

LINEE GUIDA

La delineaione di linee guida per l'annotazione è stata necessaria per definire in maniera esaustiva i parametri di ricerca - *hate speech*, *aggressività*, *offensività*, *stereotipo*, *target* – e fornire degli esempi di chiarimento. Infatti, dal momento che l'*hate speech* è un fenomeno molto vario, con numerose caratteristiche e un'ampia casistica, è stato necessario definire in maniera chiara dei parametri di annotazione, orientando la ricerca solo su alcuni specifici elementi.

La presente analisi si è focalizzata su cinque categorie – *hate speech*, *aggressività*, *offensività*, *stereotipo* e *target* – ispirate all'annotazione dell'*hate speech* razzista fatta da Poletto et al. (2017), opportunamente rielaborate per adattarsi al nuovo contesto, come si vedrà nei seguenti paragrafi. Le categorie di annotazione sono applicate a tutti i commenti e sono indipendenti l'una dall'altra. Di conseguenza, nell'annotazione, le categorie possono applicarsi ai commenti in diverse combinazioni. Per esempio, un commento può risultare caratterizzato dalla presenza di stereotipo, ma senza essere classificato come *hate speech*, né risultare aggressivo o offensivo.

Esempio 1: *Le donne sono sempre più belle degli uomini.*

Similmente, un commento può essere annotato come *hate speech* e aggressivo, ma senza presentare elementi offensivi o stereotipati.

Esempio 2: *Amber deve buttarsi giù da un ponte.*

Inoltre, aggressività, offensività e stereotipo possono presentarsi anche in assenza di *hate speech*.

Esempio 3: *Dovrebbero entrambi finire in galera, tanto sono psicopatici come tutti gli attori.*

Hate Speech

Definiamo l'*hate speech* (abbreviato in HS), riprendendo la definizione di Faloppa (2020), come:

- L'istigazione, la promozione o l'incitamento alla denigrazione (es: *che uomo vomitevole*), all'odio (es: *gli uomini come Depp sono da scuoiare*) o alla diffamazione (es: *tutte queste donne che denunciano violenze lo fanno solo per soldi*) nei confronti di una persona o di un gruppo di persone.
- Sottoporre a soprusi (es: *dovremmo prenderla e sbatterla in galera*), molestie (es: *spero che gli altri galeotti facciano la festa a Depp in cella*), insulti (es: *quella è una vera troia/lui è un porco*), stereotipi negativi (es: *sei una femminista isterica*) o minacce e auguri di morte (es: *spero che Heard si ammazzi; Depp andrebbe linciato*) tale persona o gruppo.
- Può comprendere la giustificazione di queste varie forme di espressione, fondata su una serie di motivi, quali la "razza", il colore, la lingua, la religione o le convinzioni, la nazionalità o l'origine nazionale o etnica, nonché l'ascendenza, l'età, la disabilità, il sesso, l'identità di genere, l'orientamento sessuale e ogni altra caratteristica o situazione personale.

Nel caso di questo corpus, si presentano per lo più giustificazioni legate al sesso/identità di genere (es: *Depp è un porco come tutti gli uomini/Heard è una donnetta isterica*) e alla salute mentale (es: *che razza di psicopatico; questi sono tutti matti da legare*).

L'*hate speech* si differenzia dalle opinioni, le quali esprimono un punto di vista, negativo e/o sgradevole, che può essere più o meno condivisibile, mentre il linguaggio d'odio è degradante nei

confronti della dignità del target, anche tramite l'utilizzo di slur¹, e/o incita alla violenza nei confronti del target stesso.

Esempio 4: *Amber sta facendo di tutti per avere dei soldi!*

HS no; è un'opinione negativa, non è HS perché non è degradante e ammette replica.

Esempio 5: *Amber è una puttana*

HS yes; linguaggio d'odio degradante che distrugge il terreno di comune confronto.

Viene dunque considerato hate speech quel linguaggio d'odio che presenta le seguenti caratteristiche (può essere presente solo una delle due oppure entrambe):

- È degradante nei confronti della dignità del target

Esempio 6: *Johnny Depp è un porco pervertito* → HS yes

Esempio 7: *Amber Heard è vomitevole* → HS yes

Esempio 8: *Depp è proprio uno stronzo* → HS yes

Esempio 9: *La Heard è una troia* → HS yes

- Incita alla violenza nei confronti del target

Esempio 10: *Dovrebbero sbattere in galera Amber Heard* → HS yes

Esempio 11: *Depp dovrebbe ammazzarsi* → HS yes

Esempio 12: *Dovremmo proprio insegnare alla Heard cos'è la vera violenza domestica* → HS yes

Nell'annotazione, si fa una polarizzazione binaria tra presenza di hate speech (yes) e assenza di hate speech (no).

Esempio 13: *Una frustrata del cazzo, questa signorina Amber Heard* → HS yes

Esempio 14: *Una megera la Heard... Semplicemente una che ci sta con tutti...* → HS yes

Esempio 15: *Brutta falsa vipera! Ben ti sta!* → HS yes

Esempio 16: *Non è male come attrice.* → HS no

Esempio 17: *Povero Johnny, è finito tra le mani di una strega.* → HS no

Esempio 18: *Ci sono donne che non si accontentano mai. Sono disgustato.* → HS no

Aggressività

Si ricerca nei commenti l'intenzione dell'user di essere aggressivo e violento nei confronti del target di riferimento (vedi categoria "*Target*"), arrivando anche all'incitamento alla violenza.

Generalmente, l'aggressività è associata a un'impressione di fisicità, al dire e fare qualcosa come, appunto, nel caso di riferimenti ad aggressioni fisiche o ad altre azioni che hanno un impatto negativo sul corpo del target. Non va assolutamente confusa con l'ostilità e/o l'offensività dell'user, che non implica necessariamente la presenza di una componente aggressiva. Nell'annotazione, si fa una categorizzazione ternaria tra: assenza di aggressività (*no*), aggressività debole (*weak*), aggressività forte (*strong*).

Nell'aggressività debole, il commento evoca una qualche forma di reazione fisica verso il target, ma utilizza toni non troppo accesi:

Esempio 19: *Questa donna vuole fama pubblicità e soldi di traumi non ne ha avuti che si vada a*

¹ Per *slur* si intende un termine dispregiativo nei confronti di una persona o di una categoria di persone. L'utilizzo di *slur* in frasi o commenti non implica che si stia facendo uso di *hate speech*.

Esempio: *Amber Heard è una troia* → HS yes

Ma porca troia, che brutto processo! → HS no.

nascondere. → HS no, aggressività weak

Esempio 20: *Deve andare in galera* → HS yes, aggressività weak

Esempio 21: *Dovrebbe solo finire in galera, come attrice non vale niente.* → HS yes, aggressività weak

Esempio 22: *Grande! Adesso la falsa vada a lavare scale! Gli sta solo che bene!* → HS no, aggressività weak

Esempio 23: *Questa è da chiudere in una stanza coi materassi al muro e gettare la chiave.* → HS yes, aggressività weak

Nell'aggressività forte è evidenziato un implicito o esplicito riferimento ad atti di violenza nei confronti del target di riferimento.:

Esempio 24: *Cacciamo la Heard!* → HS no, aggressività strong

Esempio 25: *C'è da prenderla a sberle!* → HS yes, aggressività strong

Alcune precisazioni:

- Invocare le manette, il rinchiodare in galera o l'applicazione di interventi fisici restrittivi sulla persona, in maniera più o meno diretta, oppure smorzati dall'uso del condizionale o altri elementi attenuativi (vorrei/ potrebbe/ forse), viene catalogato come hate speech yes e aggressività weak.

Esempio 26: *Deve andare in galera.* → invocazione diretta → HS yes, aggressività weak

Esempio 27: *TSO per Amber!* → invocazione indiretta, il verbo è assente (frase nominale) → HS yes, aggressività weak

- Invocare la morte della persona, in maniera più o meno ironica, viene catalogato come hate speech yes e aggressività strong

Esempio 28: *Devono spararle* → invocazione non ironica → HS yes, aggressività strong

Esempio 29: *Al rogo! XD* → invocazione ironica → HS yes, aggressività strong

Offensività

Sottolinea l'intenzione dell'user di infierire, in maniera degradante e violenta, sul target di riferimento. Nell'annotazione si fa una categorizzazione ternaria tra: assenza di offensività (*no*), offensività debole (*weak*) e offensività forte (*strong*).

Nell'annotazione dell'offensività, si usano come parametri due elementi:

- l'uso del termine di paragone con elementi degradanti

Esempio 30: *Attore / attrice di merda* → termine degradante, HS yes, offensività strong

Esempio 31: *Pessima attrice / pessimo attore* → termine non degradante, HS no², offensività weak

- la forma del verbo, poiché l'utilizzo della 2a persona singolare è più forte e diretta rispetto alle altre)

Esempio 32: *Amber sei penosa!* → uso della 2a persona singolare, HS yes, offensività weak

Esempio 33: *Johnny, come attore fai cagare!* → uso della 2a persona singolare, presenza termine degradante, HS yes, offensività strong

NB: poiché Amber Heard e Johnny Depp sono due figure pubbliche la cui notorietà è dovuta

² Da solo, tale giudizio può essere offensivo, ma non è così degradante da qualificarsi come HS, né è una minaccia.

soprattutto al loro mestiere di attori, in questa annotazione si è deciso di non separare la loro persona dal loro mestiere. Infatti, molti utenti non operano questa separazione e tendono a far coincidere le persone di Heard e Depp con il loro mestiere di attori e/o con i personaggi che hanno interpretato. Di conseguenza, le critiche e gli insulti rivolti alla loro figura di attori vanno trattati come se fossero rivolti alla loro persona.

Esempio 34: *Depp come attore non vale granché* → HS no, offensività weak

Esempio 35: *La Heard fa cagare come attrice* → HS yes, offensività strong

Nell'offensività debole si fa riferimento a caratteristiche offensive e spiacevoli espresse in tono non eccessivamente ostile:

Esempio 36: *Recita meglio in aula che sul set* → HS no, offensività weak

Esempio 37: *A me questa pare abbastanza disturbata* → HS no, offensività weak

Esempio 38: *Heard è indifendibile. Vergognosa* → HS no, offensività weak

Esempio 39: *Questa qui è disturbata, va sbattuta in galera* → HS yes³, offensività weak

Nell'offensività forte sono rintracciabili tono ed espressioni oltraggiose, violente e degradanti, esclamazioni volgari e l'uso degli slur.

Esempio 40: *Ha sposato Satana, poveraccio* → HS no⁴, offensività strong

Esempio 41: *Che schifo di donna si dovrebbe solo che vergognarsi* → HS yes, offensività strong

Esempio 42: *Ma ci sei o ci fai o sei la classica femminista del cazzo* → HS yes, offensività strong

Alcune precisazioni:

- Nel caso di riferimenti a malattie mentali, se il riferimento è clinico (ad esempio: psicopatica, borderline) non viene considerato hate speech, ma catalogato con offensività weak, salvo uso di rafforzativi o di escalation aggressiva. Spesso, il riferimento alle malattie mentali è catalogato con stereotipo yes, soprattutto se legato al genere (es: uso del termine isterica).

Esempio 43: *Sta tizia è psicopatica ai massimi livelli.* → HS no, offensività weak

Esempio 44: *Amber è una borderline e psicopatica.* → HS no, offensività weak

Esempio 45: *Questa è da TSO... è una nullità* → HS yes⁵, offensività weak

- Se il riferimento è invece denigratorio (ad esempio: pazza, fuori di testa), allora viene catalogato come hate speech con offensività weak.

Esempio 46: *Questa pazza fa paura è fuori di testa, lui più di lei a starci insieme* → HS yes, offensività weak

Esempio 47: *Chiaramente la pazza ha perso per chi non lo sapesse* → HS yes, offensività weak

Esempio 48: *Amber è una pazza scatenata, completamente fuori di testa!* → HS yes, offensività weak

Stereotipo

Indica la presenza di preconcetti e pregiudizi generalizzati, frasi fatte e considerazioni per sentito dire, spesso infondati, eventualmente caratterizzati da aggressività e offensività. Nell'annotazione, si ha una divisione binaria tra presenza di stereotipo (*yes*) oppure assenza di stereotipo (*no*).

³ HS yes per il riferimento a interventi fisici restrittivi, come lo sbattere qualcuno in galera.

⁴ Per quanto un paragone con Satana sia forte e pensato per essere offensivo, non è così degradante da qualificarsi come HS.

⁵ HS yes per il riferimento a interventi fisici restrittivi, come appunto il TSO.

In questo caso, l'analisi si focalizza particolarmente sugli stereotipi sessisti, ossia che producono considerazioni generalizzate e spesso infondate in relazione al genere.

Esempio 49: *Una donna così bella ha ragione a prescindere, anche se ha torto marcio.* → HS no, stereotipo yes

Esempio 50: *L'errore decisivo? È che è una donna...* → HS no, stereotipo yes

Esempio 51: *L'avvocata di Johnny Depp si vede che è professionale e soprattutto gnocca.* → HS no, stereotipo yes

Esempio 52: *Le donne sono sempre più perfide degli uomini* → HS no, stereotipo yes

Esempio 53: *Tanto si sa che gli uomini non si sanno controllare* → HS no, stereotipo yes

Esempio 54: *Le donne come lei sono solo delle troie affamate di soldi* → HS yes, stereotipo yes

Esempio 55: *Dovremmo sparare a Depp* → HS yes, stereotipo no

Esempio 56: *La Heard è una pazza isterica* → HS yes, stereotipo yes

Target

Designa il destinatario o l'oggetto del commento. Nell'annotazione sono indicati quattro target diversi: Amber Heard (AH), Johnny Depp (JD), Camille Vasquez (CV), i singoli utenti (user) che si taggano a vicenda, indirizzandosi commenti. Nel caso in cui vi siano target non specifici, l'etichetta applicata è ND. Qualora siano indicati più target contemporaneamente, vanno etichetta-ti in ordine alfabetico, separati da virgole.

Esempio 57: *Ti sta bene stregaccia* → AH

Esempio 58: *Grandissimo cinghialotto dei Caraibi!* → JD

Esempio 59: *Una donna davvero con le palle un grande avvocato!* → CV

Esempio 60: *Ma ci sei o ci fai o sei la classica femminista del cazzo?* → user

Esempio 61: *Questa pazza fa paura è fuori di testa, lui, più di lei a starci insieme!* → AH, JD