ دولارات



أمريكية مقابل المشاركة. تم إجراء الاستطلاع باللغات الإنجليزية والعربية والفرنسية. تمكنا من جمع عينة من ١١١ مشاركاً من منطقة المغرب العربي. وكان حوالي ٤٧٪ من المشاركين من تونس، و ٣٠٪ من المغرب، و ٢٣٪ من الجزائر. غالبية المشاركين في الاستطلاع عرّفوا أنفسهم بأنهم إناث تتراوح أعمارهم بين ٣٥ إلى ٢٥ عاماً.

لقد قمنا أيضاً بتنظيم حلقة نقاشية مع باحثين في معالجة اللغات الطبيعية (NLP) من منطقة المغرب العربي تحديداً والعالم العربي بشكل عام للتحدث عن اساليب معالجة التحديات التي يواجهونها في جمع البيانات ومعالجتها.



المراجع

- Abu Farha, I., & Magdy, W. (2022). The Effect of Arabic Dialect Familiarity on Data Annotation. In H. Bouamor, H. Al-Khalifa, K. Darwish, O. Rambow, F. Bougares, A. Abdelali, N. Tomeh, S. Khalifa, & W. Zaghoulani (Eds.), *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)* (pp. 399–408). Association for Computational Linguistics. [perma.cc/V94X-FMEF]
- Biddle, S. (2021, October 12). Revealed: Facebook’s Secret Blacklist of “Dangerous Individuals and Organizations.” The Intercept. [perma.cc/8JG3-3UQ5]
- Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2017). Like Trainer, Like Bot? Inheritance of Bias in Algorithmic Content Moderation. In G. L. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), *Social Informatics* (pp. 405–415). Springer International Publishing. [perma.cc/AP5E-5BQT]
- Darwish, K. (2014). Arabizi Detection and Conversion to Arabic. In N. Habash & S. Vogel (Eds.), *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)* (pp. 217–224). Association for Computational Linguistics. [perma.cc/6FSR-LX9J]
- Duwairi, R. M., Alfaqeh, M., Wardat, M., & Alrabadi, A. (2016). Sentiment analysis for Arabizi text. *2016 7th International Conference on Information and Communication Systems (ICICS)*, 127–132. [perma.cc/5BNK-J2QR]
- Dwoskin, E., Whalen, J., & Cabato, R. (2019, July 25). Content moderators at YouTube, Facebook and Twitter see the worst of the web and suffer silently. *Washington Post*. [perma.cc/8A3G-X72U]
- Elsawah, M. (2023). Online tactical innovation and stagnation: Insights from the aftermath of the Arab Spring in Syria and Tunisia [Ph.D., University of Oxford]. [perma.cc/VY3Y-5ZJ6]
- Elsawah, M. (2024, January 30). Investigating Content Moderation Systems in the Global South. *Center for Democracy and Technology*. [perma.cc/8DUF-ZTJF]
- Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1). [perma.cc/Z6YG-3GZE]
- Grimmelmann, J. (2015). The Virtues of Moderation. *Yale Law & Tech*, 17, 42–109. [perma.cc/39TM-4BV4]
- Guellil, I., Saâdane, H., Azouaou, F., Gueni, B., & Nouvel, D. (2021). Arabic natural language processing: An overview. *Journal of King Saud University - Computer and Information Sciences*, 33(5), 497–507. [perma.cc/7ABJ-RPN2]
- Hajbi, S., Chihab, Y., Ed-Dali, R., & Korchiyne, R. (2022). Natural Language Processing Based Approach to Overcome Arabizi and Code Switching in Social Media Moroccan Dialect. In Y. Maleh, M. Alazab, N. Gherabi, L. Tawalbeh, & A. A. Abd El-Latif (Eds.), *Advances in Information, Communication and Cybersecurity* (pp. 57–66). Springer International Publishing. [perma.cc/362K-WGJP]

- Harrat, S., Meftouh, K., & Smaïli, K. (2018). Maghrebi Arabic dialect processing: An overview. *Journal of International Science and General Applications*, 1. [perma.cc/6KNS-EURE]
- Lorenz, T. (2022, April 11). Internet ‘algospeak’ is changing our language in real time, from ‘nip nops’ to ‘le dollar bean.’ *Washington Post*. [perma.cc/8YZW-ZNWB]
- Lorenz, T., & Nix, N. (2024, February 11). Meta turns its back on politics again, angering some news creators. *Washington Post*. [perma.cc/67K8-PG8J]
- Nicholas, G. (2022). Shedding Light on Shadowbanning. Center for Democracy & Technology. [perma.cc/577H-P23T]
- Nicholas, G., & Bhatia, A. (2023). Lost in Translation: Large Language Models in Non-English Content Analysis. Center for Democracy & Technology. [perma.cc/7LLC-5Z5S]
- Roberts, S. T. (2018). Digital detritus: “Error” and the logic of opacity in social media content moderation. *First Monday*. [perma.cc/A7VB-4Q93]
- Satariano, A., & Isaac, M. (2021, August 31). The Silent Partner Cleaning Up Facebook for \$500 Million a Year. *The New York Times*. [perma.cc/WLY6-CMV4]
- Shenkman, C., Thakur, D., & Llansó, E. (2021). Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis. Center for Democracy & Technology. [perma.cc/9DY6-8324]
- Younes, J., Souissi, E., Achour, H., & Ferchichi, A. (2020). Language resources for Maghrebi Arabic dialects’ NLP: A survey. *Language Resources and Evaluation*, 54(4), 1079–1142. [perma.cc/V7W9-5F9D]

 cdt.org

 cdt.org/contact

 **Center for Democracy & Technology**
1401 K Street NW, Suite 200
Washington, D.C. 20005

 202-637-9800

 @CenDemTech

