

Padova, 26 marzo 2024

## **ARRIVA NET GUARDIAN L'APP CHE SEGNALE IL CYBERBULLISMO** **Università di Padova e Fondazione Carolina Onlus mettono a punto uno strumento per** **misurare il rischio di cyberbullismo tra i giovani nelle conversazioni digitali**

Il cyberbullismo, che sempre più spesso ha origine in ambito scolastico, rappresenta una fonte non trascurabile di costi per il sistema economico, sociale ed educativo del Paese. Per contrastare il problema, Il Dipartimento di Filosofia, Sociologia, Pedagogia, Psicologia Applicata - FISPPA dell'Università degli Studi di Padova, insieme a Fondazione Carolina Onlus, ha lanciato il progetto Net Guardian, finanziato da Fondazione TIM con il bando Call for Ideas- Ricerca e Istruzione.

**Il progetto, avviato a febbraio 2024 e che si concluderà a giugno 2025, ha ricevuto un finanziamento di circa 230.000 euro per realizzare una App a disposizione di studenti, servizi scolastici e associazioni attive nella prevenzione al cyberbullismo, basata su un modello di Intelligenza Artificiale in grado di misurare l'esposizione al rischio di violenze on line nelle conversazioni digitali attraverso NLP e metodologia MADIT.**

Il progetto prevede l'individuazione e la classificazione dei rischi su tutte le piattaforme di messaggistica per attivare interventi mirati in base alla gravità dell'aggressione verbale rilevata: l'App valuta e categorizza il linguaggio utilizzato su una scala da 1 a 10, dove l'1 rappresenta un minimo rischio e il 10 rappresenta il massimo rischio.

L'uso positivo della tecnologia ha uno spazio rilevante nel progetto: si predispone, infatti, la creazione di un algoritmo di Machine Learning per l'analisi automatizzata del testo, che utilizza il modello di Deep Learning BERT, e la generazione dell'indice di rischio attraverso un output associato a specifici valori emersi dall'analisi del linguaggio naturale.



*Gian Piero Turchi*

«L'App prevede anche una Chatbox, che registra l'input (testo o screenshot di chat) e restituisce un output in valore numerico misurando il rischio di cyberbullismo espresso dalla conversazione – **spiega Gian Piero Turchi, del dip. FISPPA e coordinatore del progetto**-. Se il valore oscilla tra 1 e 3 (basso rischio) verrà chiesto allo user di continuare ad inserire nell'App le successive conversazioni; se il valore è tra 4 e 7 (medio rischio) si consiglia allo user di prendere contatti con i servizi del territorio; se il valore è superiore a 7 si avvisa lo user che verrà contattato da una task forcedi cyberbullismo

Partner chiave del progetto è Fondazione Carolina, dedicata a Carolina Picchio, prima vittima riconosciuta in Italia di cyberbullismo. La Onlus è impegnata dal 2018 per il benessere digitale delle nuove generazioni.

«Nonostante la continua e trasversale opera di sensibilizzazione rispetto a queste tematiche, i casi sono in costante aument. Secondo il nostro Centro studi circa 3 ragazzi su 4 sono vittima di attacchi in Rete - **commenta il Segretario generale di Fondazione Carolina, Ivano Zoppi** -. Per tutelare i

minori online oggi è indispensabile fornire strumenti e questo progetto risponde alla necessità di agire concretamente dalla parte dei più giovani e delle loro famiglie».

Una sfida educativa che, con Net Guardian, si occupa di supportare il team del Dipartimento FISPPA nella sperimentazione della Chatbox nelle scuole e di coordinare la task force di servizi territoriali.

L'innovatività della soluzione è avvalorata dalla richiesta di associazioni di settore (es. Istituti di formazione), che esprimono i potenziali benefici della proposta per il monitoraggio delle conversazioni a scuola e la possibilità di intervenire tempestivamente ove si rilevi un rischio medio-alto di cyberbullismo.

Infatti, ad oggi, i servizi operanti nel contrasto al cyberbullismo intervengono post-segnalazione da parte di terzi, ma raramente dalle vittime, con il rischio che la violenza si sia già manifestata. Inoltre, numerose segnalazioni risultano poi essere “falsi positivi”.

**Il nuovo sistema, analizzando rigorosamente le modalità interattive e gli specifici contenuti usati nelle chat attraverso MADIT, e dunque l'impatto che il linguaggio usato genera, può agire anche preventivamente e diminuendo il numero di falsi positivi.**