information

Article

Development of Technologies for the Detection of (Cyber)Bullying Actions: The BullyBuster Project

Giulia Orrù [1], Antonio Galli [2], Vincenzo Gattulli [3], Michela Gravina [2], Marco Micheletto [1], Stefano Marrone [2], Wanda Nocerino [4], Angela Procaccino [4], Grazia Terrone [5], Donatella Curtotti [4], Donato Impedovo [3], Gian Luca Marcialis [1,*] and Carlo Sansone [2]

[1] Department of Electrical and Electronic Engineering, University of Cagliari, 09123 Cagliari, Italy; giulia.orru@unica.it (G.O.); marco.micheletto@unica.it (M.M.)
[2] Department of Electrical and Information Technology Engineering, University of Naples "Federico II", 80138 Naples, Italy; antonio.galli@unina.it (A.G.); michela.gravina@unina.it (M.G.); stefano.marrone@unina.it (S.M.); carlosan@unina.it (C.S.)
[3] Department of Computer Science, University of Bari, 70121 Bari, Italy; vincenzo.gattulli@uniba.it (V.G.); donato.impedovo@uniba.it (D.I.)
[4] Department of Law, University of Foggia, 71122 Foggia, Italy; wanda.nocerino@unifg.it (M.N.); angela.procaccino@unifg.it (A.P.); donatella.curtotti@unifg.it (D.C.)
[5] Department of History, Cultural Heritage, Education, and Society, Tor Vergata University, 00133 Rome, Italy; grazia.terrone@uniroma2.it
* Correspondence: marcialis@unica.it

Abstract: Bullying and cyberbullying are harmful social phenomena that involve the intentional, repeated use of power to intimidate or harm others. The ramifications of these actions are felt not just at the individual level but also pervasively throughout society, necessitating immediate attention and practical solutions. The BullyBuster project pioneers a multi-disciplinary approach, integrating artificial intelligence (AI) techniques with psychological models to comprehensively understand and combat these issues. In particular, employing AI in the project allows the automatic identification of potentially harmful content by analyzing linguistic patterns and behaviors in various data sources, including photos and videos. This timely detection enables alerts to relevant authorities or moderators, allowing for rapid interventions and potential harm mitigation. This paper, a culmination of previous research and advancements, details the potential for significantly enhancing cyberbullying detection and prevention by focusing on the system's design and the novel application of AI classifiers within an integrated framework. Our primary aim is to evaluate the feasibility and applicability of such a framework in a real-world application context. The proposed approach is shown to tackle the pervasive issue of cyberbullying effectively.

Keywords: bullying; deepfake detection; text analysis; keystroke dynamics

Citation: Orrù, G.; Galli, A.; Gattulli, V.; Gravina, M.; Micheletto, M.; Marrone, S.; Nocerino, W.; Procaccino, A.; Terrone, G.; Curtotti, D.; et al. Development of Technologies for the Detection of (Cyber)Bullying Actions: The BullyBuster Project. *Information* **2023**, *14*, 430. https://doi.org/10.3390/info14080430

Academic Editors: Marco Leo and Sara Colantonio

Received: 16 June 2023
Revised: 25 July 2023
Accepted: 27 July 2023
Published: 1 August 2023

1. Introduction

Bullying and cyberbullying are widespread societal concerns that harm the health and well-being of millions of children, adolescents, and adults worldwide. These negative behaviors immediately impact those affected and lead to a toxic environment in educational institutions, workplaces, and online spaces. With the rising reliance on digital platforms for communication and socialization, the necessity for creative methods to identify, prevent, and manage bullying and cyberbullying has become relevant. In particular, prevention and contrast programs are needed to neutralize the consequences and effects of these phenomena. Since prevention passes from identifying the first signs of aggression, automatic detection tools can greatly support the entities involved, such as schools and relatives. In this direction, integrating computer vision algorithms and artificial intelligence (AI) with psychological models can allow the achievement of frameworks for analyzing the behavior of groups of individuals to detect situations at risk.

This is the aim of the project, "BullyBuster—A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms," presented in this work. The BullyBuster (BB) project was funded under the tender relating to Projects of Relevant National Interest (PRIN) 2017. It involves four multidisciplinary research groups belonging to four universities in Southern Italy (University of Bari Aldo Moro, University of Cagliari, University of Foggia, University of Naples Federico II). This multidisciplinarity makes it possible to cover all aspects of implementing the bullying and cyberbullying detection and prevention framework from a technical, psychological, and legal point of view.

The BullyBuster investigation is functional not only on a theoretical level for constructing practical algorithms for identifying symptomatic behaviors of bullying and cyberbullying but also from an operational point of view. To promote strategies and policies for developing digital infrastructures, the framework could be implemented to facilitate investigative and judicial actions through the use of evidence collected at the real or virtual crime scene.

The BullyBuster project's preliminary results have led to important awards, including selection for inclusion in the "Maker Faire European Edition 10th Anniversary Book" and selection as a promising project by the "Research Centre on Artificial Intelligence under the auspices of UNESCO" (IRCAI) under the Global Top 100 list of AI projects addressing the 17 United Nations Strategic Development Goals (https://ircai.org/top100/entry/bullybuster-a-framework-for-bullying-and-cyberbullying-action-detection-by-computer-vision-and-artificial-intelligence-methods-and-algorithms/, accessed on 1 June 2023).

The aim of the paper is the presentation of the design of a framework that integrates behavior analysis systems based on artificial intelligence algorithms with psychological models and that is applicable in a real application context. Specifically, its application to school contexts was analyzed through the analysis of the Italian and European legal system. The paper is therefore placed in the area of AI for mental health and well-being. The paper is organized as follows. Section 2 describes the BullyBuster framework in terms of input context and technical details. Section 3 provides an experimental evaluation of AI-based detection modules. The prototype, use cases, and critical issues are reported in Section 4. Section 5 reports discussion and future developments of the framework.

2. Related Work and Social Context Overview

2.1. Related Work

The intersection of artificial intelligence and mental health is witnessing a promising evolution, creating a burgeoning field dedicated to harnessing AI for promoting well-being [1] (Figure 1). As a matter of fact, AI's capability to continuously collect, analyze, and interpret vast amounts of data offers unprecedented opportunities for real-time health monitoring, where subtle changes in behavior and mood can significantly impact treatment outcomes [2]. Furthermore, AI can offer personalized health advice by leveraging insights from individual data patterns. Such personalized interventions can increase the effectiveness of mental health treatments by tailoring them to each individual's unique needs and circumstances. In the broader context of mental well-being, AI applications extend into three main categories [3]: (1) personal sensing or digital phenotyping, which involves the use of mobile devices and wearables to collect and analyze data on individuals' behavior, physiological responses, and environmental factors, allowing for continuous monitoring and assessment of mental health conditions [4]; (2) natural language processing (NLP) of clinical texts and social media content [5]; (3) chatbots and conversational agents that can engage in interactive conversations with users, providing emotional support, psychoeducation, symptom tracking, and even therapeutic interventions [6].

However, the application of AI to specific areas, such as bullying and cyberbullying, is relatively new. The BullyBuster project is a prime example of pioneering work in this area, utilizing artificial intelligence and computer vision as potent tools in the fight against both cyber- and real-world bullying. In the context of physical bullying, AI is leveraged

to recognize suspicious behavior through behavioral biometrics such as body language or facial expressions [7]. Furthermore, "crowd analysis", studying the natural movements of people, groups of people, or objects, forms a crucial component of AI strategies to detect physical bullying [8]. In the digital domain, AI is deployed to detect manipulated content, also called "deep fakes" [9], a disturbingly common phenomenon among young people. Various image-quality metrics can be combined to create a model capable of distinguishing a broad range of manipulation techniques [10]. AI techniques also help in identifying verbal assaults. These systems can recognize offensive words [11] or assess emotional stress via keystroke dynamics analysis [12]. In addition, the AI within the BullyBuster project also caters to psychological aspects. For instance, the level of mobile device or internet addiction, an indicator of psychological distress, can be detected through automatic behavioral analyses [13]. Furthermore, human micro-behaviors can be identified by examining interaction and touch dynamics, enabling the definition of significant behavioral anomalies and characterizing human activities and related sentiments [14].

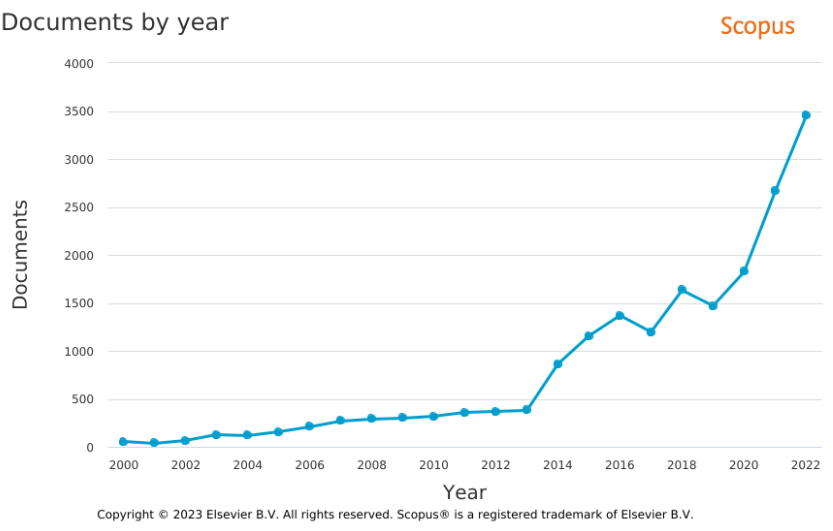

Figure 1. Annual publication trend from 2000 to 2022 for documents focusing on the intersection of artificial intelligence and well-being. The graph illustrates the increasing interest in this research area over time.

Translating these advancements into practical, effective tools, such as those deployed within the BullyBuster project, requires a comprehensive understanding of the broader context beyond the technological aspects. Crucial to this understanding are the psychological and legal implications of these applications. In fact, the implementation choices of the project are closely linked to the psychological consequences of these risky situations and to the legislative constraints of the countries in which it will be used. In particular, the Italian jurisprudence will be analyzed, as the Italian Ministry of Education, University, and Research funded the project.

2.2. Overview of Bullying and Cyberbulling from a Psychological Perspective

Bullying is a kind of anti-social behavior characterized by violent and intimidating actions of one person or a group of people over a targeted victim [15]. These actions may include verbal harassment, physical aggression, and persecution, generally carried out in school or even outside school, and have become a dramatic emergency among adolescents [16]. Cyberbullying has been described as a new form and extension of face-to-face bullying. Indeed, similar to traditional forms of face-to-face bullying, indicators of cyberbullying include aggression, repetitiveness, intentionality, and power imbalance between the victim and the perpetrator [17]. However, cyberbullying has several distinct characteristics from face-to-face bullying, such as potential perpetrator anonymity, lack

of physical interaction, accessibility of the victim, breadth of the potential audience, availability, and duration of the threat [18,19]. Despite these differences, many researchers agree that there is a significant overlap between cyberbullying and face-to-face bullying, meaning that bullies and victims appear to extend (or alternate) their roles from offline to online and vice versa [20]. Bullying and cyberbullying can influence the victim's physical, psychological, and social health. Besides relevant psychological damage due to depression (more than 57% of victims suffer from this), the victims are prone to a low educational rendition. They are unwilling to socialize with their peers. The extreme consequence is the victim's suicide [21,22]. Despite such evidence, little attention has been given to protective factors that could help mitigate some psychopathological problems associated with (cyber)bullying. In this regard, researchers are intensely interested in identifying risk and protective factors that could help to prevent and better contrast this phenomenon [23]. This goal is the cornerstone of the BullyBuster project.

2.3. Implications for the Italian Legal System

In Italy, the jurisprudence has to subsume the conduct of bullying in the context of individual cases already existing in the legal system (e.g., threat, stalking, beatings, injuries, digital identity theft, etc.) (see Court of Cassation, section V, 25 January 2021, n. 13979). However, limited to the phenomenon of cyberbullying, jurisprudence refers to the law 29 May 2017, no. 71 "Provisions for the protection of minors for the prevention and fight against the phenomenon of cyberbullying". The emphasis on cyberbullying is justified by the unique characteristics that the use of technical tools lends to bullying acts; the use of modern technologies can hide the perpetrators and greatly aggravate the harmful effects on the victim for their more significant and much more sound diffusion. It concerns only minors, both as offended persons and perpetrators.

In general, the legislator puts in place a series of remedies to be adopted once a violent episode, both bullying and cyberbullying, has occurred. Think of the duty to inform—incumbent on the headteacher who has known about the phenomena of cyberbullying—subjects exercising parental responsibility or the guardians of the victims (article 5, paragraph 1, Law No. 71/2017), or the possibility, reserved for minors of at least 14 years, of "the obscuring, removal or blocking" of personal data inserted on the net, even when the insertion does not constitute a crime. On the preventive side, the law introduces actions within educational institutions to neutralize the phenomenon from the beginning. For example, art. 3, no. 71/2017 provides for a "technical table" where all subjects (public and private) interested in the phenomenon are represented, which has the task of drawing up "an integrated action plan for the fight and prevention of cyberbullying"; art. 4 states that the Ministry of University and Research is responsible for adopting "guidelines for prevention and contrast in schools", which include the essential objective of the training school staff. In each school, a contact person is chosen from its teachers to coordinate the initiatives to prevent and combat (cyber)bullying.Despite the legislative interest, the need to guarantee a preventive approach to cyberbullying, aimed at protecting the individual and his well-being before the occurrence of any bullying event, has been identified by many authors [24].

3. The BullyBuster framework

The goal of the BullyBuster project (Figure 2) is to combine state-of-the-art AI methods with psychological models to produce tools that can accurately assess the risk of engaging in, assisting, or being the victim of bullying and violence in real and virtual scenarios.

(**a**) (**b**)

Figure 2. Logo of the project "BullyBuster—A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms" (**a**). The project has been included in the Global Top 100 list of AI projects by IRCAI (**b**).

The essence of the BullyBuster framework lies in its integrative nature (Figure 3). It brings together three crucial elements:

- Cutting-edge AI modules, each catering to a unique aspect of the detection and prevention solution: (1) crowd analysis for potential bullying incidents using video surveillance, (2) text analysis of social media accounts for possible verbal attacks, (3) keystroke dynamics analysis to gauge potential victims' emotional states, and (4) deepfake detection to counter the spread of manipulated harmful content;
- Comprehensive psychological models, developed from a data collection phase involving the BullyBuster questionnaire, which serve as the foundation for understanding the behavioral dynamics involved in bullying and inform our AI modules' functioning;
- Legal studies addressing privacy issues and other jurisdiction-specific regulations; BullyBuster pays significant attention to complying with legal constraints in the regions of its operation.

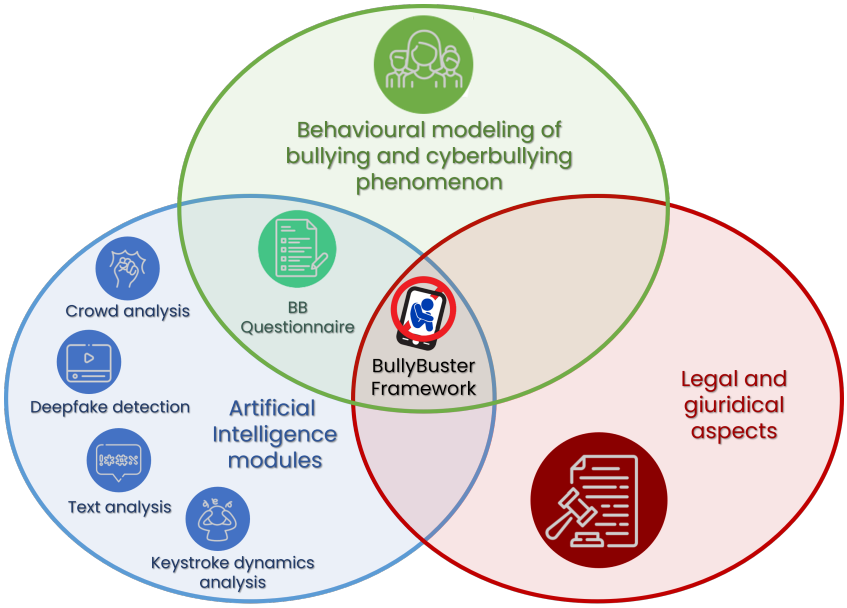

Figure 3. The BullyBuster project framework effectively integrates contributions from artificial intelligence, technology, law, and psychology.

3.1. Physical Violence Detection

Studying human activity and behavior through images and videos can be highly beneficial in the battle against bullying [25]. Physical aggression, isolation, or other physical patterns such as encirclement, are among the behavioral "markers" that might signify the presence of a problem. Based on these psychological models, a computer vision system can be trained to detect anomalous situations caused by bullying. We have developed an innovative approach to analyze crowd behavior by detecting groups of people in a given

scene and studying their temporal dynamics. Our system operates under the assumption that anomalous events occur when multiple instances of group formation and breaking up suddenly emerge in the scene. In particular, we have observed rapid and random aggregative or dispersive behavior during anomalies that induce panic among the crowd. The key aspect of our method lies in modeling the speed of these changes. This approach and its technical specifics are fully detailed in our previous work [8]. In summary, we have introduced a novel temporal descriptor specifically designed to capture variations in group numbers over time. Inspired by the one-dimensional local binary pattern [26], our descriptor utilizes a sliding time window encompassing a set of frames from a video sequence centered around a specific temporal instant. At each instant, we compare the number of groups computed with the counts before and after it. This comparison yields three possible cases: an increase, a decrease, or no change in the number of groups. These cases are encoded as a set of "trinary" states (increase/decrease/unchanged) represented by a string of "trits" (trit = trinary digit) [8]. By collecting a sufficient number of these strings, we construct histograms of the occurrences of trinary states, providing a temporal description of the group dynamics across the frames.

To determine whether an alarm should be triggered, we measure the deviation of each histogram from a baseline "quiet" state-related histogram. By continuously monitoring the variation in the number of groups over time, we can effectively detect such changes and utilize them for anomaly detection in crowd behavior. This descriptor was then used to implement an anomaly detector which warns an operator when the analyzed crowd scene shows characteristics that deviate from those defined as normality, i.e., a low number of abrupt changes in the number of groups of people.

3.2. Manipulated Video Content Detection

Deepfakes are manipulated audio, image, or video content with an alteration in someone's appearance, voice, or actions to make it appear as if they said or did something they never said or did [27]. Deepfakes can be used to harass, embarrass, or defame individuals and are, therefore, a powerful and dangerous tool in the context of bullying. To preserve the reliability of multimedia communications, deepfake detection is crucial. Deepfake detectors frequently focus on one or more specific types of manipulation, making them unable to generalize [28]. However, adequately planned ensemble learning and fusion techniques can minimize this problem. Our deepfake detection module within the BullyBuster framework integrates multiple classifier systems that leverage the complementary strengths of various models to enhance generalizability across different manipulation types [10]. We carefully selected four representative deepfake models from the state of the art to create this robust detection system. The specific implementation details for each model are beyond the scope of this paper but can be found in the respective original works:

- Visual Artifact-based Detector: This model employs a fine-tuned ResNet50 trained on artificially modified face images to simulate resolution inconsistencies. It detects visual anomalies in image content [29].
- General Network-based Detector: This model uses a fine-tuned XceptionNet that has been trained on deepfake images. It provides a broad baseline for detecting a variety of common deepfake manipulations [30].
- Frequency Analysis-based Detector: This model is built around Discrete Fourier Transform (DFT) for frequency analysis, making it adept at detecting manipulations that alter an image's frequency characteristics [31].
- DCT-based Detector: This model, described in [32], utilizes Discrete Cosine Transform (DCT) and robustness-enhancing augmentation techniques. It helps to detect manipulation in different scaling variations.

The strength of our module lies not only in the individual capacities of these models but, above all, in their integration. The module allows the operator to select one or multiple models based on the specific detection needs. This design offers flexibility and enhances the overall detection accuracy. Furthermore, we employ several fusion techniques, like

arithmetic mean (mean), Bayesian mean (bayes), accuracy-based parametric fusion (acc-based), and fusion via Multilayer Perceptrons (MLP) to combine the outputs of the different models [10]. This fusion approach further increases our detection system's robustness and adaptability to different deepfake types.

3.3. Verbal Abuse Detection

Cyberbullying is a prevalent issue, defined as an intentional aggressive act by an individual or a group of individuals, using electronic forms of contact, repeated over time against a victim who cannot easily defend themselves [33]. It poses a significant danger due to its pervasive nature, allowing the victim to be targeted anywhere [16]. This phenomenon is facilitated by the ease of disseminating textual comments, especially on social networks. The BullyBuster project aims to address this problem, featuring a module specifically designed to identify potential attacks within textual comments, such as those on social media platforms. In particular, the project incorporates a methodology derived from a previous study conducted by Gattulli et al. [11], which focuses on examining the intentionality behind text messages.

Such an approach is based on creating different features that characterize the intention of a text message. The features used in this study are the number of negative words (BW), the number of "don't/not" (NN), the use of (U), the positive/negative comment weight (PW/NW), the use of second person (SP), the presence of threats (TR), the presence of bullying terms (KW), and the comment length (L).

More specifically, based on the characteristics, it was deemed appropriate to create a dictionary as well:

- Number of negative words (BW): words that fall within a defined vocabulary "Badwords", containing 540 negative, vulgar words, insults, and humiliation [34].
- Number of "not/not" (NN): number of "not/not" within the comment.
- Uppercase (U): boolean value indicating whether the comment is capitalized. It can be interpreted as an attack on someone [35].
- Positive/negative comment weight (PW/NW): positive and negative comment weight in the range [0,1] (WordNet and SentiWordNet) [36].
- Use of the second person (SP): words that fall within a defined vocabulary, containing 24 words indicating the presence or absence of a second singular or plural form in the comment [37].
- Presence of threats (TR): words that fall within a defined vocabulary, containing 314 violent or inciting words [38].
- Presence of bullying terms (KW): words that fall within a defined vocabulary, containing 359 terms identified as insults and possible insults.
- Comment length (L): a value representing the length of the comment in terms of words.

The feature engineering phase was followed by choosing an appropriate classifier among different AI approaches, i.e., SVM with linear kernel, Random Forest, MLP, and Decision Tree. This development has equipped the BullyBuster framework with a verbal abuse detector that analyzes one or more social accounts in a reference period and returns the percentage of aggressive and vulgar comments received and the number of accounts involved.

3.4. Stress Detection

Biometrics involves using body measurement and statistical analyses to extract and quantify human characteristics. Initially employed for user authentication and identification purposes [39], this technology is now increasingly applied in several domains, including entertainment and user experience personalization [40]. In the realm of biometrics, there are two distinct categories of approaches: (i) physiological biometrics, which entails the direct measurement of physical attributes, including facial features, fingerprints, iris patterns, retinal structures, and vocal characteristics, and (ii) behavioral biometrics,

which centers around capturing particular human behaviors such as handwriting, typing, or speaking.

Among different behavioral biometrics, keystroke dynamics has emerged as a highly effective and cost-efficient method, easily implementable using commonly available hardware. It has been increasingly used in recent years to enforce user authentication by analyzing habitual rhythm patterns as they type on a physical and virtual keyboard [41].

In combating cyberbullying, we leverage keystroke dynamics for user emotion recognition. The objective is to harness its potential as a cost-effective and widely accessible approach for emotion recognition, as it solely necessitates a standard keyboard as hardware. Furthermore, since a keystroke recorder can be implemented as hardware or software, with the latter being inconspicuous, individuals using the keyboard must be made aware of the monitoring process. The primary challenge encountered in keystroke dynamics relates to the varying length of typing sessions. To address this issue, we have incorporated the time-windowing strategy proposed in our previous work [12]. This approach is particularly vital in social media contexts, where messages tend to be concise, and typing is often rapid. Implementing this method allows us to circumvent the limitations of shorter writing windows, consequently enabling more accurate analysis of user behavior. To briefly explain, our approach involves dividing a writing session, lasting for n seconds, into k sub-sessions by utilizing a time window of length t (where t is less than n), which slides over the original writing session with a stride of s. After extracting these sub-sessions, various features can be derived and employed to train a machine learning model to recognize the user's emotions experienced throughout the typing session. For a more in-depth understanding of the implementation specifics of our keystroke dynamics analysis, we encourage readers to refer to [12].

4. Experimental Evaluation

4.1. Preliminary Data Collection

Data were collected through a questionnaire among high school and university students to validate the proposed models. The student first views four animated videos with bullying and cyberbully situations and then fills in the questionnaire, the questions of which have been selected by the team of psychologists to estimate to what extent their behaviors in real life and on the internet put them at risk of acting or being bullied. In particular, the users are asked to watch the videos and then answer a question about their emotions while viewing each video. The question is designed to capture the individual's emotional experience using a 5-point Likert scale with options such as sadness, happiness, and other basic emotions [14]. The questionnaire helps acquire, in a strictly anonymous form, all possible information to identify typical "behavioural profiles" of bullies or victims. These profiles are then encoded and integrated into the AI systems for bullying and cyberbullying detection developed during the project.

The second part dealing with the questionnaire includes ninety-three questions that will then serve for labeling the individual in the primary classes (*personality index*), namely *Bullying perpetuator*, *Bullying victim*, *Cyberbully perpetuator*, and *Cyberbully-victim* [42]. To ensure that the participants are not influenced or primed in any way before watching the videos and providing their emotional responses, the labeling questions always follow those of acquiring emotional responses.

The personality index or class is assigned through objective mathematical calculations with scores for each response [42]. Each user can be classified in one or more of the four classes or as *Outsider*.

During the test, the app records the following features:

- *KeyLogger*: everything that is typed on the keyboard.
- *Touch and Multi-Touch coordinates* on the cell phone display during the test (e.g., playing on video, scrolling back and forth on video, etc.).
- *Questionnaire answers* (date and time, question, answer).

- *Sensor Values*: gyroscope, accelerometer, proximity, atmospheric pressure, magnetometer, ambient brightness, step detector (some devices do not have all sensors).

Since there is no registration phase, the data are in no way traceable to the individual who completed the questionnaire. The questionnaire is preceded by the acceptance of a license agreement specifying the data collection purposes and modality.

In the preliminary experiment reported here, we analyzed the results of 147 users; of these, 58.50% completed the entire questionnaire, and 12.24% completed more than 60%.

Only individuals who completed more than 60% of the questionnaire were considered for the personality index analysis. The results of these 103 users are shown in Figure 4. In addition to the personality index, a risk level was associated with each individual:

- Range 1 (Low Risk): Individuals falling within Range 1 are considered to have a low risk of engaging in bullying behaviors or being victimized. They are less likely to display aggressive or harmful actions towards others and are relatively less vulnerable to being targeted by bullies.
- Range 2 (Moderate Risk): Participants falling within Range 2 have a moderate risk of either perpetrating bullying behaviors or becoming victims of bullying. They might exhibit occasional instances of aggressive behaviors or encounter some bullying experiences, but their involvement is not severe or pervasive.
- Range 3 (High Risk): Individuals in Range 3 are at a higher risk of either actively engaging in bullying behaviors or experiencing significant victimization. They might display frequent or intense aggressive actions or are prone to repeated victimization and distress due to bullying encounters.

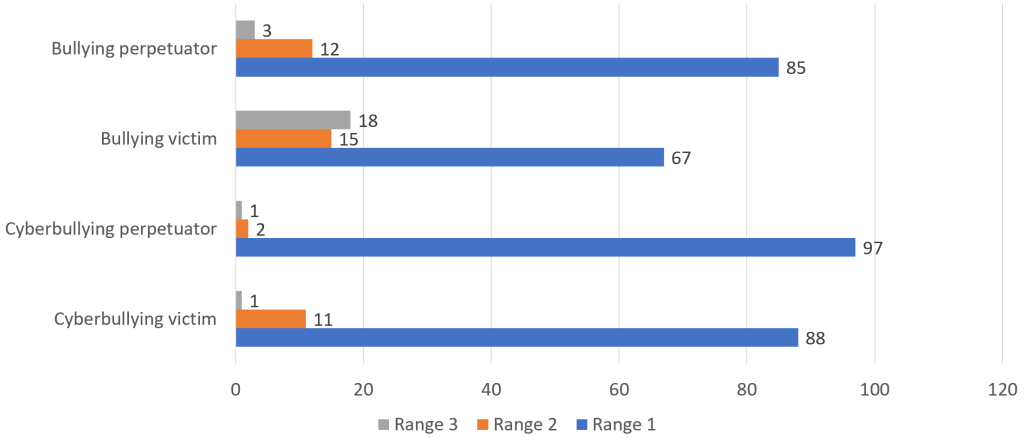

Figure 4. Number of individuals who fall into a personality index for each risk level. Range 1 corresponds to a low risk, range 2 to a moderate risk, and range 3 to a high risk for bullying behaviors or victimization.

It is important to note that some users were categorized into multiple personality indexes. The experiment denotes a shortage of *Cyberbullies* or *Bullies* and an overall higher number of *Victims*, especially of bullying. The rest of the test users were classified as unrelated to bullying.

In addition to profiling bullying and cyberbullying, it is possible to evaluate internet addiction through the CIUS-7 test [43]. The CIUS-7 test was 100% completed by 100 users. Scoring of the test showed that 97 users had no internet addiction, while 3 had a risk of internet addiction.

4.2. Experimental Protocol and Results

This section presents a consolidated summary of the experimental results obtained from the individual modules composing the BullyBuster framework. These results, derived from separate studies and explained in their respective papers, have been presented here to give an overarching view of the BullyBuster performance. Each results segment

corresponds to a particular module within our framework, and readers interested in the detailed experimental setup, methodology, and comprehensive results are encouraged to refer to the cited original papers. This synthesized summary offers insight into how each component contributes to the overall effectiveness of the BullyBuster system.

4.2.1. Physical Violence Detection

To test the crowd anomaly detector, we adopted the Motion-Emotion dataset (MED) [44]. It contains 31 video sequences totaling approximately 44,000 frames. Each video is captured at 30 frames per second with a fixed camera above individual walkways. Crowd density in the videos ranges from sparse to dense. The videos show normal and abnormal behaviors, manually labeled frame by frame as one of five classes (Panic, Fight, Congestion, Obstacle, and Neutral). The videos do not show a single behavior at a time but rather the transition from normal to abnormal. For this reason, the dataset allows us to simulate the anomaly detector's functioning in detecting bullying actions, such as in a school playground. It is important to note that this kind of system operates so far from individuals that automatic facial recognition is impossible (Figure 5). To evaluate the performance of the anomaly detection system based on the proposed descriptor, we measured the number of false alarms and the number of correctly detected anomalies based on the alarms falling within a time window corresponding to approximately 27 s and centered on the actual occurrence of the anomaly. Precision and recall are usually considered together to appreciate the system's performance. In addition, we used a summary parameter, the F1 score.

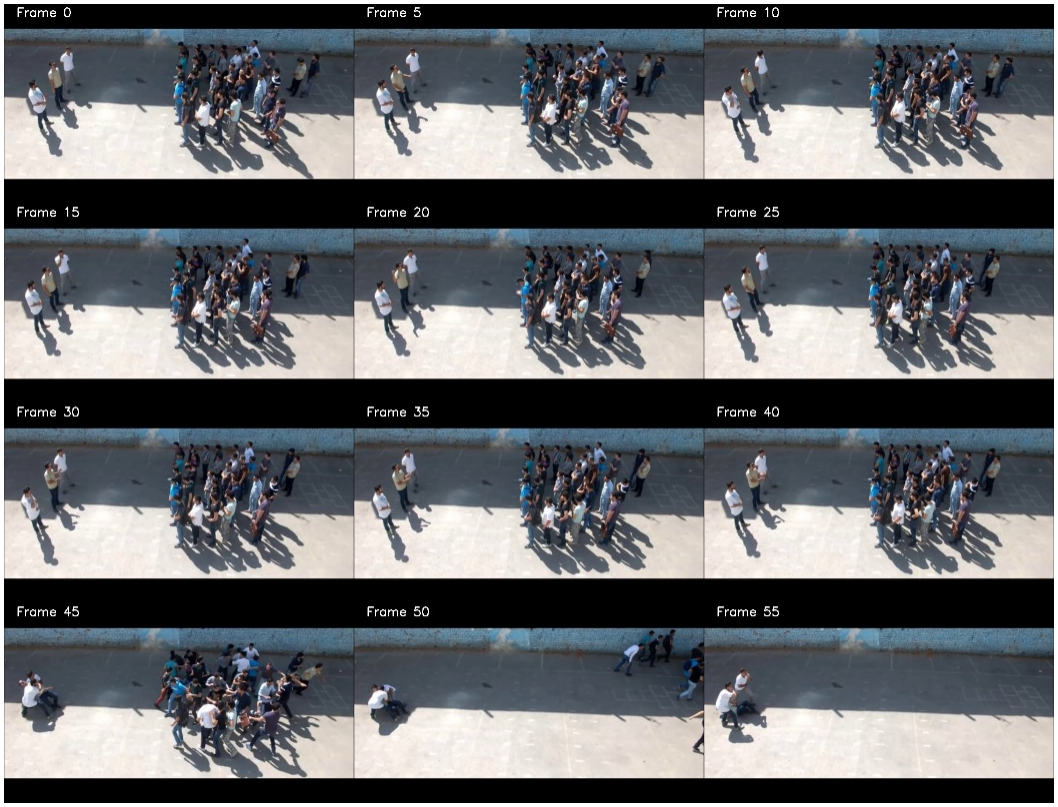

Figure 5. Example of a sequence depicting a scene of panic extracted from the MED dataset [44]; the individuals are at such a distance as to allow the analysis of the overall behavior of the crowd without allowing personal identification.

Among the range of results published in [8], we have chosen to highlight the best-performing ones within the scope of this paper. Accordingly, the anomaly detection system achieved a precision of 73.17%, a recall of 83.33%, and an F1-score of 77.92%. Albeit acknowledging the existence of false positives and negatives, these results underscore the

system's valuable assistance to a human operator in distinguishing between normal and abnormal scenes, paving the way for prompt intervention.

4.2.2. Manipulated Video Content Detection

In order to evaluate the effectiveness of the multimodal deepfake detector, we utilized the FaceForensics++ (FF++) dataset [30]. This dataset consists of 1,000 original videos and their manipulated counterparts generated using various deepfake techniques. Specifically, the manipulations are of four types; two are based on computer graphics methods (Face2Face and FaceSwap) and two on machine learning techniques (Deepfakes and Neural Textures). FaceSwap and Deepfakes predominantly involve face swap deepfakes, while Face2Face and Neural Textures consist of reenactment deepfakes. We analyzed the performance of each model and their fusion through the True Positive Rate (TPR), indicating accuracy on manipulated samples, and the True Negative Rate (TNR), denoting accuracy on genuine samples. These rates are calculated with a threshold set at 0.5, a standard choice in binary classification tasks since it provides a balanced measure of sensitivity and specificity. The detailed results from this evaluation are presented in Table 1. Each model exhibits unique strengths and weaknesses in their deepfake detection capabilities across different manipulation techniques. For example, the ResNet50 model achieves high TPRs across all manipulation types but struggles particularly with Neural Textures. Conversely, the DFT model performs strongly on deepfakes and Neural Textures but is less effective on FaceSwap and Face2Face. This observation corroborates the view of the state of the art [45], which suggests that deepfake models often have little capacity for generalization and accurately classify known types of samples (intra-dataset) but report significant errors on unknown types (cross-dataset). Nevertheless, our system's distinctive feature is the fusion of different models [10], which yields an average performance that remains robust across all tested manipulation types. For instance, the fusion of all models with a mean strategy produces consistent TPRs across all manipulation types while maintaining a TNR of 74.00%. Using an MLP-based fusion strategy results in an increased TPR across all manipulations, particularly notable with Neural Textures, albeit at a slight decrease in TNR. It is worth noting that these results represent an unpublished evaluation of the BullyBuster deepfake detection module. Through this analysis, we aim to emphasize the robustness and reliability of our system in real-world applications such as detecting and intervening in instances of bullying.

Table 1. Comparison of True Positive Rates (TPRs) and True Negative Rates (TNRs) for the individual and various combined models in the BullyBuster deepfake detection system. The combination methods are indicated in parentheses.

	TPR [%]				TNR [%]
Model/Test Set	**FaceSwap**	**Deepfakes**	**Face2Face**	**NeuralText.**	
ResNet50	97.29	99.88	96.00	12.66	97.35
XceptionNet	33.76	81.26	34.20	31.13	67.20
DFT	68.00	98.78	79.43	98.01	6.70
DCT	91.86	82.80	81.53	67.28	74.85
ResNet50 + XceptionNet (mean)	96.93	99.89	92.21	14.99	92.84
ResNet50 + XceptionNet (bayes)	76.16	99.77	86.34	13.40	95.80
ResNet50 + XceptionNet (acc-based)	97.39	99.94	95.20	12.08	97.61
ResNet50 + XceptionNet (MLP)	98.57	100.00	98.51	25.86	92.79
all (mean)	94.07	99.83	93.28	48.50	74.00
all (Bayes)	76.88	99.88	89.11	28.02	93.25
all (acc-based)	97.08	99.25	89.70	30.45	91.49
all (MLP)	98.98	100.00	97.92	72.65	73.38

4.2.3. Verbal Abuse Detection

While developing the verbal abuse detection module for the BullyBuster framework, we utilized a unique approach to emulate the conditions of textual cyberbullying. Our methodology, detailed in [11], drew upon the social network interactions of public figures frequently subjected to offensive commentary, allowing us to create the so-called "Aggressive Italian Dataset". The dataset includes interactions with four notable Italian individuals during November and December 2020, a period characterized by governmental instability and the ongoing COVID-19 pandemic. The individuals considered were the following:

- *Achille Lauro*, an Italian rapper, whose social profiles and normal fan comments are studded with offensive and sexist comments. His Twitter profile has over 57k followers.
- *Fabio Rovazzi*, an Italian Youtuber and singer. His social media profiles are full of insults considering him a poor singer. His Twitter profile has over 37k followers.
- *Matteo Renzi*, an Italian politician. He is often criticized for his political choices and made fun of with goliardic videos on the web. His Twitter profile has 3.3 million followers.
- *Giuseppe Conte*, an Italian politician, jurist, and academic. Aggressive comments are directed at his political work culminating in a recent government crisis. His Twitter profile has around 1 million followers.

The "Aggressive Italian Dataset" consists of 4028 comments, including main and nested ones, alongside user names and publication dates. These comments were manually classified into "aggressive" and "non-aggressive" categories by ten independent evaluators. The dataset was then divided into 3048 training samples and 1000 testing samples. Experimentation with this dataset generated several important observations broadly discussed in our previous study [11]. Table 2 summarizes these findings, showing the average performance metrics (Precision, Recall, and F1 score) of different classification models employed on the four profiles analyzed. The Random Forest (RF) model demonstrates the highest overall accuracy (0.93), outperforming the other classifiers. Notably, the RF model achieves an impressive 0.98 precision score in the non-aggressive class, implying that it is usually correct when it predicts a non-aggressive comment. This is a crucial metric within our BullyBuster framework, as false positives (misclassifying a non-aggressive comment as aggressive) can harm the user experience. The other models also show respectable performances, especially with SVM and MLP tying for second place in overall accuracy (0.90). The SVM model, in particular, shows a solid ability to correctly identify aggressive comments, reflected in its 0.90 recall score for the aggressive class.

Table 2. Text analyzer average results regarding the four social profiles analyzed.

Average Results [%]	**SVM**			**DT**			**RF**			**MLP**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Not-aggressive class	0.95	0.88	0.92	0.93	0.84	0.88	0.98	0.89	0.93	0.94	0.90	0.92
Aggressive class	0.73	0.90	0.81	0.66	0.83	0.73	0.78	0.96	0.86	0.77	0.84	0.81
Accuracy		0.90			0.85			0.93			0.90	

Challenges encountered during the experiments mainly centered around classifying sarcastic or vulgar comments containing grammatical errors and recognizing potentially offensive language hidden within hashtags. Despite these issues, the findings generally illustrate that our verbal abuse detection module can identify aggressive comments with a reasonable degree of accuracy. In particular, the RF model manifests an optimal equilibrium in accurately discerning both aggressive and non-aggressive comments, thereby rendering it the most suitable choice for integration within the BullyBuster framework.

4.2.4. Stress Detection

The final module we investigate within the BullyBuster framework is dedicated to stress detection. In the referenced study [12], we employed the EmorSurv dataset [46]. This dataset includes keystroke dynamics from 124 participants, with their emotional states categorized into five distinct classes: *Anger*, *Happiness*, *Calmness*, *Sadness*, and *Neutral State*. The data acquisition involved participants typing freely formulated sentences before and after being subjected to mood inductions via video stimuli in an interactive online platform. The collected data included timing and frequency information, capturing the participants' typing behavior. Given our time-windowing approach, which segments each user's writing session into smaller portions, we treated the data as multi-instance. Consequently, the Multi-Instance Learning (MIL) methodology was employed, viewing data as 'bags' of instances where each 'bag' represents a typing session. Two types of bags were considered: Fixed Bags (FBs), where the number of instances in each bag is kept constant, and Variable Bags (VBs), where the bag size can vary, mirroring the natural variability in typing session lengths. Different models' performance metrics, classification accuracy, precision, recall, and F1-score, are reported in Table 3. The results revealed that the MIL-SVM model outperformed the other techniques, especially when used with VBs. Given that emotion detection from keystrokes is a multi-instance problem with inherent variability in the length of typing sessions (thus, the number of instances), it is not surprising that the VB approach showed superior performance. Despite trying various balancing strategies to address the dataset's heavily skewed 'neutral' class data, the CNN models could not achieve comparable results. However, the slight imbalance between precision, recall, and F1-score in the MIL-SVM VB model suggests room for further refinement. Future work could involve optimizing the model to improve its ability to distinguish between different emotional states, particularly under bullying-related stress conditions.

Table 3. Comparison of the stress detector setups, in terms of classification accuracy, precision, recall, and F1-score (F1), varying the bag type (Fixed Bags—FBs, Variable Bags—VBs) and the balancing technique (Class weights—CW, Undersampling—US, Oversampling—OS, Under-oversampling—UOS). The best results are reported in bold.

Approach	Accuracy	Precision	Recall	F1-Score
CNN CW	0.48	0.58	0.48	0.50
CNN US	0.57	0.43	0.57	0.48
CNN OS	0.46	0.45	0.46	0.43
CNN UOS	0.52	0.48	0.52	0.49
MIL-SVM VB	**0.76**	**0.80**	**0.69**	**0.74**
MIL-SVM FB	0.52	0.6	0.52	0.53

5. BullyBuster Prototypes and Use Cases

The previous section presented the high-level BullyBuster framework; through a psychological modeling of the problem, artificial intelligence algorithms were identified that extract complementary information to describe a complex and transversal phenomenon. Starting from this high-level framework, it is possible to implement multiple prototypes. However, in order to create AI-assisted intervention–prevention systems, it is necessary to analyze the socio-legal context in which these systems will be used. For this reason, a brief overview of the criticalities considered for their definition will be provided before presenting the prototypes created during the project.

5.1. Critical Issues and Solutions

From a legal point of view, the main criticality of the BullyBuster project concerns the processing of data based on video recordings of the individuals present in a scene

with behavioral analysis through AI systems. The presence of minors further complicates the use of such systems. Regulation (EU) 2016/679, better known by the acronym GDPR (General Data Protection Regulation) [47], guarantees the effectiveness of protecting the person concerning the evolution of new technologies. The growing datafication has led to a definite increase in the risks associated with the improper use of personal data online through the most modern profiling techniques (and in the most severe cases, identity theft) [48]. Since it is widely understood that "privacy" means the right to protect one's identity, the legitimate exercise of data processing requires a balance with the effective right to control this activity [49].

Although minors have the right to privacy and protection of their data (Article 16 of the UN Convention on the Rights of the Child), access to network services and the consequent data entry had been left without adequate guarantees. This explains that Art. 8 GDPR regulates the "Conditions applicable to minors' consent about information society services". In this sense, any uncontrolled circulation of personal data concerning minors could expose them to hazardous situations capable of negatively affecting their developing personality [50]. This provision allows for the processing of special categories of personal data, such as biometric data used for video analysis in the context of the project, provided that it is done for archiving purposes in the public interest, scientific or historical research, or statistical purposes. This processing must be based on Union or national law, be proportionate to the intended objective, respect the essence of the right to data protection, and implement appropriate and specific measures to protect the fundamental rights and interests of the data subjects.

Regarding video analysis to detect bullying actions based on crowd movements around the victim, the processing of biometric data can be considered legitimate if proper security measures are implemented to safeguard the minors' data. These security measures should follow the privacy principles by design during the design phase of individual tools. This includes anonymizing the images of the children, pseudonymizing identification data, and minimizing residual data, retaining only the necessary information for the duration strictly required for the project's purposes. By adhering to these provisions and implementing privacy-enhancing measures, the BullyBuster project can ensure compliance with the GDPR while effectively addressing bullying issues using video analysis techniques.

To overcome the legal challenges related to analyzing messaging for identifying cyber-aggressive textual comments and behavioral analysis involving keystrokes and tactile dynamics, the BullyBuster project must implement the following measures in compliance with the GDPR. In particular, it was necessary to prepare adequate privacy information as well as the terms of the condition of the system and the service rendered at the educational institution. The disclosures comply with the principles and indications of the GDPR rules and will be released to all interested parties (e.g., families, parents, appointees, teachers). Also, in this case, the legitimacy of computer data acquisition is guaranteed through adequate security measures consisting of data encryption systems, pseudonymization, and anonymization, even of only part of the same or part of the processing. Lastly, it was necessary to prepare an impact assessment (DPIA, under Art. 35 GDPR) which takes into consideration the type and methods of treatment and the category of data processed, as well as a peremptory deadline for the cancellation of the data collected and analyzed once the research activity has ceased.

5.2. Prototypes

In the BullyBuster framework, three specific use cases and prototypes (Figure 6) are outlined:

- The BullyBuster Questionnaire (BullyBuster.pythonanywhere.com, accessed on 1 June 2023) was released as a mobile app and web app. This use case allowed data collection to develop psychological models based on tools to prevent bullying and cyberbullying behaviors. Students view animated videos depicting bullying scenarios before completing a questionnaire. The questions chosen by the BullyBuster team

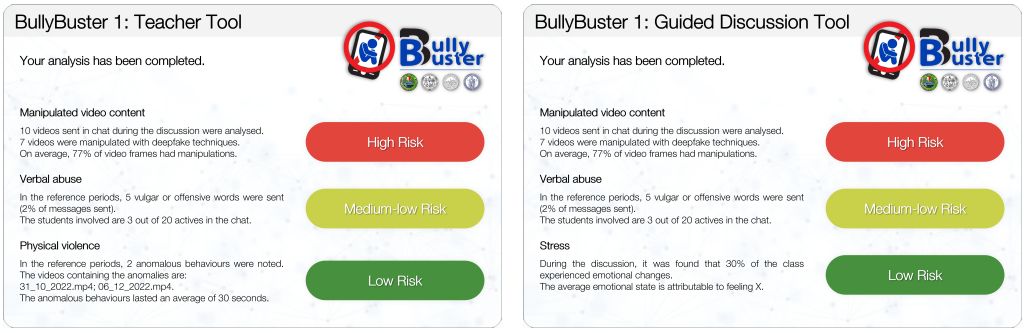

Figure 6. Final dashboards of two use cases of the BullyBuster framework, which allows the analysis of the behavior of a group of individuals to evaluate the risk of (cyber)bullying actions.

of psychologists are intended to estimate the extent to which the student's real-life and online behaviors put them at risk of perpetuating, assisting, or being a victim of bullying and violence. The BullyBuster Questionnaire is currently being used in educational institutions to collect data that will be used to improve the BullyBuster models.

- The Teacher Tool is a desktop application that allows teachers to upload data from the class chat and videos from the video surveillance system. Several modules process the uploaded data, including a deepfake detector, a text analyzer, and a crowd analyzer. The system generates a report that includes each module's risk percentages of bullying and cyberbullying actions and the overall risk percentage for the entire class. The GUI of the Teacher Tool includes a main window for data uploading and a results window for viewing the generated report.
- The Guided Discussion Tool is a desktop application addressed to students. In this use case, students must use the BullyBuster chat installed on desktop devices in the school's computer room to discuss assigned topics (such as the environment, politics, current events, and so on). The system then analyzes the chat data to detect the presence of deepfakes, violent comments, and stress levels via keystroke analysis. The system generates a report that shows the percentage risk of cyberbullying actions for each module and the overall risk for the class. The GUI results window in the Guided Discussion Tool allows the teacher to access and review the generated report.

6. Conclusions

Bullying and cyberbullying emerge as crucial social challenges that negatively affect individuals and communities, leading to the disruption of social interactions and the degradation of mental health. As our lives progressively move to digital platforms, the urgency and relevance of these issues have been accentuated.

The BullyBuster (BB) project encapsulates a comprehensive countermeasure to these prevailing concerns, merging a multidisciplinary strategy with advanced artificial intelligence technologies, computer vision, and psychological constructs. This paper presents the integration of this multidisciplinary research with the aim of one or more prevention/intervention systems in the fight against bullying applicable in a real context (in this case, in the European socio-legal context).

In terms of practical utility, the BullyBuster framework extends beyond mere theoretical contributions. Its modular architecture facilitates continuous enhancement, guaranteeing adaptability in response to the dynamic landscape of technological advancements and associated challenges. BullyBuster is a valuable resource for stakeholders like schools and families, contributing to digital infrastructure planning and legal investigations.

One notable aspect of the BullyBuster project is its ability to detect anomalous behaviors linked to bullying without collecting personally identifiable information. This approach underscores the project's strong commitment to upholding privacy and data

protection standards, establishing it as a viable and ethically considerate solution applicable across various environments.

The BullyBuster project's accomplishments, such as its endorsement by the "International Research Centre on Artificial Intelligence under the auspices of UNESCO" (IRCAI) as a promising project among the Global Top 100 AI initiatives addressing the 17 United Nations Strategic Development Goals, and its feature in the "Maker Faire European Edition 10th Anniversary Book", attest to its innovative approach and prospective influence. As we traverse the increasingly complex digital territory, this project marks notable progress toward creating safer physical and digital environments, demonstrating the transformative potential of technology and interdisciplinary collaboration to induce societal change.

In summary, the BullyBuster project brings together numerous unique and innovative facets, contributing significantly to the domain of bullying and cyberbullying detection. One of its key strengths is the interdisciplinary nature of the project. The project has successfully designed effective algorithms for detecting harmful behavior by merging juridical and psychological models. This approach embraces an expansive perspective, integrating critical knowledge from multiple fields. On the technical front, the project's design leverages cutting-edge computer vision and AI techniques specifically applied to bullying and cyberbullying. These state-of-the-art methods enable the project to accurately capture and analyze complex behavior patterns. Consequently, it can be employed effectively in realistic environments for prevention, detection, and potential legal action. This focus on practicality ensures that the project's outcomes have direct, real-world benefits in creating safer physical and digital environments. On the other hand, the BullyBuster project has faced several challenges. The first limitation was the necessity to comply strictly with GDPR, especially in video recording and behavioral analysis, to ensure individuals' rights to privacy and data control. Another crucial concern was data protection for minors. Given their involvement in the project, additional measures were needed to safeguard their privacy and data. Furthermore, a balance had to be struck between detecting bullying behavior through video analysis and preserving the privacy of minors. This was addressed by employing solutions such as privacy by design, image anonymization, and data pseudonymization. By exploring novel areas, such as the analysis of cyber-aggressive textual comments and keystroke dynamics, the project required obtaining parental consent and providing thorough privacy information alongside robust security measures. Lastly, the project involved a detailed Data Impact Assessment to consider the nature and methods of the data treatment and the types of data processed to set a firm timeline for data deletion after the research concluded. Given these considerations, we regard the BullyBuster project as not just a framework but a cornerstone for creating safer and healthier spaces. Bridging cutting-edge technologies and interdisciplinary knowledge, it sets the benchmark for new paradigms in preventing and detecting bullying behaviors. With our continued commitment to enhancing and adapting the framework to emerging challenges and technological advancements, we aspire to foster environments where respect and safety become the norm for all individuals and communities worldwide.

Author Contributions: Conceptualization, G.O., A.G., V.G., M.G., M.M., S.M., W.N., A.P., G.T., D.C., D.I., G.L.M. and C.S.; methodology, G.O., A.G., V.G., M.G., M.M. and S.M.; software, G.O., A.G., V.G., M.G., M.M. and S.M.; validation, G.O., A.G., V.G., M.G., M.M., S.M., W.N., A.P. and G.T.; formal analysis, G.O., A.G., V.G., M.G., M.M. and S.M.; investigation, G.O., A.G., V.G., M.G., M.M. and S.M.; data curation, G.O., A.G., V.G., M.G., M.M., S.M. and G.T.; writing—original draft preparation, G.O., A.G., V.G., M.G., M.M., S.M., W.N., A.P. and G.T.; writing—review and editing, G.O., D.C., D.I., G.L.M. and C.S.; visualization, G.O., A.G., V.G., M.G., M.M., S.M., W.N., A.P., G.T., D.C., D.I., G.L.M. and C.S.; supervision, D.C., D.I., G.L.M. and C.S.; project administration, D.C., D.I., G.L.M. and C.S.; funding acquisition, D.C., D.I., G.L.M. and C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Italian Ministry of Education, University and Research (MIUR) within the PRIN2017—BullyBuster—A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms (CUP: F74I19000370001). The project has been included in the Global Top 100 list of AI projects addressing the 17 UNSDGs (United Nations Strategic Development Goals) by the International Research Center for Artificial Intelligence under the auspices of UNESCO.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available at https://github.com/hosseinm/med, https://www.kaggle.com/datasets/sorokin/faceforensics, accessed on 1 June 2023. Other data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BW	Number of negative words
CNN	Convolutional Neural Network
CW	Class weights
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DPIA	Data Protection Impact Assessment
DT	Decision Tree
EU	European Union
FB	Fixed Bags
FF	FaceForensics
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
IRCAI	International Research Centre on Artificial Intelligence under the auspices of UNESCO
KW	Presence of bullying terms
L	Comment Length
MLP	Multi-Layer Perception
MED	Motion Emotion Dataset
MIL	Multi-Instance Learning
NGO	Non-governmental Organization
NN	Number of not
PRIN	Projects of Relevant National Interest
PW/NW	Positive/negative comment weight
RF	Random Forest
SP	Second Person
SVM	Support Vector Machine
TNR	True Negative Rate
TPR	True Positive Rate
TR	Presence of threats
U	Capitalization
UOS	Under-oversampling
US	Undersampling

References

1. van der Maden, W.; Lomas, D.; Hekkert, P. A framework for designing AI systems that support community wellbeing. *Front. Psychol.* **2023**, *13*. [CrossRef]
2. Li, X.; Li, J.; Zhang, Y.; Tiwari, P. Emotion recognition from multi-channel EEG data through a dual-pipeline graph attention network. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 9–12 December 2021; pp. 3642–3647.
3. D'Alfonso, S. AI in mental health. *Curr. Opin. Psychol.* **2020**, *36*, 112–117. [CrossRef] [PubMed]

4. Faurholt-Jepsen, M.; Frost, M.; Ritz, C.; Christensen, E.M.; Jacoby, A.S.; Mikkelsen, R.L.; Knorr, U.; Bardram, J.E.; Vinberg, M.; Kessing, L.V.; et al. Daily electronic self-monitoring in bipolar disorder using smartphones—The MONARCA I trial: A randomized, placebo-controlled, single-blind, parallel group trial. *Psychol. Med.* **2015**, *45*, 2691–2704. [CrossRef] [PubMed]
5. Guntuku, S.C.; Yaden, D.; Kern, M.; Ungar, L.; Eichstaedt, J. Detecting depression and mental illness on social media: An integrative review. *Curr. Opin. Behav. Sci.* **2017**, *18*, 43–49. [CrossRef]
6. Vaidyam, A.; Wisniewski, H.; Halamka, J.; Keshavan, M.; Torous, J. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Can. J. Psychiatry* **2019**, *64*, 070674371982897. [CrossRef]
7. Rescigno, M.; Spezialetti, M.; Rossi, S. Personalized models for facial emotion recognition through transfer learning. *Multimed. Tools Appl.* **2020**, *79*, 35811–35828. [CrossRef]
8. Orrù, G.; Ghiani, D.; Pintor, M.; Marcialis, G.L.; Roli, F. Detecting anomalies from video-sequences: A novel descriptor. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 4642–4649.
9. Vestman, V.; Kinnunen, T.; Hautamäki, R.G.; Sahidullah, M. Voice Mimicry Attacks Assisted by Automatic Speaker Verification. *Comput. Speech Lang.* **2020**, *59*, 36–54. [CrossRef]
10. Concas, S.; La Cava, S.M.; Orrù, G.; Cuccu, C.; Gao, J.; Feng, X.; Marcialis, G.L.; Roli, F. Analysis of Score-Level Fusion Rules for Deepfake Detection. *Appl. Sci.* **2022**, *12*, 7365. [CrossRef]
11. Gattulli, V.; Impedovo, D.; Pirlo, G.; Sarcinella, L. Cyber Aggression and Cyberbullying Identification on Social Networks. In Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2022), Vienna, Austria, 3–5 February 2022; pp. 644–651.
12. Marrone, S.; Sansone, C. Identifying Users' Emotional States through Keystroke Dynamics. In Proceedings of the 3rd International Conference on Deep Learning Theory and Applications, DeLTA, INSTICC, Lisbon, Portugal, 12–14 July 2022; SciTePress: Setubal, Portugal, 2022; Volume 1, pp. 207–214. [CrossRef]
13. Haberlin, K.A.; Atkin, D.J. Mobile gaming and Internet addiction: When is playing no longer just fun and games? *Comput. Hum. Behav.* **2022**, *126*, 106989. [CrossRef]
14. Balducci, F.; Impedovo, D.; Macchiarulo, N.; Pirlo, G. Affective States Recognition through Touch Dynamics. *Multimed. Tools Appl.* **2020**, *79*, 35909–35926. [CrossRef]
15. Wiertsema, M.; Vrijen, C.; van der Ploeg, R.; Sentse, M.; Kretschmer, T. Bullying perpetration and social status in the peer group: A meta-analysis. *J. Adolesc.* **2023**, *95*, 34–55.
16. Slonje, R.; Smith, P.K.; Frisén, A. The nature of cyberbullying, and strategies for prevention. *Comput. Hum. Behav.* **2013**, *29*, 26–32. [CrossRef]
17. Smith, P.K.; Mahdavi, J.; Carvalho, M.; Fisher, S.; Russell, S.; Tippett, N. Cyberbullying: Its nature and impact in secondary school pupils. *J. Child Psychol. Psychiatry* **2008**, *49*, 376–385.
18. Chan, H.C.O.; Wong, D.S.W. Traditional School Bullying and Cyberbullying Perpetration: Examining the Psychosocial Characteristics of Hong Kong Male and Female Adolescents. *Youth Soc.* **2019**, *51*, 3–29.
19. Sticca, F.; Perren, S. Is Cyberbullying Worse than Traditional Bullying? Examining the Differential Roles of Medium, Publicity, and Anonymity for the Perceived Severity of Bullying. *J. Youth Adolesc.* **2012**, *42*, 739–750. [CrossRef] [PubMed]
20. Kowalski, R.M.; Morgan, C.A.; Limber, S.P. Traditional bullying as a potential warning sign of cyberbullying. *Sch. Psychol. Int.* **2012**, *33*, 505–519.
21. Bauman, S.; Toomey, R.B.; Walker, J.L. Associations among bullying, cyberbullying, and suicide in high school students. *J. Adolesc.* **2013**, *36*, 341–350. [CrossRef]
22. Hu, Y.; Bai, Y.; Pan, Y.; Li, S. Cyberbullying victimization and depression among adolescents: A meta-analysis. *Psychiatry Res.* **2021**, *305*, 114198. [CrossRef]
23. Terrone, G.; Gori, A.; Topino, E.; Musetti, A.; Scarinci, A.; Guccione, C.; Caretti, V. The Link between Attachment and Gambling in Adolescence: A Multiple Mediation Analysis with Developmental Perspective, Theory of Mind (Friend) and Adaptive Response. *J. Pers. Med.* **2021**, *11*, 228. [CrossRef]
24. Mantovani, M.O. Profili penali del cyberbullismo: la L. 71 del 2017. *Indice Penale* **2018**, *2*, 475.
25. Zhan, B.; Monekosso, D.N.; Remagnino, P.; Velastin, S.A.; Xu, L.Q. Crowd analysis: A survey. *Mach. Vis. Appl.* **2008**, *19*, 345–357. [CrossRef]
26. Chatlani, N.; Soraghan, J.J. Local binary patterns for 1-D signal processing. In Proceedings of the 2010 18th European Signal Processing Conference, Aalborg, Denmark, 23–27 August 2010; pp. 95–99.
27. Zhang, T. Deepfake generation and detection, a survey. *Multimed. Tools Appl.* **2022**, *81*, 6259–6276. [CrossRef]
28. Nadimpalli, A.V.; Rattani, A. On improving cross-dataset generalization of deepfake detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 20 June 2022; pp. 91–99.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
30. Rössler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Niessner, M. FaceForensics++: Learning to Detect Manipulated Facial Images. In Proceedings of the ICCV 2019, Seoul, Korea, 27 October–2 November 2019; pp. 1–11. [CrossRef]
31. Durall, R.; Keuper, M.; Pfreundt, F.J.; Keuper, J. Unmasking DeepFakes with simple Features. *arXiv* **2020**, arxiv:1911.00686.

32. Concas, S.; Perelli, G.; Marcialis, G.L.; Puglisi, G. Tensor-Based Deepfake Detection in Scaled and Compressed Images. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 3121–3125. [CrossRef]
33. Dredge, R.; Gleeson, J.; de la Piedad Garcia, X. Presentation on Facebook and risk of cyberbullying victimisation. *Comput. Hum. Behav.* **2014**, *40*, 16–22. [CrossRef]
34. Ishara Amali, H.M.A.; Jayalal, S. Classification of Cyberbullying Sinhala Language Comments on Social Media. In Proceedings of the 2020 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 28–30 July 2020; pp. 266–271. [CrossRef]
35. Chatzakou, D.; Kourtellis, N.; Blackburn, J.; Cristofaro, E.D.; Stringhini, G.; Vakali, A. Mean Birds: Detecting Aggression and Bullying on Twitter. *arXiv* **2017**, arXiv:1702.06877.
36. Raghavan, M.; Poongavanam, M.K.; Ramachandran, S.R.; Sridhar, R. Emotion and sarcasm identification of posts from facebook data using a hybrid approach. *ICTACT J. Soft Comput.* **2017**, *7*, 1427–1435. [CrossRef]
37. Shtovba, S.; Petrychko, M.; Shtovba, O. Detection of Social Network Toxic Comments with Usage of Syntactic Dependencies in the Sentences. In Proceedings of the Conference the Second International Workshop on Computer Modeling and Intelligent Systems, CEUR Workshop 2353, Zaporizhzhia, Ukraine, 15–19 April 2019.
38. Raza, M.; Memon, M.; Bhatti, S.; Bux, R. *Detecting Cyberbullying in Social Commentary Using Supervised Machine Learning*; Springer: New York, NY, USA, 2020; pp. 621–630. [CrossRef]
39. Jain, A.; Hong, L.; Pankanti, S. Biometric identification. *Commun. ACM* **2000**, *43*, 90–98. [CrossRef]
40. Mandryk, R.L.; Nacke, L.E. Biometrics in Gaming and Entertainment Technologies. In *Biometrics in a Data Driven World*; Chapman and Hall/CRC: London, UK, 2016; pp. 215–248.
41. Karnan, M.; Akila, M.; Krishnaraj, N. Biometric personal authentication using keystroke dynamics: A review. *Appl. Soft Comput.* **2011**, *11*, 1565–1573. [CrossRef]
42. Hinduja, S.; Patchin, J. Bullying, Cyberbullying, and Suicide. *Arch. Suicide Res. Off. J. Int. Acad. Suicide Res.* **2010**, *14*, 206–221. [CrossRef]
43. Lopez-Fernandez, O.; Griffiths, M.D.; Kuss, D.J.; Dawes, C.; Pontes, H.M.; Justice, L.; Rumpf, H.J.; Bischof, A.; Gässler, A.K.; Suryani, E.; et al. Cross-cultural validation of the compulsive internet use scale in four forms and eight languages. *Cyberpsychol. Behav. Soc. Netw.* **2019**, *22*, 451–464. [CrossRef]
44. Rabiee, H.; Haddadnia, J.; Mousavi, H.; Kalantarzadeh, M.; Nabi, M.; Murino, V. Novel dataset for fine-grained abnormal behavior understanding in crowd. In Proceedings of the 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Colorado Springs, CO, USA, 23–26 August 2016; pp. 95–101.
45. Malik, A.; Kuribayashi, M.; Abdullahi, S.M.; Khan, A.N. DeepFake detection for human face images and videos: A survey. *IEEE Access* **2022**, *10*, 18757–18775. [CrossRef]
46. Maalej, A.; Kallel, I. Does Keystroke Dynamics tell us about Emotions? A Systematic Literature Review and Dataset Construction. In Proceedings of the 2020 16th International Conference on Intelligent Environments (IE), Madrid, Spain, 20–23 July 2020; pp. 60–67. [CrossRef]
47. Cuffaro, V.; Ciommo, F.D.; Gambini, M.L.; Alessandro, M.; D'Orazio, R. Trattamento dei dati personali e Regolamento UE n. 2016679. In *Il Corriere Giuridico—Monografie*; Ipsoa: Rome, Italy, 2018; ISBN 9788813366667.
48. Veale, M.; Borgesius, F.Z. Adtech and real-time bidding under European data protection law. *Ger. Law J.* **2022**, *23*, 226–256. [CrossRef]
49. Warren, S.; Brandeis, L. The right to privacy. In *Killing the Messenger*; Columbia University Press: New York, NY, USA, 1989; pp. 1–21.
50. Costello, C.R.; McNiel, D.E.; Binder, R.L. Adolescents and social media: Privacy, brain development, and the law. *J. Am. Acad. Psychiatry Law* **2016**, *44*, 313–321. [PubMed]