إرشادات الشرح

معلومات أساسية:

نحن نبتكر طرقًا أفضل لاختبار نماذج التعلم الآلي من أجل اكتشاف خطاب الكراهية عبر الإنترنت باستخدام الذكاء الصناعي. لهذا الغرض، قمنا بتجميع مجموعة بيانات تتكون من بضعة آلاف من المدخلات النصية القصيرة. يعكس كل إدخال جانبًا معينًا من جوانب الكراهية على الإنترنت. قمنا باستخدام اللهجة المصرية لأنها مستخدمة بشكل كبير على الانترنت.

المطلوب منك:

1. حدد ما إذا كان كل إدخال مخصص لك يحض على الكراهية أم لا.
2. ضع علامة على أي إدخالات تعتقد أنها غير واقعية.

ماذا نعني بكلمة "كره"؟

نحن نعرّف الكراهية على أنها إساءة تستهدف مجموعة محمية أو لأعضائها لكونهم جزءًا من تلك المجموعة. تستند المجموعات المحمية إلى العمر والإعاقة والعرق (اللون والجنسية والأصول العرقية أو القومية) والدين أو المعتقد والجنس والتوجه الجنسي بالإضافة إلى الهوية الجنسية.

ماذا نعني ب "غير واقعي"؟

نترك هذا الأمر لك عمدًا. لم ننشئ أي إدخالات غير واقعية ، لذا فإن هذه العلامة الاختيارية تهدف فقط إلى ضمان جودة البيانات.

بعض الأشياء التي يجب وضعها في الاعتبار أثناء التعليقات التوضيحية:

- **يمكن أن يكون للغة البغيضة في بعض السياقات استخدامات غير كريهة** (على سبيل المثال ، الخطاب المضاد الذي يشير إلى الكراهية: "ليس من المقبول وصف السود بالعبيد.")
- **الهدف من الإساءة مهم لمعرفة إذا ما كانت تحض على الكراهية أم لا**. الإساءة بين الأشخاص وإساءة المعاملة ضد الجماعات غير المحمية مثل المهن والانتماءات ليست كراهية. (على سبيل المثال ، عبارة "أنا أكره اليهود" بغيضة ولكن "أنا أكرهك" ، "أنا أكره مائدتي" و "أنا أكره الأطباء" ليست كذلك).

أجزاء أخرى من الإرشادات:

- لا تضع علامة على الإدخالات على أنها غير واقعية لمجرد أنك تعتقد أنه من غير المرجح أن تظهر على الإنترنت كثيرًا. ضع علامة فقط على الإدخالات غير المنطقية وغير الصحيحة نحويًا.
- يرجى إكمال التعليقات التوضيحية بشكل مستقل وعدم التحدث عنها مع الآخرين.
- معظم العبارات قصيرة جدًا لذا من فضلك لا تفرط في التفكير فيها.

شكرا جزيلا لجهودك!