نهى المصرية: أداة لرصد وتصنيف خطاب الكراهية

تتعرض النساء للعديد من التعليقات المسيئة دائمًا ولحملات كراهية ممنهجة بين حين وآخر، ويؤدي تباين اللهجات العربية إلى صعوبة رصدها في ظل تعامي خوارزميات المنصات الكبرى عن السياق المحلي، لهذا السبب أخذت الجمعية الأردنية للمصدر المفتوح على عاتقها مشروع صناعة أداة "نهى" التي لا تكتفي بكشف الكراهية فحسب، بل تشرّحها وتصنّفها لتساعد المعنيين على مجابهتها. في هذا المقال نتطلع على التفاصيل الدقيقة لقصة نهى المصرية

بقلم
نادين بني ياسين

٤ آذار ٢٠٢٦

المقدمة

جاءت فكرة مشروع نهى، وهي أداة تصنيف وتحليل خطاب الكراهية الرقمي القائم على النوع الاجتماعي، من إدراكنا بأن أدوات التصنيف المماثلة في سياق المحتوى العربي شحيحة، وبالأخص تلك التي تصنف العنف ضد النساء في الفضاء الرقمي، والسبب الآخر هو أن منصات التواصل الاجتماعي العملاقة لا توفر أداة تعمل بفعالية على رصد خطاب الكراهية باللغة العربية وتقدر على أن تحلل وتفهم السياق المحلي والإقليمي وتميّز أشكال العنف المختلفة الموجهة ضد النساء. لذلك تتضمن المشروع تعليم نهى على عدة لهجات محلية وقد أنجزنا النسخة المصرية والأردنية وبدأنا بالعمل على نسخ تعتني باللهجة التونسية والعراقية والكردية. قد يثير فضولكم لماذا سمينا الأداة نهى؟ السبب هو أننا رغبنا بتسميتها باسم عربي أصيل، ومعنى نهى في اللغة العربية يرتبط بالعقل والحكمة والتمييز.

في سنة ٢٠٢١ بدأت الجمعية الأردنية للمصدر المفتوح (جوسا) بالعمل على نموذج نهى الأردنية وطورت الجمعية النموذج ليحلل خطاب الكراهية ويعطي نتائج ثنائية تبيّن ما إذا احتوى النص المكتوب باللهجة الأردنية على خطاب كراهية أم لا، ويأخذ النموذج بعين الاعتبار السياق المستخدم في المجتمع الأردني، ففي تلك الفترة رصدت جوسا كثرة خطاب كراهية على منصات التواصل الاجتماعي واستهداف العديد من حملات خطاب الكراهية للنساء، وكان الخطاب في هذه الحملات يحمل أشكالًا متعددة من العنف، فأصبحت نهى الأردنية بناءً على النموذج الذي طورناه أول أداة تصنيف وتحليل خطاب الكراهية الموجه ضد النساء على الإنترنت في اللغة العربية وبالتحديد اللهجة الأردنية، وهي بالتأكيد أداة مفتوحة المصدر ومجانية.

كانت الرؤية عندها أنه من الضروري والمهم بناء نموذج يحلل الخطاب بشكل ثنائي، بمعنى أن تتمكن هذه الأداة من التمييز بين خطاب الكراهية والخطاب المحايد، وكان الاكتفاء بهذه الثنائية لسببين رئيسيين، الأول هو افتقارنا إلى الموارد والخبرة في هذا المجال مما صعّب علينا تحديد قدرات النموذج، والثاني هو اعتبارنا بأن هذه الثنائية هي نقطة الانطلاق فقط، مع إدراكنا بأهمية تصنيف العنف على اختلاف أنواعه، وبالطبع من أجرى عملية التصنيف كُنَّ نساءً أردنيات على دراية بالسياق واللغة المستخدمة في المجتمع الاردني.

ثم جاءت فكرة نهى المصرية، وبدأت الجمعية الأردنية للمصدر المفتوح بالعمل على هذه الفكرة منذ مطلع عام ٢٠٢٥ مع الأخذ بعين الاعتبار الدروس المستفادة والخبرة والمعرفة التي اكتسبها الفريق أثناء العمل على نهى الأردنية، وكانت الرغبة حاضرة لدى فريق العمل بتحسين جودة الأداة.

اكتشفنا نتيجة الخبرة والتجربة وبعد البحث في التعريفات والأدبيات أنه لم يعد من الكافي تحليل الخطاب على ثنائية الكراهية أو الحياد فحسب بل يجب أن نفصّل أيضًا بين أنواع خطاب الكراهية المستخدم وحِدّته، وذلك لنتمكن من تسهيل الكشف عنه ومواجهته. لذلك اعتمد نموذج نهى المصرية على منهجية تتكون من عدة تصنيفات رئيسية وفرعية —وسيأتي ذكرها في آخر المقالة— والتي مكنت النموذج من تحديد نوع خطاب الكراهية المستخدم وليس فقط الكشف عن وجود خطاب كراهية بشكل عام. ولضمان الشفافية والمصداقية لهذه الأداة قمنا بوضع نسبة تأكد الاداة من نتيجة كل تصنيف تجريه لأننا نؤمن بأن الباحثين/ات والأشخاص المهتمين باستخدام نهى لديهم العدسة النقدية والتحليلية الخاصة بهم، ويمكنهم أن يميزوا ما إذا تناسب المحتوى مع التصنيف الذي تقدمه الأداة، وكذلك لنجعل الإنسان مركزيًا في العملية بدلًا من التواكل على الأداة كليًا، لهذا تَعرِض الأداة نسبة التأكد (درجة الثقة) من دقة التصنيف ويُترَك الحُكم النهائي للأشخاص وفق ما يرونه مناسبًا في كل سياق.

الأهداف الرئيسية لنهى

تزويد الباحثين/ات والنشطاء على الإنترنت والمدافعين/ات عن حقوق الإنسان بأداة لتحليل خطاب الكراهية القائم على النوع الاجتماعي، وذلك لتحليل خطاب الكراهية المكتوب باللهجة المحلية بسرعة وسهولة، وتحديد أنماط العنف في خطاب الكراهية في النص العربي الذي يستهدف النساء، ولتعمَّ الفائدة لا بد أن تكون الأداة مجانية ومفتوحة المصدر.

ومن أحد أهداف هذه الأداة هي تقليل الوقت، والجهد، والاحتراق النفسي الذي قد يتعرض له الأشخاص عند التعرض لهذا الكم من المحتوى الضار والعنيف، فتحمل نهى العبء النفسي نيابة عن الأفراد، ويبقى القرار النهائي بالنسبة للتصنيف الأدق لدى مستخدمة الأداة.

المنهجية

أولًا جرت مرحلة جمع البيانات عبر مسح صفحات على منصات التواصل الاجتماعي، ومع التركيز على حسابات الأفراد والمجموعات والمبادرات المصرية أو المقيمين في مصر، وذلك بهدف جمع التعليقات من هذه الحسابات على فيسبوك وإنستغرام. وتنوعت هذه الصفحات فبعضها يتلقى خطاب كراهية وعنف ممنهج وبعضها ينشر خطاب الكراهية ويشن الحملات المنهجة ضد النساء والأفراد، وكان المجموع الكلي ١٦٤ حسابًا.

يُعدّ حماية الحق في الخصوصية وأمن البيانات أمرًا أساسيًا في جميع مراحل تطوير ونشر واستخدام نموذج نهى. ولتقليل مخاطر انتهاك الخصوصية التزمنا بعدة اعتبارات أخلاقية في مرحلة جمع البيانات منها:

- **مبدأ تقليل البيانات:** لم تُجمع أي بيانات شخصية إلا إذا كانت مطلوبة بشكل صارم لأغراض العملية المعنية. فعلى سبيل المثال، لقد حرصنا خلال هذه العملية على أخذ التعليق فقط وأزلنا أسماء المستخدمين لحفظ الخصوصية ولم نحتفظ بأي بيانات تعريفية عند بناء أداة التصنيف.
- **مبدأ الموافقة:** وهذا يعني وجوب الحصول على موافقة واضحة ومكتوبة من الأفراد الذين استلزم الأمر جمع بياناتهم الشخصية، أي في حال الاضطرار إلى أخذ أي معلومات شخصية أو خاصة، كنا نأخذ الموافقة من المعنيين لحماية خصوصيتهم وأمانهم الشخصي ولتفادي أي إشكاليات قانونية.

نوع الحسابات التي تم جمعها: حرصنا على تنويع الحسابات فشملت الأنواع التالية:

الحسابات الشخصية: النساء (الأشخاص الذين يُعرّفون أنفسهم على أنهم فتيات أو نساء) وقد جمعنا حسابات مختلفة لنساء على اختلاف وظائفهم وأدوارهم على مواقع التواصل الاجتماعي مثل مؤثرات، ونسويات، وناشطات سياسيات، ومدافعات عن حقوق الانسان، وصحفيات، وجمعنا حسابات المنظمات والمبادرات والجماعات النسوية المختلفة التي تنشر وتبحث في قضايا النساء في مصر. وحرصنا على التنوّع في عينة البيانات لكي نستوعب الخطابات ولغة العنف المستخدمة والحملات بأشكالها المختلفة ضد النساء إذ لاحظنا التقاطع بين أشكال العنف.

كما لاحظنا بعض الاختلافات في نوعية الخطاب الموجّه ولغة العنف المُستخدَمة ضد النساء وأن الهدف من تلك الأساليب باختلاف أشكالها كان إلحاق أكبر ضرر ممكن بناءً على توجهاتهن، وكان من المهم تسليط الضوء على هذه الممارسات والأساليب.

جمعنا حسابات لمؤثرين معروفين في المجتمع المصري وأغلب ما ينشرونه على صفحاتهم هو تحريض وتنميط وتكفير وعنف تجاه المرأة باسم الدين أو العادات والتقاليد والعرف المجتمعي، وهذا بالتقاطع مع حسابات لمنصات هجوم ذكورية. وكانت هذه الحسابات تستخدم خطاب يحتوي على عنف ممنهج تجاه النساء وكان من المخيف واللافت أن هذه الصفحات لديها عدد هائل من المتابعين الذين يتفاعلون مع المحتوى العنيف، وقد تضمنت المنشورات والتعليقات خطاب كراهية وكانت عبارة عن حلقة وخطاب لا متناهي من العنف الذي لا يُبَلّغُ عنه ورصده وتسميته، وبدا لنا أن منصات التواصل الاجتماعي الكبيرة تتفلت من مسؤوليتها في رصد هذا المحتوى ولا تعتمد آلية تبليغ فعالة لحذفه.

حددنا وسوم «هاشتاغ» ذات صلة بحقوق المرأة والنسوية والإساءة عبر الإنترنت في السياق المصري على الفيسبوك وإنستغرام، باعتبارها مسارات مهمة ومحتملة لاكتشاف الحسابات والنقاشات ذات الصلة، وساعدنا هذا في البداية على تحديد هذه الحسابات المذكورة سابقًا، وطُوّرت قائمة شاملة للكلمات المفتاحية باللغتين العربية والإنجليزية لتسهيل عملية البحث عبر فيسبوك وإنستغرام، وشملت القائمة مصطلحات متعلقة بما يلي:

- الخطاب النسوي
- نسوية (Feminism)
- حقوق المرأة (Women's Rights)
- مدافعات عن حقوق المرأة (Women's Rights Defenders)
- العنف والإساءة.
- الهوية و الأدوار الجندرية.
- امرأة مصرية، فتيات مصر، المرأة في المجتمع المصري، دور المرأة.
- الخطاب المضاد: ضد النسوية، آراء معارضة لحقوق المرأة، الرجعية، الأصولية، ذكورية.

ما هي عينة البيانات التي اخترناها وجمعناها؟ ولماذا؟

تقتصر عينة البيانات على التعليقات والمنشورات من الحسابات المختارة على منصات Meta (فيسبوك وإنستغرام).

اخترنا التعليقات لأنها الوسيلة الأكثر مرونة ووضوحًا للتواصل المباشر مع صاحب المنشور وهي وسيلة تواصل علنية، ولأن مدخلات البيانات التي استخدمناها لتدريب نموذج التصنيف كانت نصوصًا مكتوبة، واخترنا التعليقات التي تحتوي على النصوص وليست تلك التي تحتوي حصرًا على الرموز التعبيرية «إيموجي» لأنها ذوات قدرات تعبيرية محدودة.

على ماذا ركزنا في عملية جمع البيانات؟

ركز فريق جمع البيانات على مسارات مختلفة لبناء وتغذية نموذج التصنيف، بهدف جمع أكثر من ٥٣ ألف نقطة بيانات (تعليق) لتدريب النموذج. استخدم فريق جمع البيانات الجمع اليدوي لمراقبة الحسابات ولاستخراج التعليقات من هذه الحسابات المذكورة سابقًا وذلك وفق نوعية الحسابات التي راقبناها، حيث مسحنا كل حساب واخترنا خمس منشوراتٍ فقط لنحرص على التنوع في البيانات ولتستطيع الأداة فهم التعليقات المختلفة وتحليلها وفهم لغة أشكال العنف المختلفة في السياق المصري. جُمعت هذه التعليقات النصية من قائمة الحسابات التي قمنا بمراقبتها بشكل دوري باستخدام أداة Export Comments التي تستخرج التعليقات من المنشورات على وسائل التواصل الاجتماعي.

ما وراء كواليس تدريب النموذج الآلي نهى

قد يبدو بالظاهر أن صنع النماذج اللغوية الآلية سهل، ولكن في الحقيقة تقف خلف هذه الآلة السريعة مئات من ساعات العمل وعشرات من الجلسات النقاشية والتحليل بشكل دوري للوصول لأدق التصنيفات، وهذا الجهد الكبير بذله فريق التصنيف بالتعاون مع فريق جوسا.

وقد بذلت المحللة والباحثة وفاء هيكل الاستشارية القائمة على المشروع جهدًا كبيرًا في تحليل وتفكيك الخطاب والسياق المصري بما يتناسب مع المشروع، كما ساهمت في بناء وتدريب نهى وفي الكتابة والتحليل وتعيين التحديات وتقديم الحلول لها وإعطاء الإرشادات.

صنفنا التعليقات وفقًا لجدول تصنيفات معتمد لدينا، ويعتبر هذا الجدول أحد المعايير الأساسية في عملية التصنيف بهدف تدريب نموذج نهى الآلي، وحتى تتمكن الآلة من فهم الخطاب العنيف فهي أيضًا تحتاج إلى فهم الخطاب غير العنيف لتميّز بينهما، ويستمد الخطاب العنيف القائم على النوع الاجتماعي المكتوب على وسائل التواصل الاجتماعي من الخطاب السائد في المجتمع، إذ يتأثر بالخطاب السياسي والديني بشكل وثيق. وقد لاحظنا أن أشكال العنف كثيرًا ما تتقاطع مع بعضها، ما يوحي بأنه لا يوجد شكل أو قالب واحد لخطاب الكراهية، بل يبدو متجذرًا أكثر مما توقعنا. وحرصنا على أن نحافظ على مفهوم حرية الرأي والتعبير، إذ أن الفصل بين مساحة حرية التعبير وخطاب الكراهية هو تحدٍ ضمن مجموعة تحديات واجهناها وسنوضحها في القسم التالي.

تحديات تصنيف محتوى العنف القائم على النوع الاجتماعي

كان من المهم والمفيد إدراك التحديات وتحديدها من بداية العمل على نهى المصرية وتدوين التحديات واستقاء الدروس المستفادة من مشروع نهى الأردنية أثناء عملية بناء النموذج من الصفر، وقد وفّر ذلك لنا أدواتًا وحلولًا استخدمناها كي نتغلب على هذه التحديات التي صادفناها خلال رحلة العمل السابقة، وقد كانت هذه التحديات الخمسة أصعبها:

١. **اختلاف اللهجات:** لا يتحدث المصريون بلهجة واحدة فقط، بل تختلف اللهجات من شمال مصر إلى جنوبها وبين شرقها وغربها، وبالإضافة لذلك يجذب المحتوى المصري على وسائل التواصل الاجتماعي العديد من المعلقين من جنسيات عربية غير المصرية مثل بلاد الخليج والشام.
القرار عما إذا كان التعليق باللهجة المصرية أم لا تُرك للمصنفات/ ين لأنهن أدرى باللهجة المصرية واختلافها، وقد ركّزن على التعليقات باللهجات المصرية حصرًا، وذلك لأن النموذج مبني لفهم وتحليل اللهجة المصرية ولا يمكننا أن نشمل لهجات أخرى دون أن يؤثر ذلك على كفاءة وأداء النموذج.
٢. **استغلال الخطاب الديني:** وجدنا استغلالًا للخطاب الديني بكثرة واستخدامه كوسيلة لنشر خطاب الكراهية، فمثلًا بعض التعليقات اكتفت بنشر آية قرآنية أو حديث نبوي لكنه كلام حق يراد به باطل، وكنا على وعي وحذر في التعامل مع هذه الحالات إذ أننا لا نريد أن نبني نموذجًا لغويًا يُصنّف النصوص الدينية كخطاب كراهية.
٣. **الخطاب العنيف الموجه ضد الرجال:** صادفتنا تعليقات تحتوي على خطاب كراهية موجه ضد الرجال واحتوت هذه التعليقات على عنف لم نستطع غض البصر عنه، فنحن ندرك تمامًا أن خطاب الكراهية وأشكال العنف الممارس على النساء يختلف بشكله سواء في لغة العنف والأثر الناتج عن العنف عن خطاب الكراهية ضد الرجال، وتناقش فريقا العمل (فريق جوسا وفريق المصنفين/ات) مسألة تصنيف خطاب الكراهية الموجه ضد الرجال وكانت النتيجة هي أنه خطاب كراهية ويجب تصنيفه.
٤. **المحتوى الساخر:** يمكن استخدام المحتوى الساخر في التعبير عن الرأي، ويمكن استخدامه لتمرير الخطاب العنيف وهذا يصعّب عملية التصنيف.
٥. **التشفير:** أحد أصعب التحديات التي واجهتنا هي عند رصد كتابة تشفّر الشتائم أو الخطاب العنيف، مثل تبديل حروف الكلمة، أو حذف نقاط وحروف، أو إضافة فواصل، أو إضافة «إيموجي».

ما هي الإرشادات التي طورناها للتغلب على هذه التحديات؟

خرجنا من إطار تصنيفات "خطاب الكراهية" إلى تصنيفات "الخطاب العام"، وذلك لتتمكن الآلة من أن تفرق بين الخطاب العنيف وغيره.

اعتمدنا على المعايير الستة التي طورتها خطة عمل الرباط[1] بشأن حظر التحريض على الكراهية والتي أعدّتها مفوضية الأمم المتحدة السامية لحقوق الإنسان (OHCHR) كنقطة مرجعية دولية واضحة لتحديد حالات التحريض على التمييز أو العنف التي تستدعي تقييد حرية التعبير، وذلك بما يتوافق مع المادة ٢٠ من العهد الدولي الخاص بالحقوق المدنية والسياسية (الفقرة ٢)[2] الذي ينص على حظر دعوة الكراهية التي تشكل تحريضًا على التمييز أو العداء أو العنف.

المعايير الستة التي نصّت عليها خطة الرباط هي:

١. **السياق الاجتماعي والسياسي:** يجب أخذ السياق في عين الاعتبار حيث تختلف التعبيرات المعينة التي يمكن أن تحرض على التمييز أو العنف ضد المجموعة المستهدفة أو يمكن أن يكون لها تأثير مباشر على النية أو العلاقة السببية على حد سواء. وينبغي وضع فعل وتأثير الخطاب في السياق الاجتماعي والسياسي السائد عند صدور الخطاب ونشره.
٢. **موقعية المتكلم:** ينبغي دراسة موقعية المتكلم أو حالته في المجتمع، وعلى وجه الخصوص مركزه الفردي أو مركز منظمته في بيئة الجمهور الذي يوجه إليه الخطاب، حيث يجب قياس حجم الضرر. على سبيل المثال تحديد ما إذا كان المتكلم شخصية ذات نفوذ وتأثير أو شخصية سياسية معروفة او اجتماعية ينشر خطابًا يحتوي على تحريض أو عنف أو كراهية ضد فئات معينة من المجتمع، وبالتالي يعظم حجم الأثر والضرر الذي نتج عن هذا الخطاب.
٣. **النية:** ربط العلاقة بين غرض الخطاب وموضوعه بالإضافة الى الجمهور المتلقي وتحديد ما إذا وجدت النية للتحريض ضد مجموعة ما.
٤. **المحتوى وشكل الخطاب:** يشكل محتوى الخطاب إحدى النقاط الأساسية التي نركز عليها وهو عنصر هام في التحريض، وقد يتضمن تحليل المحتوى مدى كون الخطاب استفزازيًا ومباشرًا، بالإضافة إلى التركيز على الشكل والأسلوب وطبيعة الحجج المستخدمة في الخطاب.
٥. **مدى نشر خطاب الكراهية:** ويتضمن ذلك عناصر مثل تأثير الخطاب، وطبيعته العامة، وكبر جمهوره وحجمه، وما إذا كان الخطاب علنيًا، وما هي وسائل النشر، والنظر فيما إذا كان الخطاب اقتصر على منشور وحيد أم أنه خطاب واسع المدى عن طريق وسائل الإعلام السائدة أو شائع عبر الإنترنت، وما إذا كان لدى الجمهور أي وسيلة للتصدي للتحريض، ونظرنا ما إذا كان البيان قد عُمم في بيئة محصورة أم مفتوحة على نطاق واسع وعلى مرأى الجميع.
٦. **أرجحية وقوع الضرر، بما في ذلك الخطر الوشيك:** ليس من الضروري ارتكاب الفعل الذي دعا إليه خطاب التحريض حتى يعتبر ذلك الخطاب جريمة، لكن يجب تحديد درجة المخاطر والضرر الناجم عنه. وما إن كان هناك احتمال معقول بأن ينجح الخطاب في التحريض على جريمة فعلية ضد المجموعة او الفرد المستهدف، مع الإقرار بأن الصلة السببية ينبغي أن تكون مباشرة.

اخترنا ودربنا فريقًا من المصنفين والمصنفات المصريين من مناطق مختلفة داخل مصر، وجرى اختيارهم بناءً على فهمهم للسياق الاجتماعي والسياسي والثقافي والخطاب التداول في مصر، وكان لهم دور محوري ومهم في عملية تصنيف التعليقات وبناء جدول التصنيف بما يتناسب مع السياق المصري، وكان لهم أيضًا دور مهم في عملية المسح للمنصات والحسابات على منصات التواصل الاجتماعي التي تنشر خطاب الكراهية ضد النساء والحسابات والصفحات التي تتعرض لخطاب كراهية، وساعدنا ذلك في عملية جمع البيانات مما وفّر علينا الوقت والجهد من الناحية التقنية.

أحد أهم الإرشادات في عملية التصنيف هي عدم اختيار التصنيف بناءً على افتراضات أو مشاعر، إذا لم يتضمن التعليق كلمات تَحُض على الكراهية، أو شتائم، أو تعبير بشكل مباشر عن العنف والكراهية، فيفضّل تصنيفه على أنه "اعتراض" أو "محايد" حسب موضع الكلام، مثل النصوص الدينية أو مثل "تم"، "ايه الهبل"، "مبتذل / لا يوجد قبول"، "مغرورة".

عند تصنيف التعليقات الساخرة، طبقنا نفس الإرشادات. إذا احتوت التعليقات على لغة تُشير الى التنمر، أو التمييز الجنسي، أو التحريض على العنف، أو التهديد بالقتل، صُنِّفت على أنها خطاب كراهية. أو على أنها "اعتراض"، هنا اعتمدنا مقياس الحدّة بالسخرية ومحتوى التعليق والكلمات التي استخدمت للسخرية.

استخدمنا خاصية الإشارة وذلك للمحتوى المشفّر الذي يحتوي على اللغة العنيفة، عندما يشفّر المعلقون الكلام للتحايل على أدوات الرصد مثل وضع «إيموجي» بين الكلام أو أرقام أو نجمة. حيث قام المصنفون باستخدام علامة مخصصة له "غير منتظم" لتمييزه من بين التعليقات العادية الأخرى، واستطاع النموذج في طوره الأخير على تحليل وتصنيف المحتوى المشفّر. و**استخدمنا خاصية الإشارة للخطاب الديني** حتى يستطيع المصنفين/ات استخدامها بدلًا من تصنيف الخطاب الديني الذي يحتوي على نصوص قرآنية.

أضفنا تصنيف "تجاهل التعليق" وذلك للتعليقات بأي لهجة غير المصرية، للحفاظ على خصوصية النموذج.

شهادات من المصنفات والمصنفين عن تجربتهم في تصنيف التعليقات وجمع البيانات لبناء "نهى المصرية"

ملاحظة: كتبن المصنفات والمصنف عن التجارب الشخصية من منظورهن وبأسلوبهن، وللحفاظ على أصالة التجارب لم يُعدِّل فريق جوسا على الشهادات أي تعديل يُذكَر.

مريم البرعي:

"خلال عملي مع الجمعية الأردنية للمصدر المفتوح (جوسا) على أداة نهى لرصد خطاب الكراهية باللهجة المصرية، شاركتُ في جمع وتصنيف التعليقات ضمن فريق واعٍ بالأبعاد الجندرية والثقافية والدينية داخل السياق المصري. وما وجدناه على المنصات لم يكن منفصلًا عن الواقع الاجتماعي بل انعكاسًا مباشرًا له، حيث تتجلى بوضوح نفس خطابات الأنظمة القمعية كالأبوية والعنصرية والرأسمالية والتي تنتج العنف والوصم ضد النساء بشكل مكثف على المساحات الرقمية.

ورغم التعرض الكثيف لخطاب الكراهية أثناء عملية الجمع والتصنيف، لم تكن التجربة صعبة نفسيًا كما توقعت. وقد ساعدني على ذلك وعيي بمنظومات القمع والنقد والتحليل النسوي لها، بالإضافة لبيئة العمل الداعمة التي أوجدها فريق التصنيف حيث استخدمنا السخرية كأداة مقاومة لخلق مسافة آمنة بين تصنيف التعليقات وسبل التفاعل معها، إضافة إلى تقسيم ساعات العمل لتجنّب الاستنزاف. هذه الاستراتيجيات لم تُلغِ تمامًا الآثار السلبية المحتملة لاستهلاك كم كبير من خطاب الكراهية الذي يتطابق مع الواقع المعاش، لكنها جعلت التعامل معه أكثر أمانًا واستدامة.

كان الوصول إلى تعليقات تتضمن خطاب الكراهية أمرًا في غاية السهولة، ما يكشف محدودية استجابة منصات التواصل الاجتماعي لخطاب الكراهية المكتوب باللغة العربية عليها، وتم ذلك عبر صفحات الأخبار المحلية والكلمات المفتاحية المتعلقة بقضايا قتل النساء وحسابات المشهورات وصانعات المحتوى، حيث وجدت تعليقات تتضمن تهديدات وتحرشات جنسية ومسبات صريحة، بالإضافة لدعوات تحريضية لممارسة القمع عبر الأسرة أو مؤسسات الدولة الرسمية.

الأمر الأكثر تعقيدًا كان آليات الخطاب المتخفي الذي يصعب رصده وتصنيفه آليًا كعبارات المدح التي تُستخدم للإشارة إلى غياب ما يتم مدحه مثل "ونعم التربية والأخلاق"، والاقتباسات الدينية خارج السياق التي تستخدم للتأثيم والتلميحات الجنسية غير المباشرة واستدعاء شخصيات خيالية سينيمائية لتوجيه اتهامات أخلاقية للنساء على المساحات الرقمية. لذلك، لا يكفي رصد خطاب الكراهية الصريحة فقط، بل يجب إيجاد وتطوير أدوات قادرة على فهم السياق بالتوازي مع بناء معرفة حقوقية وتقنية تواكب تحولات خطاب الكراهية الرقمي."

نادي فرج:

"قبل مشاركتي في تدريب نهى، كنت اتعامل مع خطابات العنف الجندري باعتبارها أحد أطياف العنف، مفترضًا ضمنيًا أن العنف اللفظي أقل خطورة من العنف الجسدي، لاسيما في تلك الخطابات الحاضرة والمعتادة في حياتنا اليومية. وأدركت بعد إعادة النظر أن هذا الافتراض ينبع من إستراتيجية كنت قد تبنيتها لأبقى مواجهًا للموضوع بطريقة تجعلني بليدًا تجاهه. حيث لم أرغب في التعايش مع وهني تجاه القوة الجارحة للغة علينا اجتماعيًا ونفسيًا.

خلال المقابلة، طُرح عليّ سؤال عن الحد الفاصل بين خطابات الكراهية وحرية الرأي والتعبير، وأجبت مترددًا بصوت متقطع من خلف الشاشة: الأذى. غير أن نهى يتبنى إطارات تحليلية وتصنيفات متفاوتة ذات مقاييس أوسع لمفهوم العنف - شاملًا جوانبه التحريضية والتهديدية والجنسية - واضعًا إياها ضمن الفئات الأشد والأعلى، على عكس الخطابات التمييزية والمهينة - مثل التأثيم والاتهام والصور النمطية الضارة والسب والتنمر- في الفئات الأدنى شدة. يتجلى لي الأمر على الصعيدين الشخصي والسياسي، ضرورة التعامل مع خطابات الكراهية لا كفئة واحدة بل كأطياف متعددة ومترابطة لآليات القمع الاجتماعي.

وبعدها جاء الجزء الأصعب وهو التعليقات المصنّفة كـ "اعتراض/رفض" خاصةً في السياقات الدينية من قبيل "إذا لم تستحي فافعل ما شئت". حيث أتاح لي تدريب نهى فرصة لإعادة التعلم ومساحة للتعمق في سياسات وأخلاقيات المعيشة داخل حيز غير مريح —ولكن مهم— لمواقف الاختلاف والاعتراض وعدم الاتفاق. وكان من المحبط التعايش في هذا الحيز خصوصًا عند استناد الفوقية الأخلاقية على الدين لإتاحة أو تبرير العنف ضد النساء والأقليات، ومع ذلك وجب الفصل بين الأذى والاختلاف - وتجنّب تضخيم هذا الأخير - لتخيل طرق مغايرة للمعيشة المشتركة مع الآخر.

ستبقى نهى مفتاحًا لبابًا يورد المزيد من الأسئلة، ومنهجية للتفكر فرادًا وجمعًا، وأحيانا للمواجهة، موفرة لنا مساحة لنتوقف ونفكك الخطابات المؤذية التي تتسرب إلى كينونتنا، خطابات عاشت مُتجاهلة ودامت في أدق تفاصيل حياتنا اليومية."

فاطمة البرعي:

"أثناء فترة عملي كمُصنِّفة على أداة نهى مع الجمعية الأردنية للمصدر المفتوح، صُدمت من الحضور الكثيف والفجّ لمختلف أشكال خطاب الكراهية الموجّه ضد النساء. لم تكن الصدمة نابعة من مواجهة أنماط جديدة من هذا الخطاب بقدر ما كانت نتيجة كثافته وتمركزه اللافت على منصات التواصل الاجتماعي. كفتاة مصرية نشأت في الأرياف، اعتدت منذ طفولتي على هذا الخطاب في الشارع، وعلى شاشات التلفاز، وحتى في المناسبات الاجتماعية؛ حيث كانت تعليقات مثل "ايه ده! تختنق كده ليه؟" أو "شوفيلك عريس بدل ما تعنسي" تُقال بلا حرج. ولم يكن هذا الخطاب معزولًا، بل يرافق أشكالًا مختلفة من الاضطهادات الجنسية والجسدية. هذا الاعتياد كان مصحوبًا بغياب شبه تام لآليات قانونية وثقافية تحمي النساء من خطاب الكراهية، سواء كان مبطنًا أو معلنًا. إلا أن الصدمة الأكبر تمثلت في سهولة انتشار هذا الخطاب في الفضاء الرقمي، وغياب أطر قانونية وتنظيمية فعّالة للحد منه، ما يسهم في بعض الأحيان في تأجيج العنف ضد النساء في الحياة الحقيقية، كما شهدنا في حالات فتيات «تيك توك»، حيث لعب ولا زال يلعب الرأي العام الرقمي دورًا مباشرًا في تقييد حرياتهن.

واجهتُ أيضًا عددًا من التحديات، كان من أبرزها استخدام الخطاب الديني عبر توظيف نصوص أو مفاهيم دينية لتبرير الكراهية وضبط سلوك النساء وإضفاء شرعية أخلاقية على أشكال من العنف ضدهن. كما برز شيوع الألفاظ المبطّنة التي تعتمد على لغة إيحائية وغير مباشرة تحمل تحقيرًا أو تهديدًا دون اللجوء إلى شتائم أو لغة تهديد صريحة، ما يجعل رصدها بالذكاء الاصطناعي ومحاسبتها بشكل شبه موضوعي أكثر صعوبة. في نهاية هذه التجربة، تعلمتُ أن الفضاء الرقمي هو وسيلة يمكن استخدامها لإعادة إنتاج العنف أو لتفكيكه، وأن المشكلة ليست في وجود هذا الخطاب العنيف ضد النساء فقط، بل في اعتيادنا عليه. ولذلك يتحمّل كل فرد منا مسؤولية مساءلة ومقاومة أشكال هذا الخطاب، سواء في الفضاءات الرقمية أو في الواقع، إذ إن نهى ليست مجرد أداة بل نموذج يجب أن يتجسد في ممارساتنا جميعًا."

منى أبو العنين:

"في بداية عملي كمُصنِّفة على أداة نهى مع الجمعية الأردنية للمصدر المفتوح، كنت أعتقد أن كوني مصرية وعلى احتكاك دائم بمنصات التواصل الاجتماعي، خاصة بالتعليقات الموجَّهة إلى النساء، سيجعل عملية التصنيف أسهل وأكثر سلاسة. إلا أن الواقع جاء مغايرًا تمامًا لهذه التوقعات.

في المراحل الأولى، واجهت صعوبة حقيقية في تحديد التصنيفات المناسبة، إذ إن غالبية التعليقات لم تكن مباشرة أو صريحة، بل جاءت في صورة تشبيهات، تلميحات، أو عبارات غير واضحة المعنى. هذا النمط من الخطاب تطلّب وعيًا عميقًا بالسياق الثقافي المصري، ومتابعة مستمرة لتغيراته السريعة، حيث تظهر يوميًا تريندات جديدة، أو كلمات تبدو عادية في ظاهرها، لكنها تُستخدم داخل سياق رقمي يحمل دلالات مهينة أو غير أخلاقية. كما أن بعض الأساليب كانت تحمل معاني تحرّش مبطّنة، مثل عبارات من قبيل "شوف الاستوري" وغيرها، وهي أمثلة لا يمكن فهمها إلا من خلال السياق.

مع مرور الوقت، أصبحت عملية التصنيف أكثر سهولة، لا سيما فيما يتعلق بالتعليقات التي تندرج تحت فئتي الاعتراض أو الحياد. في البداية، كنت أستغرق وقتًا طويلًا في تحليل كل تعليق، لكن مع زيادة عدد التعليقات المصنّفة، أصبحت أكثر سرعة في اتخاذ القرار، دون الحاجة للتوقّف طويلًا أمام كل تعليق.

في المراحل النهائية من المشروع، بدأت تظهر أنماط متكررة من التعليقات، وتمت مناقشتها ومشاركة الآراء حولها داخل فريق التصنيف، لتتحول إلى ما يشبه الكلمات المفتاحية التي تُسهِّل التعرف على التصنيف، مع الرجوع المستمر إلى الدليل الإرشادي المكتوب في بداية المشروع. ومع ذلك، أصبح واضحًا أن حصر خطابات الكراهية في تعبيرات ثابتة أمر بالغ الصعوبة، نظرًا للتطور السريع وتنوع الأساليب المستخدمة.

أما على المستوى النفسي، فلم تكن هذه الخطابات صادمة بالنسبة لي، إذ إن كثيرًا من التعليقات لم يكن غريبًا عن السياق المصري أو عن البيئة والمجتمع الذي نشأت فيه. هذا القرب المسبق من هذا النوع من الخطاب جعلني، إلى حدٍّ ما، أكثر قدرة على التعامل معه بوعي ومسافة، ومحاولة حماية حدودي النفسية قدر الإمكان أثناء العمل. ورغم قسوة بعض التعليقات، فإن النظر إليها من زاوية تصنيفية وتحليلية، ساعدني على تقليل أثرها النفسي والتعامل معها باعتبارها مادة عمل أكثر من كونها تجربة ذاتية.

في المجمل، أتاحت لي هذه التجربة رؤية خطابات الكراهية الموجَّهة ضد النساء من زاوية مختلفة تمامًا؛ زاوية لا تقوم على التلقّي أو التأثّر المباشر، بل على التحليل والتصنيف والفهم. الانتقال من موقع المتلقية إلى موقع المُصنِّفة غيّر طريقة تعاملي مع المحتوى الرقمي، وجعلني أكثر وعيًا بتعقيد هذه الخطابات، وتنوّع أساليبها، وسرعة تطورها داخل السياق المصري. هذه التجربة أعادت تشكيل نظرتي لما يُكتب على منصات التواصل الاجتماعي، ورسّخت لدي قناعة بأهمية العمل المنهجي في رصد خطابات الكراهية بدل الاكتفاء بالتعامل معها كجزء اعتيادي من الواقع الرقمي."

من سيستفيد من نهى وكيف؟

- الباحثون والمؤسسات البحثية.
- الصحفيون والمؤسسات الإعلامية .
- المنظمات النسوية ومنظمات حقوق الانسان.
- مجموعات المتطوعين التي تعمل على مكافحة خطاب الكراهية أو التحرش الجنسي الرقمي أو العنف الممارس ضد النساء والفتيات في الفضاء الرقمي.
- الجهات الرسمية في مصر والمهتمة برصد محتوى العنف القائم على النوع الاجتماعي خلال الفترات الانتخابية، مثل:
  - الهيئة الوطنية للانتخابات.
  - مراكز استطلاع الرأي.

نموذج نهى يحمل العبء النفسي والاحتراق الذي قد يتعرض له الباحثين/ات الذين يتعاملون مع خطاب كراهية و محتوى ضار وعنيف بشكل متكرر في مجال عملهم في الفضاء الرقمي، حيث يوفر النموذج قدرة معالجة أكبر للبيانات، مما يسمح بتحليل كمية كبيرة منها في وقت قصير، ويسهّل هذا على الحصول على نتائج أكثر دقة في الدراسات التي تعتمد على تحليل محتوى منصات التواصل الاجتماعي، وتنتج مخرجات وصفية إحصائية قابلة للاستخدام في:

- التحليل البحثي
- التقارير الصحفية
- حملات المناصرة

أداء النموذج

نموذج نهى يصنف خطاب الكراهية بنسبة عالية من النجاح ٧٧,٣٪ والدقة بنسبة ٧٩,١٪ والاسترجاع (Recall) بنسبة ٧٧,٣٪ ومؤشر ال F1[3] بنسبة ٧٧,٨٤٪ وتدل هذه المؤشرات على نجاح وفاعلية النموذج من ناحية تقنية حيث حقق النموذج الهدف المرجو منه. على عكس النماذج العربية الأخرى المُصَنِّفة لخطاب الكراهية، تختص نهى باللهجة المصرية، وتستطيع أن تفهم السياق واللهجة المصرية، وتستطيع أن تصنف خطاب كراهية وتفرّق بين خمسة فئات رئيسية وعشر فئات فرعية، وهي:

١. العنف: ويتفرّع إلى تهديدات مباشرة بالعنف أو القتل | التحريض على العنف أو القتل / استدعاء السلطات | العنف الجنسي.
٢. محتوى جنسي: ويتفرّع إلى التحرّش الجنسي اللفظي | الشتائم الجنسية.
٣. اللغة التمييزية أو المهينة: ويتفرّع الى التأثيم أو الاتهام | اللغة التي تحتوي على الشتائم والإهانة | صور نمطية ضارة.
٤. الاعتراض/الرفض
٥. محايد

أخيراً وليس آخرًا، سيتطور أداء نهى مع الاستخدام المستمر من قبل مجتمع الباحثات والباحثين، وهم يمثلون الفئة المستفيدة الأولى والأهم!

الخاتمة

يعد نموذج نهى أول أداة تصنيف لخطاب الكراهية القائم على النوع الاجتماعي باللهجات المحلية. وتستطيع نهى تحليل السياق وتصنيف أشكال مختلفة من العنف ورصد اللغة المشفرة المستخدمة للتحايل على أدوات الرصد الأخرى، والتي تعتبر شحيحة نظرًا لأنه لا يوجد أدوات رصد قادرة على فهم وتحليل سياقاتنا المحلية واللغات واللهجات المختلفة في كل سياق عربي، وبطبيعة الحال لا يمكننا وضع لهجات أو لغات أو سياقات شتى في قالب واحد وأن نحللها باستخدام أداة رصد واحدة لأن كل سياق له خصوصيته، وتطبيقًا لذلك بنى فريق جوسا نماذج مختلفة لتحليل خطاب الكراهية على حسب البلد واللغة والسياق. وتشارك فريق جوسا مع الجمعيات والمؤسسات النسوية المحلية في كل بلد لتقديم الدعم التقني في مرحلة بناء النموذج وتدريبه، ويترك اختيار المصنفين/ات وبناء جدول التصنيفات والتصنيف للمؤسسات المحلية في البلد المعني لتصنيف وتحليل خطاب الكراهية القائم على النوع الاجتماعي.

يمكنكم الآن تجربة واستخدام نهى في عملكم فهي مفتوحة المصدر ومجانية الاستخدام:
https://nuharesearch.com/analyze

---

[1] https://www.ohchr.org/en/freedom-of-expression
[2] https://www.ohchr.org/en/freedom-of-expression
[3] مقياس F1 هو مقياس لتقييم التعلم الآلي يقيس دقة النموذج. وهو يجمع بين درجات الدقة والاستدعاء للنموذج.

بقلم
نادين بني ياسين

عن جوسا
عن جوسا
فريق جوسا
الشركاء والداعمون
الوظائف
قواعد السلوك
الشفافية
استراتيجيتنا
سياسة الخصوصية

عملنا
البرمجيات مفتوحة المصدر
العتاد مفتوح المصدر
الإنترنت المفتوح
الخصوصية
الأمن الرقمي
المعرفة المفتوحة
الحكومة الشفافة
مدونة جوسا
الفعاليات
الأبحاث والمنشورات

تواصل
اتصل بنا
دعم الأنشطة الطلابية
jordanopensource/
josa@josa.social@
jordanopensource/
jo_osa@
jordanopensource/
jordanopensource@

الجمعية الأردنية للمصدر المفتوح

محتوى هذا الموقع مرخص برخصة المشاع الإبداعي نسب المصنف - الترخيص بالمثل 3.0 دولي (CC-BY-SA 3.0)، إلا إذا ورد خلاف ذلك.