

Deepfake Forensics Workshop (DFF-2026): Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media

Luca Guarnera
Department of Mathematics and
Computer Science,
University of Catania
Catania, Italy
luca.guarnera@unict.it

Francesco Guarnera
Department of Mathematics and
Computer Science,
University of Catania
Catania, Italy
francesco.guarnera@unict.it

Giovanni Puglisi
Department of Mathematics and
Computer Science,
University of Cagliari
Cagliari, Italy
puglisi@unica.it

Zahid Akhtar
Department of Electrical and
Computer Engineering,
State University of New York
Polytechnic Institute
Utica, NY, USA
akhtarz@sunypoly.edu

Fernando Pérez-González
atlanTTic, University of Vigo
Vigo, Spain
fperez@gts.uvigo.es

Sebastiano Battiato
Department of Mathematics and
Computer Science,
University of Catania
Catania, Italy
sebastiano.battiato@unict.it

Abstract

The rapid evolution of generative artificial intelligence is profoundly transforming the multimedia ecosystem, enabling the creation of highly realistic synthetic and manipulated content while simultaneously introducing unprecedented challenges for the authenticity, reliability, and trustworthiness of digital media. As generative technologies continue to evolve, multimedia forensics is no longer limited to deepfake detection, but is progressively expanding toward a broader research landscape encompassing source attribution, content provenance, explainability, adversarial robustness, multimodal forensic analysis, and evaluation under realistic deployment conditions. This evolution requires forensic methodologies capable of operating across heterogeneous generation pipelines while remaining reliable, interpretable, and applicable in real-world scenarios.

The 2nd Deepfake Forensics Workshop (DFF-2026) provides an interdisciplinary forum for researchers and practitioners from multimedia, computer vision, machine learning, cybersecurity, and digital forensics to discuss these emerging challenges and foster the development of next-generation forensic solutions. Building upon the success of the inaugural edition, the second workshop broadens its scientific scope to address the entire lifecycle of AI-generated multimedia, from content generation and manipulation to forensic analysis, authentication, provenance verification, and trustworthy deployment.

This paper reviews the scientific directions that emerged from the first edition of the workshop, discusses how they have shaped the evolution of the DFF initiative, and presents the objectives and research themes that characterize DFF-2026, highlighting its role as a meeting point for the multimedia forensics community in the

era of generative AI. Further information about the workshop is available at <https://iplab.dmi.unict.it/mfs/acm-dff-ws-2026/>.

CCS Concepts

• **Applied computing** → **Computer forensics**.

Keywords

Deepfake Detection, Deepfake Attribution, Adversarial Attacks, Generative Models, Digital Forensics

ACM Reference Format:

Luca Guarnera, Francesco Guarnera, Giovanni Puglisi, Zahid Akhtar, Fernando Pérez-González, and Sebastiano Battiato. 2026. Deepfake Forensics Workshop (DFF-2026): Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media. In *Proceedings of the 34th ACM International Conference on Multimedia (MM '26)*, November 10–14, 2026, Rio de Janeiro, Brazil. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3767308.3842136>

1 Introduction and Motivation

The creation and manipulation of multimedia content has always been a central topic in the multimedia community. For decades, images, videos, and audio recordings have been edited, enhanced, or tampered with using both professional tools and consumer software. However, recent advances in artificial intelligence and generative modeling have radically changed the scale, realism, and accessibility of media manipulation. Multimedia content can be altered through a wide spectrum of techniques, ranging from classical post-processing and compositing to fully synthetic generation and hybrid editing pipelines combining traditional tools with AI-based methods. Highly realistic manipulations can now be produced with minimal effort, sometimes relying only on high-level semantic instructions such as text prompts. As a result, the boundary between authentic and manipulated content is becoming increasingly blurred.



This work is licensed under a Creative Commons Attribution 4.0 International License. *MM '26, Rio de Janeiro, Brazil*

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2213-4/2026/11
<https://doi.org/10.1145/3767308.3842136>

This technological change is already having significant social, legal, and security implications [7, 10]. Manipulated and synthetic media are often used for malicious purposes, including disinformation and propaganda, non-consensual and abusive content (e.g., deepfake pornography), fraud and identity theft, reputational attacks, and the fabrication or alteration of digital evidence. Beyond these direct threats, the widespread diffusion of highly realistic synthetic media is progressively eroding public trust in digital content, giving rise to the phenomenon known as *Impostor Bias* [7], where even authentic media may be perceived as potentially manipulated. This growing uncertainty undermines the credibility of digital evidence and poses significant challenges for journalism, legal investigations, public institutions, and society as a whole.

In this context, the role of multimedia forensics is becoming both more critical and more challenging [2, 3]. The forensic problem is no longer limited to detecting low-quality or visually inconsistent forgeries, but instead concerns the reliable analysis of highly realistic and semantically coherent manipulations. Modern forensic methods are required to detect and localize manipulations across heterogeneous editing and generation pipelines, and to attribute synthetic or manipulated content to specific models, tools, or generation processes. At the same time, forensic techniques must remain effective in a technological landscape where generation and editing methods are rapidly evolving and continuously optimized to remove detectable traces. Therefore, there is a growing need for new tools that are not only accurate, but also robust, interpretable, and suitable for use in real investigative and judicial contexts. Consequently, explainability and transparency become essential requirements, enabling investigators and legal authorities to better understand, verify, and trust the evidence produced by forensic systems. Beyond detection and localization, identifying the origin of manipulated content is becoming an increasingly important forensic objective. Attribution methods aim to determine the generative model, editing pipeline, or software responsible for producing a given piece of synthetic media, providing valuable information for forensic investigations by enabling source identification and provenance analysis [1, 11, 13]. Addressing these challenges requires not only advances in detection algorithms [21], but also comprehensive forensic frameworks integrating robust evaluation protocols, explainable decision-making, and methodologies suitable for deployment in real-world investigative and judicial environments.

At the same time, the field is increasingly challenged by adversarial machine learning and counter-forensic strategies [4, 12, 14]. Existing forensic systems often suffer from limited generalization and remain vulnerable to targeted attacks designed to evade detection or mislead attribution. As generation and manipulation technologies continue to evolve, artifact-specific solutions are becoming progressively less effective, highlighting the need for forensic methods capable of adapting to previously unseen manipulation techniques.

To foster discussion on these emerging challenges and stimulate the development of more robust, generalizable, and trustworthy multimedia forensic methodologies, we organized the 2nd Deepfake Forensics (DFF-2026) Workshop. The workshop brought together researchers working on the forensic analysis of manipulated and AI-generated multimedia content, including images, videos, and

audio, addressing the full spectrum of challenges ranging from detection and localization to attribution, robustness, explainability, and legal implications. The workshop welcomed contributions addressing both the forensic analysis of existing manipulation and generation techniques and the development of novel synthesis and editing methods, to better understand how different manipulation pipelines influence forensic traces and affect the effectiveness of modern forensic methods. Particular attention was devoted to adversarial machine learning, which represents one of the major threats to the robustness of current forensic approaches, as well as to the evaluation of forensic methods under realistic operating conditions, including multimedia content shared through social media platforms and processed by common compression, re-encoding, and post-processing pipelines. Furthermore, the workshop promoted the study of generative and manipulation methods not only as targets of forensic analysis, but also as valuable research tools for discovering novel forensic traces, designing more challenging benchmarks, and investigating new attack strategies. This perspective fosters the development of forensic methods that are robust, generalizable, and suitable for real-world deployment, moving beyond solutions tailored to specific generative models or isolated manipulation artifacts.

2 From DFF-2025 to DFF-2026: Emerging Research Directions

The first edition of the Deepfake Forensics Workshop (DFF-2025) [5] was conceived as a forum to bring together researchers working on multimedia forensics, computer vision, machine learning, and digital security to address the growing challenges posed by AI-generated media. While the primary focus initially centered on deepfake detection and attribution, the accepted contributions naturally expanded the discussion toward a broader set of problems, highlighting the increasing complexity of multimedia forensic analysis in the era of foundation generative models. Rather than focusing exclusively on improving detection accuracy, the contributions emphasized the need for robust, generalizable, explainable, and practically deployable forensic solutions capable of operating in realistic scenarios.

One of the most prominent research directions emerging from the first edition concerns the generalization ability of deepfake detectors. Several contributions addressed the limitations of methods relying on generator-specific artifacts by proposing more transferable representations and evaluation strategies. CLIP-Flow [26] formulated deepfake detection as an anomaly detection problem, learning only from authentic images to improve robustness against previously unseen generators. Similarly, REVEAL [16] introduced a retrieval-based framework exploiting contextual visual similarity instead of conventional discriminative classifiers, demonstrating strong performance under both unseen generators and degraded image conditions. The importance of evaluating detectors beyond controlled laboratory settings was further emphasized by studies investigating detector benchmarking [20], realistic social-network compression pipelines [19], and comparisons between lightweight handcrafted features and deep learning solutions [18]. Collectively, these works highlighted that improving generalization has become

one of the fundamental challenges for next-generation multimedia forensic systems.

A second research direction focused on source attribution and media provenance [13]. While binary detection determines whether content is authentic or manipulated, attribution aims to identify the specific generation pipeline or model responsible for producing synthetic media. This topic was explored from multiple complementary perspectives. Ragaglia *et al.* [22] demonstrated that the latent representation produced by the emerging JPEG AI compression standard already preserves semantic and generator-specific information, enabling efficient deepfake source attribution directly in the compressed domain. In parallel, Di Pierno *et al.* [9] extended attribution to the audio domain by proposing a hierarchical framework capable of recognizing both the generation technology and the specific synthesis model. Attribution was also investigated from the perspective of proactive media authentication through robust diffusion-model watermarking techniques [17]. Together, these contributions demonstrate the growing importance of provenance analysis as a complement to traditional detection systems.

The workshop also highlighted the increasing relevance of adversarial robustness and security-oriented evaluation. As deepfake detectors become widely deployed, understanding their vulnerabilities has become equally important. Cao *et al.* [6] proposed transferable adversarial perturbations capable of attacking multiple deepfake generation pipelines, while Ciftci *et al.* [8] introduced an adversarial framework specifically targeting frame-based deepfake detectors under both white-box and transfer attack settings. Rather than considering adversarial attacks solely as a threat, these contributions emphasized their role in exposing weaknesses of existing forensic systems and guiding the development of more resilient detection methodologies. Beyond visual deepfake detection, the accepted papers further demonstrated the progressive expansion of multimedia forensics toward multimodal analysis and emerging application domains. Several contributions addressed audio deepfake detection and attribution through multi-stream learning [25], continual adaptation strategies assisted by large language models [23], multimodal audiovisual analysis [24], and hierarchical attribution of synthetic speech generators [9]. At the same time, new benchmark datasets and application scenarios were introduced, including large-scale datasets for anime image manipulation detection and localization [27], as well as frequency-based approaches for improving the robustness of face recognition systems against morphing attacks [15]. These works illustrate how multimedia forensics is progressively evolving beyond conventional image-level deepfake detection toward a broader ecosystem involving multimodal evidence, domain-specific datasets, and application-oriented forensic solutions.

The evolution of the DFF initiative was further reinforced through the organization of the *Adversarial Attacks on Deepfake Detectors (AADD) Grand Challenge* [4], which complemented the workshop by promoting standardized and reproducible evaluation of adversarial robustness. By focusing on realistic attack scenarios and benchmarking protocols, the challenge highlighted that future multimedia forensic systems must not only detect manipulated media, but also remain reliable under intentional attacks and previously unseen manipulations. Taken together, the contributions presented at DFF-2025 and the outcomes of the AADD Grand Challenge suggest that

multimedia forensics is evolving from a collection of independent detection algorithms toward a comprehensive discipline encompassing robustness, provenance, explainability, multimodal reasoning, and standardized evaluation protocols. Compared with the inaugural edition, the second DFF Workshop embraces a broader and more interdisciplinary perspective on multimedia forensics. While DFF-2025 established a strong foundation in deepfake detection, attribution, adversarial robustness, and multimodal forensics, the emerging research trends highlighted the need to address additional challenges, including explainability, trustworthy AI, foundation models, content provenance, evaluation under realistic deployment conditions, and the legal and ethical implications of AI-generated media. Building upon these developments, DFF-2026 expands its scientific scope by encouraging contributions that advance not only the performance of forensic algorithms, but also their transparency, reliability, interpretability, and applicability in real-world multimedia ecosystems.

3 Future Research Directions in Multimedia Forensics

The rapid evolution of generative artificial intelligence is continuously reshaping the multimedia ecosystem, introducing increasingly sophisticated synthesis and editing technologies while simultaneously raising new challenges for multimedia forensics. As highlighted by the scientific contributions presented at DFF-2025 and reflected in the broader scope of DFF-2026, future forensic systems must evolve beyond conventional deepfake detection toward comprehensive frameworks capable of addressing authenticity, provenance, robustness, explainability, and trustworthiness across heterogeneous multimedia modalities. In this context, several research directions are expected to play a central role in the next generation of multimedia forensic methodologies.

3.1 Generalization Across Evolving Generative Models

One of the major limitations of current forensic approaches remains their limited ability to generalize across unseen generators and manipulation pipelines. Existing methods are frequently optimized for specific architectures or datasets, making them vulnerable to the rapid evolution of generative AI. Future research should therefore focus on developing model-agnostic representations capable of capturing intrinsic forensic traces that remain stable across different synthesis technologies, including diffusion models, foundation models, multimodal generation systems, and future AI architectures yet to emerge. Improving cross-dataset and cross-generator generalization will remain one of the primary challenges for multimedia forensics.

3.2 Explainability, Trustworthiness, and Multimedia Provenance

As multimedia forensic technologies are increasingly adopted in investigative, judicial, and regulatory environments, explainability becomes a fundamental requirement rather than an optional capability. Future forensic systems should not only provide accurate predictions, but also generate transparent and interpretable

explanations that allow investigators to understand the evidence supporting each decision. At the same time, binary detection alone will no longer be sufficient. Future research will progressively shift toward comprehensive provenance analysis, integrating detection, source attribution, manipulation localization, passive and active authentication, watermark verification, and content credentialing into unified forensic frameworks capable of reconstructing the complete lifecycle of multimedia content.

3.3 Robust Multimedia Forensics in Real-World Environments

Multimedia shared through online platforms undergoes multiple transformations, including compression, transcoding, resizing, filtering, and repeated re-uploading, all of which may significantly alter forensic traces. Furthermore, adversarial machine learning continues to expose vulnerabilities in current forensic systems by producing manipulations specifically designed to evade detection or mislead attribution algorithms. Future research should therefore prioritize the development of forensic methodologies capable of maintaining reliable performance under realistic operating conditions, combining robustness against adversarial attacks with strong generalization across heterogeneous acquisition, editing, and distribution pipelines.

3.4 Agentic AI for Multimedia Forensics

Recent advances in foundation models and agentic artificial intelligence suggest a promising paradigm shift for multimedia forensic analysis. Rather than relying on a single detector or attribution model, future forensic systems may evolve into collaborative ecosystems composed of multiple specialized AI agents, each responsible for a specific forensic task, such as manipulation detection, localization, source attribution, metadata analysis, provenance verification, adversarial robustness assessment, or legal evidence validation. By orchestrating these complementary analyses through reasoning capabilities provided by large foundation models, agentic forensic systems could autonomously aggregate heterogeneous evidence, resolve conflicting hypotheses, retrieve contextual information, and generate comprehensive, human-readable forensic reports. Such intelligent assistants would not replace forensic experts, but rather support investigators throughout the entire analytical process by improving transparency, consistency, and decision-making. This emerging research direction represents one of the most promising opportunities for the multimedia forensics community and may naturally inspire the scientific vision of future editions of the DFF Workshop.

4 DFF-2026 Submissions and Accepted Contributions

DFF-2026 received 18 paper submissions, of which 12 were accepted for presentation at the workshop, corresponding to an acceptance rate of 66.67%. The accepted contributions reflect the broadened scientific scope of the second edition and cover several complementary directions in multimedia forensics. These include generalizable and explainable deepfake detection, synthetic image detection under compression, image manipulation localization, and morphing attack detection. Several works further explore multimodal and

domain-specific scenarios, including audio and audio-visual deepfake analysis, document image tampering, multi-view geometric consistency, and forensic image–claim verification. The program also includes contributions investigating person-specific behavioral and physiological cues, as well as broader regulatory and forensic perspectives on emerging forms of synthetic media. Overall, the accepted papers highlight the increasing diversity of multimedia forensic problems and the shift from binary deepfake detection toward more comprehensive approaches addressing localization, interpretation, robustness, multimodal evidence, and real-world applicability.

5 Discussion and Conclusions

The rapid progress of generative AI is reshaping the multimedia ecosystem and introducing new challenges for digital trust. In response, multimedia forensics is evolving beyond manipulated-content detection toward a broader discipline encompassing attribution, provenance analysis, explainability, robustness, multimodal reasoning, and trustworthy deployment. This evolution requires forensic methodologies that can operate reliably across heterogeneous generation pipelines while remaining interpretable and applicable in real-world scenarios.

The contributions presented at the first DFF Workshop reflected this transition by emphasizing generalization to unseen manipulations, adversarial robustness, explainability, and realistic evaluation settings. These directions motivated the broader scope of DFF-2026, which extends the workshop focus from detection toward the full lifecycle of AI-generated multimedia, including manipulation analysis, authentication, provenance verification, and responsible deployment. Looking ahead, multimedia forensics is expected to move toward integrated forensic ecosystems combining evidence from images, videos, audio, metadata, and contextual information. Advances in foundation models and agentic AI may further support this transition by enabling systems that coordinate specialized forensic tools, interpret heterogeneous evidence, and assist investigators in complex analyses.

The DFF initiative aims to accompany and stimulate this scientific evolution by providing a multidisciplinary forum where researchers from multimedia, computer vision, machine learning, cybersecurity, digital forensics, and related disciplines can exchange ideas and define future research directions. As multimedia generation technologies continue to evolve, future editions of the workshop will continue to broaden their scientific scope, embracing emerging paradigms such as agentic AI, trustworthy multimedia analysis, and human-centered forensic systems. We believe that fostering collaboration across these research communities will be essential for developing the next generation of multimedia forensic methodologies and ensuring the authenticity, integrity, and accountability of digital content in the era of generative AI.

Acknowledgments

This work was supported by the DEFORM project, funded by the European Union’s Horizon Europe Research and Innovation Programme under Grant Agreement No. 101308502, and by the Xunta de Galicia and the European Regional Development Fund under Project ED431C 2025/41.

References

- [1] Wasim Ahmad, Yan-Tsung Peng, and Yuan-Hao Chang. 2025. FAME: A Lightweight Spatio-Temporal Network for Model Attribution of Face-Swap Deepfakes. *Expert Systems with Applications* (2025), 128571.
- [2] Irene Amerini, Mauro Barni, Sebastiano Battiato, Paolo Bestagini, Giulia Boato, Tania Sari Bonaventura, Vittoria Bruni, Roberto Caldelli, Francesco De Natale, Rocco De Nicola, Luca Guarnera, Sara Mandelli, Gian Luca Marcialis, Marco Micheletto, Andrea Montibeller, Giulia Orrù, Alessandro Ortis, Pericle Perazzo, Giovanni Puglisi, Davide Salvi, Stefano Tubaro, Claudia Melis Tonti, Massimo Villari, and Domenico Vitulano. 2025. Deepfake Media Forensics: State of the Art and Challenges Ahead. In *Advances in Social Networks Analysis and Mining*, I-Hsien Ting, Reda Alhajj, Panagiotis Karampelas, and Min-Yuh Day (Eds.). Springer Nature Switzerland, Cham, 33–48.
- [3] Irene Amerini, Mauro Barni, Sebastiano Battiato, Paolo Bestagini, Giulia Boato, Vittoria Bruni, Roberto Caldelli, Francesco De Natale, Rocco De Nicola, Luca Guarnera, et al. 2025. Deepfake media forensics: Status and future challenges. *Journal of Imaging* 11, 3 (2025), 73.
- [4] Sebastiano Battiato, Mirko Casu, Francesco Guarnera, Luca Guarnera, Giovanni Puglisi, Orazio Pontorno, Claudio Vittorio Ragaglia, and Zahid Akhtar. 2025. Adversarial Attacks on Deepfake Detectors: A Challenge in the Era of AI-Generated Media (AADD-2025). In *Proceedings of the 33rd ACM International Conference on Multimedia*. 13714–13719.
- [5] Sebastiano Battiato, Mirko Casu, Francesco Guarnera, Luca Guarnera, Giovanni Puglisi, Orazio Pontorno, Claudio Vittorio Ragaglia, and Zahid Akhtar. 2025. (DFF'25) 1st Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 14317–14319.
- [6] Zhi Cao, Zhongyuan Wang, Run Wang, Yuhong Yang, Feng Tian, Gang Wu, and Atsushi Suzuki. 2025. MGGA: Universal Perturbations against Deepfake via Multiple Model-based Gradient-Guided Feature Layer Attack. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 73–82.
- [7] Mirko Casu, Luca Guarnera, Pasquale Caponnetto, and Sebastiano Battiato. 2024. GenAI mirage: The impostor bias and the deepfake detection challenge in the era of artificial illusions. *Forensic Science International: Digital Investigation* 50 (2024), 301795.
- [8] Umur Aybars Ciftci, Nicholas Solar, Emily Greene, Anthony Rhodes, and Ilke Demir. 2025. Adversarial Reality for Frame-based Deepfake Detectors. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 118–127.
- [9] Andrea Di Piermo, Luca Guarnera, Dario Allegra, and Sebastiano Battiato. 2025. Towards Reliable Audio Deepfake Attribution and Model Recognition: A Multi-Level Autoencoder-Based Framework. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 101–109.
- [10] Europol. 2022. Facing reality? Law enforcement and the challenge of deepfakes. In *An observatory report from the Europol Innovation Lab*, Publications Office of the European Union, Luxembourg.
- [11] Claudio Giusti, Luca Guarnera, and Sebastiano Battiato. 2026. Proto-LeakNet: Towards signal-leak aware attribution in synthetic human face imagery. *Computer Vision and Image Understanding* (2026), 104848.
- [12] Francesco Guarnera, Luca Guarnera, Alessandro Ortis, Sebastiano Battiato, and Giovanni Puglisi. 2025. A Novel Adversarial Gray-Box Attack on DCT-Based Face Deepfake Detectors. *IEEE Access* (2025).
- [13] Luca Guarnera, Oliver Giudice, Matthias Nießner, and Sebastiano Battiato. 2022. On the Exploitation of Deepfake Model Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 61–70.
- [14] Luca Guarnera, Francesco Guarnera, Alessandro Ortis, Sebastiano Battiato, and Giovanni Puglisi. 2024. Evasion Attack on Deepfake Detection via DCT Trace Manipulation. In *International Conference on Pattern Recognition*. Springer, 157–169.
- [15] Marco Huber, Anh Thi Luu, and Naser Damer. 2025. Morphing Resilient Face Recognition by Informed Frequency Selection. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 55–64.
- [16] Mamadou Keita, Wassim Hamidouche, Hassen B Eutamene, Abdelmaïk Taleb-Ahmed, and Abdenour Hadid. 2025. REVEAL: A Retrieval-Augmented Generation Approach for Contextual Identification of Synthetic Visual Content. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 12–20.
- [17] Qiushi Li, Stefano Berretti, and Roberto Caldelli. 2025. Visual Quality Improved Watermarking based on Dual-Reference Loss for Deepfake Attribution. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 128–134.
- [18] Giuseppe Mazzola, Liliana Lo Presti, and Marco La Cascia. 2025. How Well Do Simple Features Detect Fake Faces? A Comparison with Deep Learning. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 37–44.
- [19] Andrea Montibeller, Dasara Shullani, Daniele Baracchi, Alessandro Piva, and Giulia Boato. 2025. Bridging the Gap: A Framework for Real-World Video Deepfake Detection via Social Network Compression Emulation. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 29–36.
- [20] Alan Perotti, Marco Nurisso, and Mirko Zaffaroni. 2025. No Detector to Rule Them All. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 65–72.
- [21] Orazio Pontorno, Mattia Litrico, Luca Guarnera, Mario Valerio Giuffrida, and Sebastiano Battiato. 2026. μ Flow: Leveraging Average Images for Improving Generalisation of Deepfake Faces Detectors. In *European Conference on Computer Vision*. Springer.
- [22] Claudio Vittorio Ragaglia, Lorenzo Catania, Francesco Guarnera, Dario Allegra, and Sebastiano Battiato. 2025. What if Retrieval Could Work Before Decoding? The case of JPEG AI Latents for Deepfake Source Attribution. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 21–28.
- [23] Utkarsh Venaik, Akash Kushwaha, Nabeel Koya A, and Rajiv Ratn Shah. 2025. SASDN: A Generalizable and Minimal-Intervention LLM-Integrated Framework for Continual Adaptation in Spoofed Speech Detection. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 93–100.
- [24] JinYu Wang, Xin Jin, Huaye Wang, and Longteng Jiang. 2025. VAD-Lip: Visual and Audio Deepfake Detection via Lip Features. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 110–117.
- [25] Taiba Majid Wani and Irene Amerini. 2025. Learning to Fuse: A Gated Multi-Stream Framework for Generalized Audio Deepfake Detection. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 84–92.
- [26] Zhipeng Yuan, Kai Wang, Weize Quan, Dong-Ming Yan, and Tieru Wu. 2025. CLIP-Flow: A Universal Discriminator for AI-Generated Images Inspired by Anomaly Detection. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 3–11.
- [27] Chenyang Zhu, Xing Zhang, Yuyang Sun, Ching-Chun Chang, and Isao Echizen. 2025. AnimeDL-2M: Million-Scale AI-Generated Anime Image Detection and Localization in Diffusion Era. In *Proceedings of the 1st on Deepfake Forensics Workshop: Detection, Attribution, Recognition, and Adversarial Challenges in the Era of AI-Generated Media*. 45–54.