



BullyBuster

A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms

Wanda Nocerino

wanda.nocerino@unifg.it

Università di Foggia



Partnership

- ▶ [Università degli Studi di Napoli «Federico II»](#)
 - ▶ Carlo Sansone, Stefano Marrone, Michela Gravina, Antonio Galli
- ▶ [Università degli Studi di Bari «Aldo Moro»](#)
 - ▶ Donato Impedovo, Vincenzo Gattulli, Lucia Sarcinella
- ▶ [Università degli Studi di Foggia](#)
 - ▶ Donatella Curtotti, Angela Procaccino, Grazia Terrone, Wanda Nocerino
- ▶ [Università degli Studi di Cagliari](#)
 - ▶ Gian Luca Marcialis, Giulia Orrù, Giovanni Puglisi, Sara Concas, Marco Micheletto, Gianpaolo Perelli



Il progetto Bullybuster



1

Analisi basata su video, tramite un processo di segmentazione e caratterizzazione della scena mediante descrittori strutturali temporali e spaziali, permetterà di rilevare specifiche azioni di bullismo sulla base dei movimenti della folla.



2

Analisi testuale durante le comunicazioni digitali nei social network permetterà, mediante l'individuazione di parole e frasi tipiche di molestie, oppressione e stalking informatici



3

Analisi comportamentale mediante il rilevamento della dinamica della sequenza di tasti e l'analisi della pressione dei tasti o del touchscreen nei dispositivi elettronici sarà di supporto per evidenziare il verificarsi di una condizione emotiva anomala.





Collecting significant data:the bB-Questionnaire

BullyBuster.pythonanywhere.com

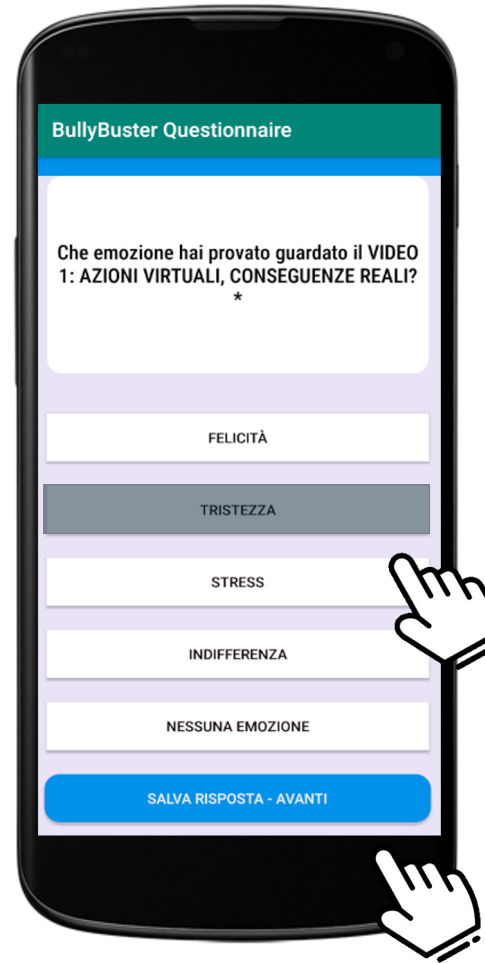


android



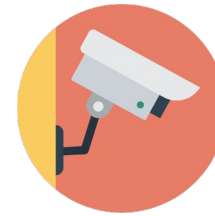
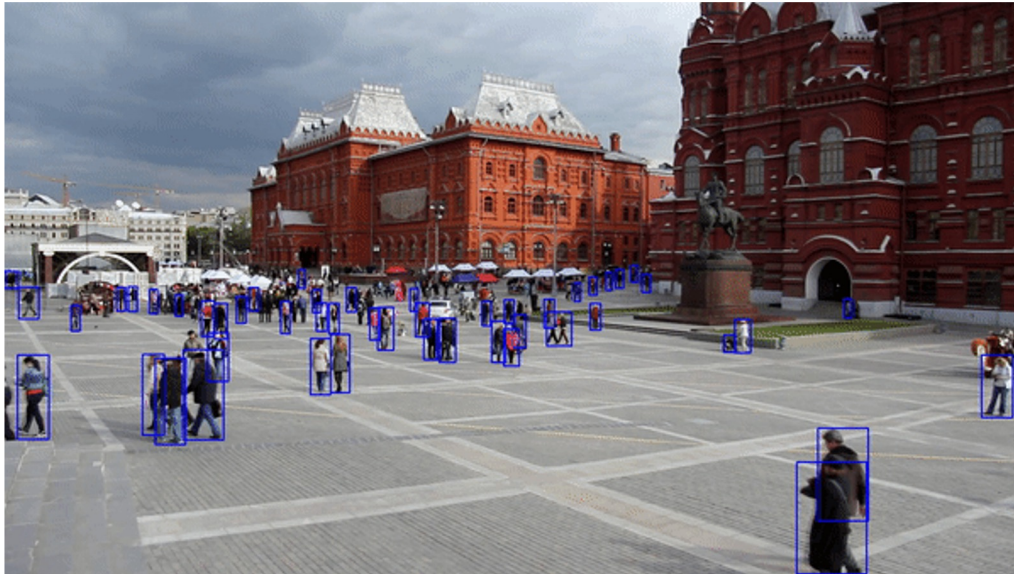
G. Terrone, A. Gori, E. Topino, A. Musetti, A. Scarinci, C. Guccione, V. Caretti, The Link between Attachment and Gambling/Internet Addiction in Adolescence: A Multiple Mediation Analysis with Developmental Perspective, Theory of Mind (Friend) and Adaptive Response, Journal Personalized Medicine, vol. 11, no. 3, 2021; <https://doi.org/10.3390/jpm11030228>.







Analisi delle "folle"



Comportamenti
violenti

Il tool di crowd analysis permette di osservare gruppi di soggetti da videocamere disposte in modo da non poter identificare i singoli, ma in grado di fornire sufficienti informazioni per segnalare, in base a modelli comportamentali opportunamente codificati, degli eventi "anomali" come episodi di violenza o panico.



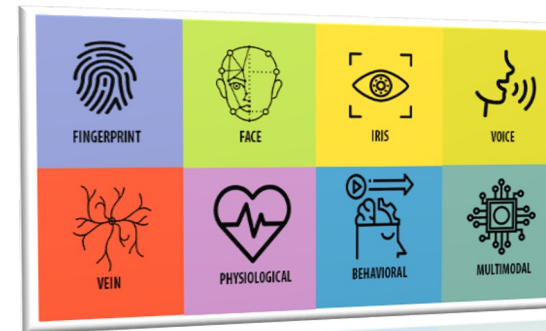
Text analyser

Il framework BullyBuster è dotato di un text analyzer che permette di analizzare chat di social network come Instagram, Facebook o una normale chat di gruppo (es. Telegram) di una classe scolastica con il fine di filtrare da essi solo commenti aggressivi/negativi.



Keystroke

- **Biometria** è un termine che si riferisce alle misurazioni del corpo e alle analisi statistiche volte a estrarre e quantificare le caratteristiche umane.
- Utilizzata principalmente per l'autenticazione o l'identificazione degli utenti, è stata sempre più utilizzata per altri scopi, tra cui l'intrattenimento e il benessere.
- Tra tutte, la dinamica dei tasti (KEYSTROKE) è considerata una delle biometrie comportamentali più efficaci ed economiche (cioè facili da implementare utilizzando hardware già disponibili).





I deepfake

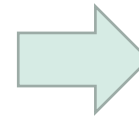
Il termine deepfake viene in genere utilizzato per riferirsi a un video che è stato modificato utilizzando un algoritmo di intelligenza artificiale per sostituire la persona nel video originale con qualcun altro in modo che il video sembri autentico.





...i deepfake "swap"

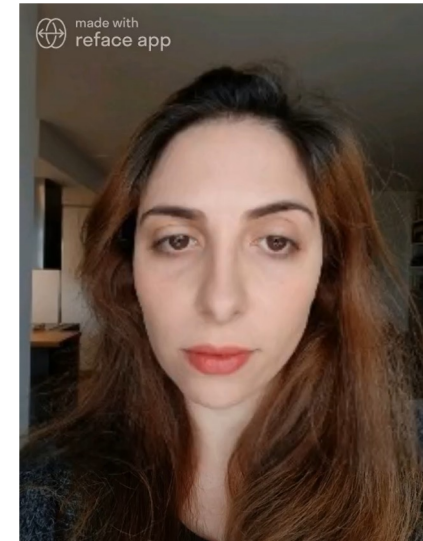
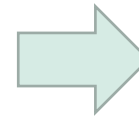
La tecnica "swap" permette di inserire il volto di un individuo target sul corpo di un secondo individuo.





... i deepfake "animations"

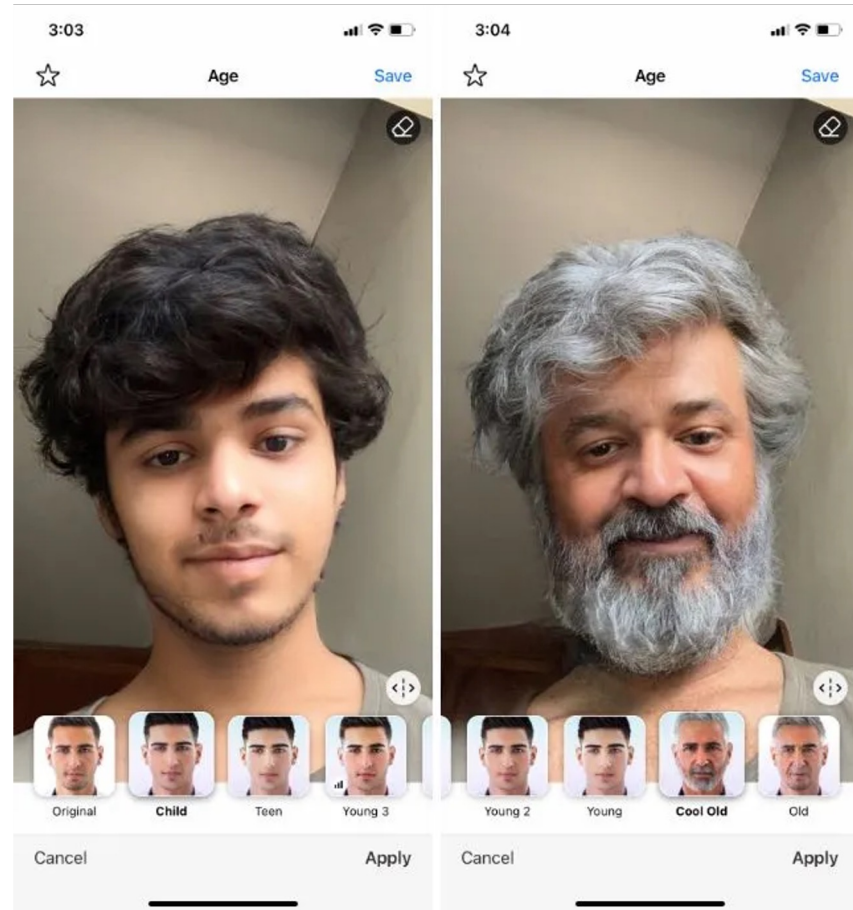
La tecnica "animation" permette di animare una foto a partire da un audio (musica o parlato)





... i deepfake "editing"

La tecnica "editing" permette di modificare alcune caratteristiche dell'individuo come l'età, il colore dei capelli, la presenza o meno di barba o capelli, e molto altro.





... i deepfake "sintesi"

La tecnica "sintesi" permette di generare artificialmente il volto di una persona che non esiste.



[<https://this-person-does-not-exist.com/en>]

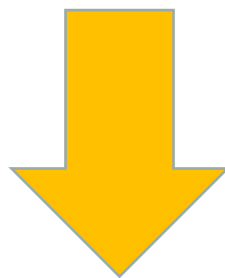




Quali rischi per la privacy



La principale criticità attiene al trattamento dei dati basato su videoregistrazione e riconoscimento facciale dei volti dei ragazzi presenti in una classe di scuola con analisi comportamentale attraverso sistemi di AI.



Regolamento (UE) 2016/679, GDPR (General Data Protection Regulation), attuato in Italia con il d.lgs. 10 agosto 2018, n. 101





Quali soluzioni?



► Art. 9, comma 2, lett. j, GDPR, per cui è consentito “il trattamento [...] a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici [...] che è proporzionato alla finalità perseguita, rispetta l'essenza del diritto alla protezione dei dati e prevede misure appropriate e specifiche per tutelare i diritti fondamentali e gli interessi dell'interessato”.



...con riferimento all'analisi video

La legittimità del trattamento dei dati biometrici dei minori coinvolti può essere garantita dalle c.d. misure di sicurezza.



- ▶ privacy by design nella fase di progettazione dei singoli tools;
- ▶ l'anonimizzazione delle immagini dei ragazzi;
- ▶ la pseudonimizzazione dei dati identificativi;
- ▶ la minimizzazione dei dati residuali da conservare solo per il periodo strettamente necessario alle finalità del progetto.



...con riferimento all'analisi testuale e comportamentale

Fermo l'obbligo del consenso dei genitori o del tutore, si è reso necessario predisporre delle adeguate informative privacy nonché i terms of condition del sistema e del servizio reso all'istituto scolastico. Le informative rispondono ai principi e alle indicazioni delle norme del GDPR, essere rilasciate a tutti i soggetti interessati.

Il bullybuster tool



Il bullybuster tool

BullyBuster : Tool chat di gruppo



Rischio contenuti multimediali manipolati

Rischio alto

Sono stati analizzati: 10 video inviati in chat nel periodo di riferimento
Di questi 7 video sono stati manipolati con tecniche deepfake
In media il 77% dei frame dei video presentava manipolazioni

Rischio violenza verbale

Medio-basso

Nel periodo di riferimento sono state mandate 5 parole volgari o offensive (2% dei messaggi inviati)
Gli studenti coinvolti sono 3 su 20 attivi nella chat

Rischio di azioni di violenza fisica

Basso

Nel periodo di riferimento sono stati rilevati 2 comportamenti anomali
I video contenenti le anomalie sono:
31_10_2022.mp4
6_12_2022.mp4
I comportamenti anomali sono durati in media 30 secondi





Contatti PI:

Prof.ssa Donatella Curtotti – donatella.curtotti@unifg.it

Prof. Donato Impedovo – donato.impedovo@uniba.it

Prof. Gian Luca Marcialis – marcialis@unica.it

Prof. Carlo Sansone – carlosan@unina.it

ConfGARR23

SAPERI INTERCONNESSI