The BullyBuster Research Center: A Multidisciplinary Approach to Digital and Social Violence

Giulia Orrù[1,*], Guido Colaiacovo[4], Sara Concas[1], Antonio Galli[3], Vincenzo Gattulli[2], Michela Gravina[3], Stefano Marrone[3], Marco Micheletto[1], Wanda Nocerino[4], Angela Procaccino[4], Giovanni Puglisi[1], Lucia Sarcinella[2], Enrico Santoro[3], Grazia Terrone[5], Alessandro Valenti[4], Donatella Curtotti[4], Donato Impedovo[2], Gian Luca Marcialis[1] and Carlo Sansone[3]

[1]*University of Cagliari, piazza d'Armi, 09123, Cagliari, Italy*
[2]*University of Bari, Via Edorado Orabona 4, 70121, Bari, Italy*
[3]*University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy*
[4]*University of Foggia, Via Antonio Gramsci 89, 71122 Foggia, Italy*
[5]*Tor Vergata University, Via Columbia 1, 00133 Roma, Italy*

Abstract

The persistence and evolution of bullying and cyberbullying across physical and digital spaces continue to pose a major threat to individual and societal well-being. Building upon the foundations of the original PRIN 2017 BullyBuster project, the PRIN 2022 PNRR BullyBuster2 project is born from the need for advanced solutions aimed at the prevention of bullying and cyberbullying events. Compared to the previous project, it extends its scope with advanced artificial intelligence techniques and a broader interdisciplinary vision. A major achievement of BB2 is the establishment of the Interuniversity Research Centre "BullyBuster" (BB Centre), designed as a permanent multidisciplinary observatory to drive research, dissemination, and policy innovation in this domain. This paper presents the objectives, structure, and societal mission of the BB Center, situating it as a strategic initiative to promote human-centered, AI-powered solutions for the prevention and mitigation of bullying-related harm.

Keywords

bullying detection, cyberbullying, artificial intelligence, interdisciplinarity, research center

1. Introduction

Bullying and cyberbullying represent persistent and increasingly complex social challenges that extend beyond school-aged populations, affecting individuals in both private and professional contexts. Despite legislative progress, these phenomena continue to evolve, fueled by new digital platforms and behavioral dynamics that require innovative, scalable responses. In this context, artificial intelligence (AI) emerges as a promising tool to support prevention, detection, and intervention strategies. The project BullyBuster 2 – The ongoing fight against bullying and cyberbullying with the help of artificial intelligence for the human wellbeing (BB2), funded under the PRIN 2022 PNRR initiative and the European Union's NextGenerationEU program, represents a major evolution of the original BullyBuster framework created during the BullyBuster project (BB1) and funded under the PRIN 2017. The two projects involve multidisciplinary teams from four Southern Italian universities: Bari Aldo Moro, Cagliari, Foggia, and

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
giulia.orru@unica.it (G. Orrù); guido.colaiacovo@unifg.it (G. Colaiacovo); sara.concas90C@unica.it (S. Concas); antonio.galli@unina.it (A. Galli); vincenzo.gattulli@uniba.it (V. Gattulli); michela.gravina@unina.it (M. Gravina); stefano.marrone@unina.it (S. Marrone); marco.micheletto@unica.it (M. Micheletto); wanda.nocerino@unifg.it (W. Nocerino); angela.procaccino@unifg.it (A. Procaccino); puglisi@unica.it (G. Puglisi); lucia.sarcinella@uniba.it (L. Sarcinella); enrico.santoro.n2@gmail.com (E. Santoro); grazia.terrone@uniroma2.it (G. Terrone); alessandro.valenti@unifg.it (A. Valenti); donatella.curtotti@unifg.it (D. Curtotti); donato.impedovo@uniba.it (D. Impedovo); marcialis@unica.it (G. L. Marcialis); carlo.sansone@unina.it (C. Sansone)

Figure 1: BullyBuster2 project logo.

Naples Federico II. While the BB1 project focused on developing proof-of-concept technologies to detect bullying actions through computer vision and AI, BB2 expands this effort by integrating novel detection modules and extending the system's applicability to adult victims and complex social scenarios. Among the core innovations of BB2 is the creation of the Interuniversity Research Center "BullyBuster" (BB Center), a permanent multidisciplinary laboratory combining expertise in AI, cybersecurity, psychology, sociology, law, and health economics. The BB Center is conceived as a reference hub for institutions and stakeholders, dedicated to the development, dissemination, and ethical deployment of AI-based solutions against interpersonal aggression. The establishment of the center signals a shift from isolated technological development to a sustainable, ecosystemic approach aimed at protecting human dignity and promoting collective well-being.

2. The BullyBuster 2 Framework

The BullyBuster 2 (BB2) framework builds upon the technological foundations established in the original BB1 project by integrating multiple detection modules aimed at recognizing various forms of bullying and emotional distress in both digital and physical contexts. The upgraded architecture is structured around five core components (Figure 2:

- Manipulated Multimedia Detection Module: Analyzes visual and audio content to detect the presence of deepfakes [1, 2], including synthetic faces, voices, and expressions that can be used to harass or misrepresent individuals [3]. Compared to BB1, this module explores novel visual detection strategies aimed at enhancing cross-manipulation generalization, and introduces audio-based classification capabilities to detect voice cloning and speech manipulation.
- Behavioral Anomaly Detection Module: Leverages crowd analytics and surveillance camera data to identify anomalous patterns that may signal physical bullying or social exclusion. BB2 adopts data synthesis techniques based on generative models, such as those developed through diffusion models [4] to simulate panic or aggression situations in real environments, overcoming the limitations related to the scarcity of annotated datasets.
- Textual Aggression Detection Module: Identifies verbal abuse in online communications through text analysis. New datasets and language models are being adapted to detect aggression in adults [5], acknowledging the linguistic and psychological differences across age groups.
- Stress and Emotion Detection Module: Uses keystroke dynamics to assess user stress and emotional state in real time. Behavioral biometrics, such as typing rhythm and touch interaction, are explored as nonintrusive indicators of distress in victims [6]. Compared to BB1, this module has been extended to support input texts of varying lengths, and it incorporates the evaluation of new machine learning models to improve accuracy and robustness in emotion inference tasks.
- Digital Device Addiction Module: Tracks user behavior to detect signs of excessive digital engagement, through self-assessment tests and psychometric variables related to device usage [7]. This component provides indirect insight into psychological vulnerability and risk of cyberbullying involvement.

Each module is developed to be compliant with the legal and ethical regulations of the European Union and is based on a psychological modeling that allows to define specific use cases.

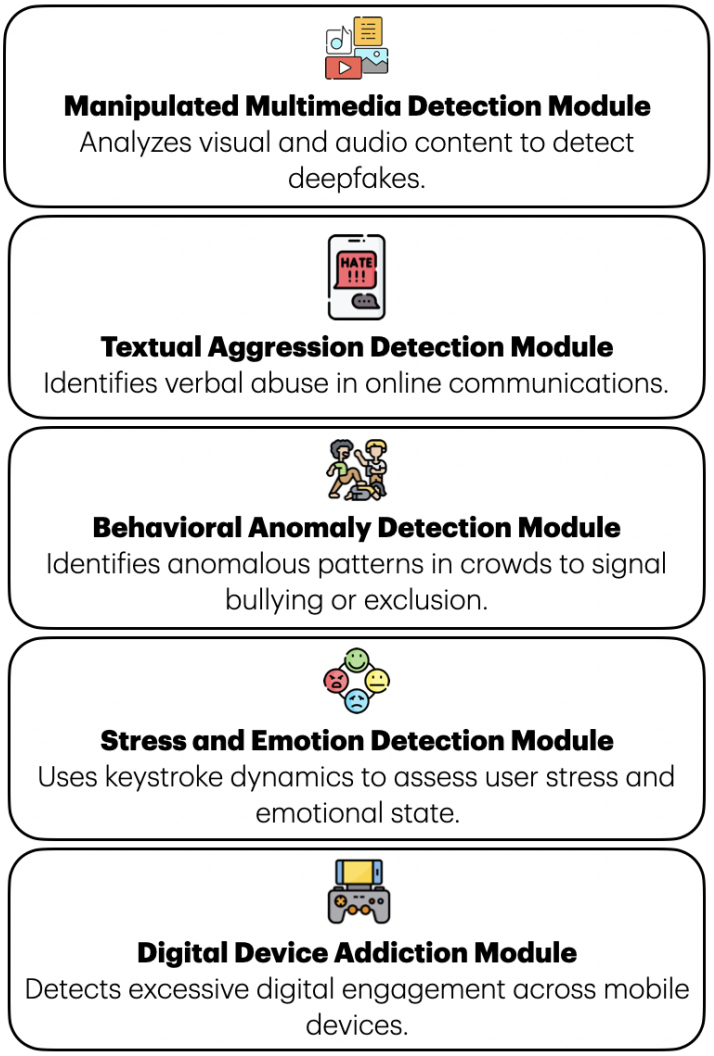

Figure 2: Schema of the BullyBuster 2 Framework showcasing its five core components: (1) Manipulated Multimedia Detection Module, (2) Behavioral Anomaly Detection Module, (3) Textual Aggression Detection Module, (4) Stress and Emotion Detection Module, and (5) Digital Device Addiction Module.

The BullyBuster 2 (BB2) project aims to deliver a range of concrete, functional outputs that go beyond theoretical development and translate research into practical tools for prevention, detection, and education in the context of bullying and cyberbullying. These outputs reflect the interdisciplinary nature of the project and are designed for deployment in real-world settings such as schools, public institutions, and research environments. In particular, three proof-of-concepts are under development and improvement:

- Teacher Tool: a desktop application allowing educators to upload class chat logs and video surveillance data. The software processes this data using AI modules for deepfake detection, textual aggression analysis, and behavioral anomaly detection in crowds. The output is a detailed report showing risk levels for each student and aggregated risk assessments for the class.
- Guided Discussion Tool: a desktop chat platform for students to discuss assigned topics. During interaction, the system monitors communication for verbal aggression and detects stress through keystroke dynamics. Results are compiled into an accessible risk report for teachers, offering insights into classroom dynamics and early warning signs of cyberbullying.
- Unified Web Application: designed to centralize and streamline access to multiple detection and analysis tools within a single coherent interface. The platform incorporates:

  1. The BullyBuster Questionnaire compilation: Users, particularly students, can complete emo-

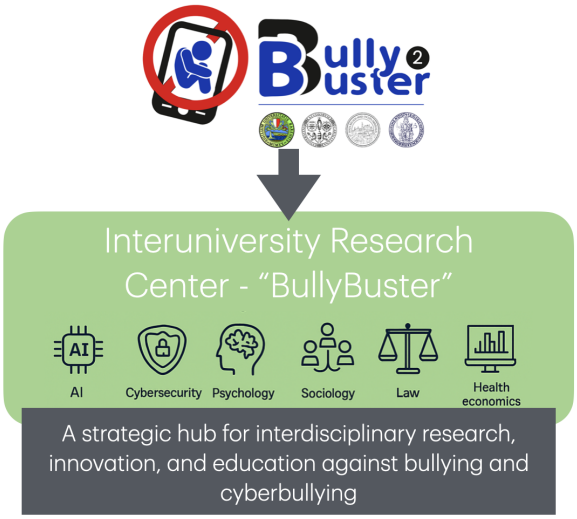

Figure 3: Schema describing the structure and disciplinary domains of the "BullyBuster" Interuniversity Research Center, established within the BB2 project as a national hub for AI-based, interdisciplinary action against bullying and cyberbullying.

tional questionnaires designed to assess risk profiles related to bullying and cyberbullying;

2. AI-Powered Analysis and Reporting: A module that processes questionnaire results using explainable clustering techniques and large language models, categorizing users based on behavioral patterns and risk factors.
3. Human Activity Recognition System: A mobile sensor-based framework for monitoring and analyzing behavioral indicators related to bullying incidents.
4. Visual Deepfake Detection System to analyze uploaded video content and flag synthetic or manipulated media.

From a legal perspective, the project pays particular attention to the ethical and legal aspects. In line with the GDPR and European regulations, each module is designed to protect the privacy and fundamental rights of users through measures such as data anonymization, impact assessment (DPIA), and local processing of information. The approach is also based on the principle of proportionality. Data collection occurs only if strictly necessary, and the most sensitive systems, such as biometrics, are used only in controlled contexts. In this way, BB2 guarantees a balance between effectiveness in prevention and respect for individual freedom, promoting a responsible, transparent, and ethical use of artificial intelligence.

3. The Interuniversity Research Center "BullyBuster"

A central outcome of BB2 is the establishment of the Interuniversity Research Center "BullyBuster", a permanent interdisciplinary initiative aimed at coordinating long-term efforts in combating bullying and cyberbullying (Figure 3). The BB Center brings together research units from four Southern Italian universities, the University of Bari Aldo Moro, the University of Cagliari, the University of Foggia and the University of Naples Federico II. It is structured to support collaborative work in six primary domains:

- Technology: that is, through experts in cybersecurity, computer vision, and artificial intelligence, who want to create and propose models and methods to be implemented in suitable demonstration products (proof-of-concept). Some examples of tools for preventing or combating bullying include the detection of counterfeit multimedia content, or the study of facial expressions, the analysis of keywords in chats or voice messages, and dynamics of typing under stress.

- Psychology: to model the behaviors of both perpetrators and victims, with a focus on understanding age-diverse psychological dynamics;
- Sociology: to assess the broader impact of aggression in digital communication environments, including social networks and workplace settings;
- Law: to create a permanent observatory on the legislation in this area, including European regulations and other sources of supranational law, to analyze the limits of the use of technologies for preventive or repressive purposes;
- Health Economics: to evaluate the socio-economic impact of bullying-related mental health outcomes on individuals and systems;
- Education and Dissemination: to design outreach strategies for schools, police forces, and companies, ensuring broad access to knowledge and prevention tools.

The BBCenter also aims to support interdisciplinary publications and research projects, organize national and international scientific events, promote technology transfer through partnerships with public institutions and private stakeholders, and act as a national reference point for decision-makers and regulators seeking expertise in anti-bullying technologies and policies.

This permanent disciplinary architecture allows the BB Center to act as a scientific and institutional interlocutor for schools, law enforcement agencies, public institutions, and private stakeholders. Its outputs are not limited to technical deliverables but include scientific publications, ethical frameworks, legal analysis, and validated tools designed for integration in real-world scenarios. The BB Center's originality lies not in any one domain, but in its ability to stabilize interdisciplinary cooperation over time, embedding artificial intelligence research within a broader ecology of legal responsibility, social awareness, and psychological realism. It is, in essence, a structural attempt to govern the development of technological responses to digital violence from within the academic system—rather than around or after it.

4. Conclusions

BullyBuster 2 (BB2) represents a significant evolution in the application of artificial intelligence to the prevention and management of bullying and cyberbullying. Building on the successful foundations of the BB1 project, BB2 broadens both the technological scope and the societal relevance of the original framework. By integrating modules for behavioral monitoring, verbal abuse detection, multimedia content analysis, stress inference, and digital addiction assessment, BB2 delivers a more comprehensive system for recognizing harmful interactions across age groups and environments.

A major milestone in this ongoing effort is the creation of the 'BullyBuster' Interuniversity Research Center, which institutionalizes a permanent multidisciplinary approach and offers a sustainable platform for research, education and technological innovation. This center functions not only as a hub for scientific collaboration but also as a reference point for public institutions, legal experts, educators, and mental health professionals.

In an era where interpersonal aggression often transcends physical boundaries, BB2 underscores the necessity of human-centered, ethically responsible AI solutions. By anchoring its development in psychological, legal, and social insights, the project seeks not merely to detect and react to aggression, but to contribute proactively to a culture of prevention, resilience, and well-being.

Future directions include validating the system in real world contexts, improving linguistic and cultural adaptability, and strengthening legal safeguards to ensure compliance with privacy and human rights standards. As bullying continues to evolve in form and impact, the BB Center offers a replicable model of how technology, when embedded within a robust ethical and institutional framework, can help shape a safer and more supportive society.

Acknowledgments

This work is supported by the European Union – Next Generation EU under the PRIN 2022 PNRR

project “BullyBuster 2 – the ongoing fight against bullying and cyberbullying with the help of artificial intelligence for the human wellbeing” (CUP: P2022K39K8).

Declaration on Generative AI

During the preparation of this work, the authors utilised ChatGPT and Grammarly to verify grammar and spelling, as well as to paraphrase and reword content. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

[1] S. Concas, S. M. La Cava, R. Casula, G. Orru, G. Puglisi, G. L. Marcialis, Quality-based artifact modeling for facial deepfake detection in videos, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3845–3854.

[2] J. Gao, M. Micheletto, G. Orrù, S. Concas, X. Feng, G. L. Marcialis, F. Roli, Texture and artifact decomposition for improving generalization in deep-learning-based deepfake detection, Engineering Applications of Artificial Intelligence 133 (2024) 108450.

[3] M. Gravina, A. Galli, G. De Micco, S. Marrone, G. Fiameni, C. Sansone, Fead-d: Facial expression analysis in deepfake detection, in: G. L. Foresti, A. Fusiello, E. Hancock (Eds.), Image Analysis and Processing – ICIAP 2023, Springer Nature Switzerland, Cham, 2023, pp. 283–294.

[4] G. Orru, R. Lecca, R. Puddu, S. M. La Cava, M. Micheletto, G. Marcialis, et al., Data generation via diffusion models for crowd anomaly detection, in: 35th British Machine Vision Conference 2024, BMVC 2024. Worshops Proceedings, BMVA press, 2025.

[5] V. Gattulli, D. Impedovo, A. Monaco, L. Sarcinella, Analyzing cyberaggression: Comparative model performance on social media comments with italian dataset, in: Proceedings of the 2024 Conference on Human Centred Artificial Intelligence-Education and Practice, 2024, pp. 15–21.

[6] S. Marrone, C. Sansone, Identifying users’ emotional states through keystroke dynamics, in: Proceedings of the 3rd International Conference on Deep Learning Theory and Applications - DeLTA, INSTICC, SciTePress, 2022, pp. 207–214. doi:10.5220/0011367300003277.

[7] F. Cozzolino, V. Moscato, A. Picariello, G. Sperlì, A community detection approach for smart-phone addiction recognition, 2019, pp. 53–64. doi:10.5220/0007839100530064.