

# A Multi-Platform Arabic News Comment Dataset for Offensive Language Detection

Shammur A Chowdhury, Hamdy Mubarak, Ahmed Abdelali,  
Soon-gyo Jung, Bernard J Jansen, Joni Salminen

Qatar Computing Research Institute  
Hamad Bin Khalifa University, Doha, Qatar  
{shchowdhury, hmubarak, aabelali, sjung, bjansen, jsalminen}@hbku.edu.qa

## Arabic Annotation Guidelines for Offensive Post Detection and its Translation

### Provided Guideline (Arabic):

هذا مشروع بحثي يهدف لتمييز التعليقات التي تحتوي على لغة غير مقبولة. سيتم إظهار تعليقات تم اختيارها بشكل عشوائي، والمطلوب تمييز ما إذا كانت تحتوي على لغة غير مقبولة (كلمات بذيئة أو لغة عدائية) كما هو موضح في الأمثلة التالية. نظرا لطبيعة هذا المشروع، نعتذر عن وجود أمثلة وتعليقات بها ألفاظ نابية!

تشمل اللغة غير المقبولة ما يلي:

#### 1- ألفاظ بذيئة

مثال: عرص، قحبة، شرموط، خرا، اتفو

#### 2- لغة عدائية أو هجومية سواء كانت بشكل صريح أو خفي وهذا يشمل:

- لغة التهديد والتحريض على العنف، مثل: احرقوا مقرات المعارضة، هذا المنافق يجب قتله

- لغة التحقير والإهانة والتقليل من شأن الآخرين، مثل:

التشبيه بالحيوانات: يا كلب يا بهيم يا جربوع، كل تين

التلفظ على أفراد العائلة: أختك بلا شرف، أبو اللي جابك، يا روح أمك

الجنس: بس يا ديوث يا شاذ يا وسخ

الكفر: يا كافر يا ابن الملحد يا زنديق

الكره العنصري: يا عبد يا خادم يا زنجي

السباب بالدين: يلعن دينك

الإهانة: يا سافل يا عديم التربية يا وقح يا حقير

الأمراض: يا جربان يا أعمى يا مجنون

اللعن: الله يلعنك، الله يحرقك في نار جهنم

المذهب أو الطائفة أو الحزب: يا شيوعي، يا إخواني

شكرا لكم على مراعاة الدقة والمشاركة في هذا البحث الهام

مع تحيات فريق البحث

---

### **Translated Guideline (English):**

**This is a research project aimed at marking comments that contain unacceptable language.**

Randomly chosen comments will be shown, and it is required to distinguish whether they contain **unacceptable language (obscene words or hostile language)** as shown in the following examples. Due to the nature of this project, we apologize for the presence of examples and comments with profanity!

Unacceptable language includes:

#### **1- Profanity**

Example: Ars, Kahba, Shamroot, Khera, Atfou

#### **2- Hostile or offensive language, whether explicit or implicit.**

This includes:

- The language of threats and incitement to violence, such as:

Burn the headquarters of the opposition, this hypocrite must be killed

- The language of humiliation and insulting others, such as:

**Similarities to animals:** O dog, beast, O jerboa, all hay

**Pronouncing family members:** Your sister without honor, Father who beget you, O soul of your

mother

**Sex:** Enough O cuckold, gay, dirty

**Infidelity:** O infidel, son of an atheist, heretic

**Racial hatred:** O servant, butler, nigger

**Cursing with religion:** Curse your religion

**The insult:** O foolish, non behaving, shameless, despicable

**Diseases:** O mangy, O blind, O insane

**Cursing:** God curse you, God burns you in Hellfire

**Doctrine, caste, or party:** O Communist, O Ikhwangy

Thank you for taking into account the accuracy and participation in this important research

Greetings from the research team