

## Richtlijnen voor annotatie

### **Achtergrond:**

We ontwikkelen een verbeterde manier om modellen te testen die automatisch haatdragende taal detecteren. Hiervoor hebben we een dataset gemaakt die bestaat uit enkele duizenden korte tekstfragmenten. Ieder fragment reflecteert een bepaald aspect van online haat.

### **Jouw taken:**

1. Geef bij ieder fragment dat jou is toegewezen of het **haatdragend** of **niet haatdragend** is.
2. Markeer alle fragmenten die jij **onrealistisch** acht.

### **Wat bedoelen we met “haat”?**

Wij definiëren **haat** als **beledigingen gericht aan een beschermde groep of aan diens leden omdat ze deel uitmaken van die groep**. Beschermde groepen zijn gebaseerd op leeftijd, handicap, ras (kleur, nationaliteit, etniciteit of nationale oorsprong), religie of geloof, geslacht en seksuele oriëntatie evenals genderidentiteit.

### **Wat bedoelen we met “onrealistisch”?**

Dit laten we met opzet aan jou over. We hebben geen fragmenten opgesteld om onrealistisch te zijn, dus deze markering is enkel bedoeld om de datakwaliteit te waarborgen.

### **Houd hier rekening mee tijdens het annoteren:**

- **Taal die haatdragend is in bepaalde context kan ook niet haatdragend gebruikt worden** (bijvoorbeeld tegenspraak op haatdragende scheldwoorden: *“Het is echt niet oké om mensen negers te noemen”*)
- **Het doelwit van beledigingen is belangrijk bij het bepalen of iets haatdragend is.** Interpersoonlijke beledigingen en beledigingen jegens niet-beschermde groepen zoals beroepen of hobby's zijn geen haat. (bijvoorbeeld *“Ik haat Joden”* is haatdragend maar *“Ik haat jou”*, *“Ik haat mijn tafel”* en *“Ik haat dokters”* zijn dat niet).

### **Overige belangrijke richtlijnen:**

1. Markeer fragmenten niet als onrealistisch alleen omdat jij denkt dat het onwaarschijnlijk is dat ze online verschijnen. Markeer enkel onzinnige en grammaticaal incorrecte fragmenten.
2. Voltooi je annotaties alsjeblieft onafhankelijk en praat er niet over met anderen.
3. De meeste fragmenten zijn vrij kort dus denk er niet te veel over na

**Enorm bedankt voor je bijdrage!**