GenText-Forensics: Challenge on Explainable Forensics and Adversarial Generation for Text-Centric Images

Fanwei Zeng
Ant Group
Hangzhou, China
riczuefwsct@alu.uestc.edu.cn

Jianshu Li
Ant Group
Singapore, Singapore
jianshu.l@antgroup.com

Changtao Miao
Ant Group
Hangzhou, China
miaoct1024@gmail.com

Chenqi Kong
NTU
Singapore, Singapore
chenqi.kong@ntu.edu.sg

Youru Li
BJUT
Beijing, China
liyouru@bjut.edu.cn

Zhi Cai
BJUT
Beijing, China
caiz@bjut.edu.cn

Haiyan Yin
A*STAR,CFAR
Singapore, Singapore
Yin_Haiyan@a-star.edu.sg

Joey Tianyi Zhou
A*STAR, CFAR
Singapore, Singapore
joey.tianyi.zhou@gmail.com

Ran He
Institute of Automation, CAS
Beijing, China
rhe@nlpr.ia.ac.cn

Anderson Rocha
University of Campinas
São Paulo, Brazil
anderson.rocha@unicamp.br

Alex Kot
Shenzhen MSU-BIT University
Shenzhen, China
eackot@ntu.edu.sg

Abstract

The rapid evolution of AIGC has intensified threats from text-centric image manipulation. Current forensic research and challenges predominantly prioritize facial Deepfakes, leaving the complex domain of text-centric forgery significantly under-explored. Moreover, existing methodologies mainly focus on binary classification or coarse localization, overlooking the critical need for fine-grained interpretability and sophisticated text editing generation. To bridge these gaps in both research scenarios and methodologies, we introduce **GenText-Forensics**, **the first AI security challenge dedicated to text-centric multimedia forensics**. Supported by **RealText-V2**, a large-scale multilingual dataset spanning diverse real-world scenarios, the challenge establishes a unified adversarial framework. **Track 1: (Defense) Forgery Analysis Report Generation** requires generating comprehensive **forgery analysis reports** that integrate detection, spatial grounding, and natural language explanation. **Track 2: (Attack) AIGC Text-Image Editing** focuses on adversarial **AIGC text editing** to expose detection vulnerabilities. This initiative aims to advance the development of **Multimedia Security** against evolving text-centric forgeries.

CCS Concepts

• **Computing methodologies** → Computer vision; • **Security and privacy** → Social aspects of security and privacy.

Keywords

Text-Centric, MLLM, Report generation, Forgery Analysis, AIGC

MM '26, November 10–14, 2026, Rio de Janeiro, Brazil

ACM ISBN 979-8-4007-XXXX-X/26/11
https://doi.org/10.1145/XXXXXXX.XXXXXXX

ACM Reference Format:
Fanwei Zeng, Jianshu Li, Changtao Miao, Chenqi Kong, Youru Li, Zhi Cai, Haiyan Yin, Joey Tianyi Zhou, Ran He, Anderson Rocha, and Alex Kot. 2026. GenText-Forensics: Challenge on Explainable Forensics and Adversarial Generation for Text-Centric Images. 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, 7 pages. https://doi.org/10.1145/XXXXXXX.XXXXXXX

1 Introduction

Sophisticated text-centric forgeries, fueled by rapid advancements in AIGC, pose a significant threat to information authenticity and security. To draw attention to these dangers and accelerate the development of defensive strategies, the research community has organized several notable detection competitions. In response, the research community has organized several detection challenges to advance defensive capabilities. Early efforts such as the *ICDAR 2023 DTT in Images* [2]focused on classifying text manipulations, while the *Image Forgery Detection Challenge (CVPR 2022)* [14] addressed general image tampering localization. In facial manipulation, and the *Deepfake Detection Challenge* [7]. Additionally, the *ALASKA2 Image Steganalysis Challenge* [1] targeted the detection of subtle steganographic traces.

Despite these efforts, current challange face three critical limitations: (1) Most approaches treat image forgery detection as a binary classification problem, lacking fine-grained localization and interpretability. Detection, grounding, and explanation are inherently interrelated subtasks of forgery analysis; unifying them into a comprehensive analysis report can better leverage their synergies and offers greater practical utility than isolated decision scores. (2) Research has largely focused on facial deepfakes or image manipulation in natural scenes, with limited attention to text-centric images. However, manipulations in document-rich images, such as financial statements and medical records, pose security risks comparable to those of facial forgeries. (3) While AIGC-based general image editing has matured, fine-grained text-image synthesis

and tampering remain underexplored. Generating text that is both visually coherent and semantically consistent demands greater precision than generic style transfer, posing unique challenges for both generation and detection. Consequently, comprehensive and fine-grained analysis benchmarks are essential to drive coordinated progress in this domain.

To address these gaps and promote comprehensive adversarial attack-and-defense research in digital media security, we introduce **GenText-Forensics**, built upon our newly released **RealText-V2** dataset. This large-scale, multilingual benchmark covers seven languages across five real-world scenarios: Finance, Education, Medical, E-commerce, and Scenes in the Wild. The challenge establishes an adversarial framework with two complementary tracks.

Track 1: (Defense) Forgery Analysis Report Generation goes beyond binary prediction by requiring participants to produce a structured forensic report that jointly performs three subtasks: (i) forgery detection (authenticity judgment), (ii) forgery grounding (localization via bounding boxes $[x_0, y_0, x_1, y_1]$), and (iii) forgery explanation (natural language descriptions of manipulation clues). Performance is evaluated using a composite metric in Sec 3.4, which combines the classification F1-score, localization mF1, and text similarity scores, with the final assessment conducted by an LLM-based judge guided by rigorous rubrics.

Track 2: (Attack) AIGC Text-Image Editing focuses on the offensive counterpart of defense. Participants employ AIGC techniques to generate high-fidelity adversarial text forgeries that preserve visual quality and text legibility while evading state-of-the-art forensic detectors.

By integrating offensive and defensive perspectives within a single challenge, **GenText-Forensics** aims to raise researchers' awareness of the risks associated with AI-generated text-centric image forgeries through an adversarial attack-and-defense framework and accelerate the research and deployment of deepfake prevention and detection strategies, aligning with ACM MM's "AI for Good" mission to protect digital information integrity.

2 Advisory and Organizer

Advisory Committee:

Anderson Rocha (Fellow, IEEE) received the Ph.D. degree. He is a Full Professor of Artificial Intelligence and Digital Forensics with the Institute of Computing, University of Campinas (Unicamp), Brazil. He is an Elected Affiliate of the Brazilian Academy of Sciences (ABC) and the Brazilian Academy of Forensic Sciences (ABCF). His main research interests include artificial intelligence, digital forensics, reasoning for complex data, and machine intelligence. He is a three-term elected member of the IEEE Information Forensics and Security Technical Committee (IFS-TC) and served as its Chair. In 2023, he was re-elected as IFS-TC Chair for the 2025–2026 term. He is a Microsoft Research Faculty Fellow and a Google Research Faculty Fellow—prestigious academic recognitions awarded by Microsoft Research and Google, respectively. In 2016, he was awarded the Tan Chin Tuan Fellowship by the Tan Chin Tuan Foundation in Singapore. Since 2023, he has also been a Fellow of the Asia Pacific Artificial Intelligence Association (AAIA).Prof. Rocha is ranked among the top 2% of research scientists worldwide according to studies by *PLOS ONE*/Stanford and Research.com. He is currently a LinkedIn Top Voice in Artificial Intelligence for his ongoing efforts to raise public awareness about AI and its societal impacts.**Webpage:** https://www.ic.unicamp.br/~rocha/

Ran He (Fellow, IEEE/IAPR) received the PhD degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences (CASIA), China, in 2009. He has been a professor at the National Laboratory of Pattern Recognition since December 2016. He is now directing the visual perception and machine learning group. He has published two books and more than 200 papers in refereed journals and conference proceedings in the areas of computer vision, pattern recognition, and image processing. He is an editorial board member of *IEEE TPAMI*, *TIP*, *TIFS*, *TCSVT*, *TBIOM*, *IJCV*, and *Pattern Recognition*. He has served as an area chair for ICCV, CVPR, ECCV, ICML, and NeurIPS. His research has received the IEEE SPS Young Author Best Paper Award (2020), the IAPR ICPR Best Scientific Paper Award (2020), and the IEEE Biometrics Council Best Doctoral Dissertation Award (Chaoyou Fu). He was the recipient of the CAS Outstanding Tutor Award in 2022 and 2023.**Webpage:** https://rhe-web.github.io/

Alex Kot (Fellow, IEEE) has been with Nanyang Technological University (NTU), Singapore, since 1991. He headed the Division of Information Engineering at the School of Electrical and Electronic Engineering (EEE) for eight years. He served as Vice Dean (Research) and Associate Chair (Research) for the School of EEE for three years, overseeing research activities across a faculty of over 200 members. He was also Associate Dean (Graduate Studies) for the College of Engineering (COE) for eight years. He is currently the Director of the ROSE Lab [Rapid (Rich) Object SEarch Lab] and the Director of the NTU–PKU Joint Research Institute. He has published extensively, with over 300 technical papers in the areas of signal processing for communications, biometric recognition, authentication, image forensics, machine learning, and artificial intelligence. He holds two U.S. patents and one Singapore patent. Dr. Kot has served as Associate Editor for numerous IEEE Transactions, including *IEEE TSP*, *IMM*, *TCSVT*, *TCAS-I*, *TCAS-II*, *TIP*, *SPM*, *SPL*, *JSTSP*, *JASP*, and *TIFS*. He was a Technical Committee (TC) member for several IEEE Technical Committees in the Signal Processing Society (SPS) and the Circuits and Systems Society (CASS). He has served IEEE in various leadership roles, including General Co-Chair of the 2004 IEEE International Conference on Image Processing (ICIP) and Area/Track Chair for multiple IEEE flagship conferences. He also served as Coordinator of the IEEE Signal Processing Society Distinguished Lecturer Program and as Chapters Chair for IEEE Signal Processing Chapters worldwide. His honors include the Best Teacher of the Year Award at NTU, the Microsoft MSRA Award, and several best paper awards as a co-author. He was elected IEEE CAS Distinguished Lecturer in 2005 and has held senior leadership positions, including Vice President of the IEEE Signal Processing Society and IEEE Signal Processing Society Distinguished Lecturer. He is currently a Fellow of the Academy of Engineering, Singapore; a Life Fellow of IEEE; and a Fellow of the Institution of Engineers, Singapore (IES). He also serves as a Distinguished Professor and Chief Scientist at Shenzhen MSU–BIT University (SMBU).**Webpage:** https://personal.ntu.edu.sg/eackot/index.html

Figure 1: Visual samples from the RealText-V2 dataset used in Track 1. The dataset covers diverse scenarios, including Finance, Medical, Education, E-commerce, and Scenes in the Wild, and multiple languages, such as English, Thai, and Handwritten scripts. The top row shows forged images, while the bottom row presents the corresponding pixel-level ground truth masks. The dashed box on the right displays a sample Forgery Analysis Report for the first image (a bank statement), illustrating the multimodal output requirement that integrates detection results, grounding coordinates, and a natural language explanation.

Organizer Committee:

Fanwei Zeng obtained the M.S. degree from the College of Signal and Information Processing, University of Electronic Science and Technology of China (UESTC). His research interests primarily focus on computer vision, deep learning, MLLMs, and AI security, particularly in facial analysis, spatio-temporal action detection, multimodal reasoning, and forgery analysis. He is currently working as an Algorithm Expert at Ant Group, where he focuses on multimedia forensics and multimodal reasoning algorithms and their industrial applications. He received the Ant Group CEO Special Award from the Security Department in 2021 and 2022. He has published papers in prestigious international journals and conferences, including Pattern Recognition, IEEE ICME, The Web Conference (WWW), and IJWMIP. His research works have received over 200 citations. He actively serves the academic community as a reviewer for conferences such as ICCV, IROS, and IEEE ICME. **Webpage:** https://orcid.org/0000-0003-4218-164X

Jianshu Li obtained the Ph.D. degree from the School of Computing, National University of Singapore in 2019, advised by Prof. Terence Sim and Prof. Shuicheng Yan. His research interests lie primarily in computer vision and image understanding, particularly in face and human analytics, semantic segmentation, and object detection. He has been working as an algorithm expert at Ant Group since 2018, focusing on face analysis algorithms—including face recognition, face liveness detection, face quality assessment, and facial attribute recognition. He received the T-Star Award (Ant Group's Annual Excellent Engineer Award) in 2023, the 1st Prize of the Wu Wen Jun AI Natural Science Award in 2022, and the First Place Award in the CelebA-Spoof Challenge at ECCV 2020. His earlier achievements include the Gold Award at PREMIA 2019 (Singapore), the Best Student Paper Award at ACM MM 2018, the Winner Prize in the Object Localization Track of ILSVRC 2017, and the Winner Prize in the Emotion Recognition Challenge at ICMI 2016. He has published over 20 papers in top-tier journals and conferences and has served as an invited reviewer for CVPR, ECCV, NeurIPS, IJCAI, FG, ICMI, IEEE TIP, TCSVT, and TMM. **Webpage:** https://sites.google.com/view/li-js/home

Changtao Miao received the Ph.D. degree in Cyberspace Security from the University of Science and Technology of China in 2025. He is currently working at Ant Group, Hangzhou, China. His research interests include digital media forensics, multimodal large language model, and AI security. He has published more than 20 papers in international journals or conferences such as IEEE Transactions on Information Forensics and Security (TIFS), IEEE Transactions on Multimedia (TMM), and ACM Transactions on Multimedia Computing Communications and Applications (TOMM). **Webpage:** https://scholar.google.com/citations?hl=zh-CN&user=oWhfPTgAAAAJ

Joey Tianyi Zhou is the Deputy Director and Principal Scientist at the A*STAR Centre for Frontier AI Research (CFAR), Singapore. He also holds a joint appointment as Principal Scientist with the Centre for Advanced Technologies in Online Safety (CATOS). Prior

to joining CFAR, he was a Senior Research Engineer at the Sony US Research Center in San Jose, USA. Dr. Zhou received his Ph.D. degree in Computer Science from Nanyang Technological University (NTU), Singapore. His research focuses on improving the efficiency and robustness of machine learning algorithms. In these areas, he has published over 100 papers and received several awards, including a Best Student Paper Nomination at the European Conference on Computer Vision (ECCV 2016), the Best Paper Award at IEEE SmartCity 2022, and awards at IJCAI workshops. He also received the Best Poster Award and a Runner-up Prize in the HANDS workshop and its competition, respectively, at the International Conference on Computer Vision (ICCV 2019). Dr. Zhou regularly organizes workshops and tutorials at top-tier international conferences such as CVPR, IJCAI, and ICDCS. He serves on the editorial boards of leading journals including *Artificial Intelligence Journal (AIJ)* and *IEEE Transactions*, and has served as an Area Chair for premier machine learning conferences including ICLR, ICML, NeurIPS, and IJCAI. He is also an Associate Programme Chair for IJCAI 2025. He is listed among the Top 2% Scientists Worldwide in 2023 by Stanford University. **Webpage:** https://joeyzhouty.github.io/

Haiyan Yin is a Senior Research Scientist and Early-Career Principal Investigator at the A*STAR Centre for Frontier AI Research (A*STAR CFAR). She received her Ph.D. in Computer Science from Nanyang Technological University (NTU Singapore) and has worked as a Research Scientist at Baidu Research USA and Sea AI Lab. Her research spans agentic AI, reinforcement learning, meta-learning, and trustworthy decision-making, with a focus on building intelligent systems that generalize across tasks, adapt rapidly to new environments, and exhibit self-directed reasoning and planning. Her work on continual, resource-efficient, and trustworthy reinforcement learning has been supported by competitive research grants, and her publications appear in leading AI conferences including *NeurIPS*, *ICLR*, *AAAI*, *IJCAI*, and *AAMAS*. **Webpage:** https://www.a-star.edu.sg/cfar/about-cfar/our-team/dr-yin-haiyan

Youru Li is a Professor with the College of Computer Science, Beijing University of Technology, China. He has received the Ph.D. and M.E. degrees at the Department of Computer Science, Beijing Jiaotong University, Beijing, China, in Jun. 2025 and Jun. 2020, respectively. From Jun. 2020 to Aug. 2021, he worked for Ant Financial, Alibaba, as a machine learning algorithm engineer. His current research interests include data mining, multimodal knowledge graph and their applications on Health Informatic, RecSys and FinTech. He has received Chinese Association for Artificial Intelligence (CAAI) WU Wenjun AI Science and Technology Award, the Chinese Institute of Electronics (CIE) Excellent Master Dissertation Award, and the CIE Excellent Doctoral Dissertation Forum Best Poster Award. He has published over 40 research papers in refereed conferences and journals (e.g., TPAMI, TKDE, TKDD, TOMM, TORS, KDD, AAAI, MICCAI). **Webpage:** https://cs.bjut.edu.cn/info/1408/3871.htm

Zhi Cai is a Professor with the College of Computer Science, Beijing University of Technology, China. He received the MSc degree from the School of Computer Science in the University of Manchester, in 2007, and the PhD degree from the Department of Computing and Mathematics of the Manchester Metropolitan University, U.K, in 2011. His research interests include multimodal information retrieval, ranking in relational databases, keyword search, and intelligent transportation systems. He has published over 50 research papers in refereed conferences and journals (e.g., TPAMI, TKDE, VLDB, ICDE, SIGMOD). **Webpage:** https://cs.bjut.edu.cn/info/1408/1167.htm

Kong Chenqi is currently a Research Fellow at the ROSE Lab, Nanyang Technological University (NTU), where he works on image manipulation detection and localization under the supervision of Prof. Alex C. Kot. He received his B.S. and M.S. degrees from Harbin Institute of Technology, Harbin, China, in 2017 and 2019, respectively. He earned his Ph.D. from the Department of Computer Science at City University of Hong Kong, Hong Kong SAR, China, in 2023. His research interests encompass AI security and multimedia forensics. He has authored or co-authored several papers published in *IEEE TPAMI*, *IEEE TIFS*, *IEEE TDSC*, *IEEE TCSVT*, *IEEE TIP*, as well as top-tier conferences including ICCV, ICML, and ACL. His academic excellence has been recognized with numerous awards and scholarships, including the National Scholarship, a Gold Medal at the International Exhibition of Inventions in Geneva, and the Research Tuition Scholarship. **Webpage:** https://chenqikong.github.io/

3 Competition Format

3.1 Guidelines

The challenge will be hosted on the Kaggle platform (https://www.kaggle.com/). Participants must register on this platform, which will provide training and test data, evaluation scripts, and real-time leaderboards. The competition features two tracks: **Track 1 (Defense): Forgery Analysis Report Generation** and **Track 2 (Attack): AIGC Text-Image Editing**. Participants may choose to compete in either track. Additionally, substantial prizes will be offered to encourage active participation from researchers, engineers, educators, students, and independent developers.

3.2 Schedule

The tentative timeline for the challenge is as follows:

- **March 15, 2026**: Competition registration opens.
- **April 1, 2026**: Release of training and validation datasets.
- **May 1, 2026**: Release of testing datasets.
- **May 15, 2026**: Top 5 teams invited to evaluate on hidden test datasets for each track.
- **May 23, 2026**: Announcement of the top 3 teams for each track.
- **June 1, 2026**: Paper submission deadline.

3.3 Previous Challenges

We previously hosted three successful challenges:

- 2022 Image Forgery Detection Challenge with over 700 global teams.
- 2024 Global Multimedia Deepfake Detection Challenge with more than 1,500 global teams.
- 2025 IJCAI Challenge on Deepfake Detection and Localization with 184 global teams.

Figure 2: Visual demonstrations of Track 2: AIGC Text-Image Editing (Attack). The figure illustrates the diversity of the generation task across different languages (e.g., English, Thai) and scenarios (Scenes in the wild, E-commerce, Education). The top row presents the original images, while the bottom row displays the AIGC-edited adversarial samples. The modifications showcase various editing granularities, ranging from semantic content replacement (Left: "There" → "Here"), fine-grained non-Latin character editing (Middle: Thai text modification), to style attribute transfer (Right: changing text color to blue). These high-fidelity samples are designed to challenge the robustness of forensic models.

3.4 Rules and Evaluation

Competition Rules. We warmly invite researchers from diverse backgrounds, including women, minorities, and young scholars, to participate. To ensure fairness and reproducibility, all participants must adhere to the following rules:

- **Model Constraints:** Each team is allowed only one final model submission. The total number of model parameters must not exceed 32 billion (32B) to encourage efficiency.
- **External Data:** No additional private datasets are allowed. Participants may generate their own training data based on the official dataset provided by the organizers, with no restrictions on the quantity of such generated data.
- **Open Science:** Finalists must submit their training and inference code in Docker format, accompanied by a technical report. Winners are required to publicly release their solutions. All participants must register on the official Kaggle competition page.

Evaluation Framework. To comprehensively assess the performance of both defense and attack models in the context of text-centric image forgery, we have designed rigorous evaluation protocols for each track.

Track 1:(Defense) Forgery Analysis Report Generation In this track, participants are required to generate a structured analysis report in Markdown format. The evaluation measures performance across four distinct dimensions: Detection, Grounding, Explanation, and Overall Report Quality via Rubrics. **Metrics Definition** is as follows:

- **Detection (S_{Det}):** This is a binary classification task determining whether the image is tampered. We use the standard **F1-Score** to measure the trade-off between Precision and Recall.
- **Grounding (S_{Loc}):** To evaluate localization precision, we calculate the **Pixel-level F1-Score** (mF1) and mean Intersection over Union (mIoU) based on the intersection between predicted masks and Ground Truth masks, strictly adhering to the protocol in TruFor [3].
- **Explanation (S_{Exp}):** This metric evaluates the linguistic quality and semantic fidelity of the generated explanation. We utilize **BERTScore** [13] to calculate the semantic similarity between the participant's generated text and the expert-annotated ground truth explanation.
- **Report Quality Rubrics (S_{Rep}):** To assess the logical coherence and professional depth of the report, we employ an advanced LLM Judge (e.g., Qwen3-MAX or GPT-4o) to conduct a rubric-based evaluation. Inspired by the paradigm [4], we define fine-grained scoring rubrics (normalized to 0-100) covering three critical dimensions: **Factuality** (accuracy of

Table 1: Comparison of RealText-V2 with existing text-centric forgery datasets. RealText-V2 is the first large-scale benchmark to support multi-lingual analysis with comprehensive annotations for detection (Det), grounding (Mask), and explanation (Expl).

Dataset	Total	Text Line	Multi Lang	Det	Mask	Expl
T-IC13 [10]	462	✓	✗	✗	✗	✗
T-SROIE [11]	986	✓	✗	✗	✗	✗
OSFT [9]	2938	✓	✗	✗	✗	✗
DocTamper [8]	170K	✗	✗	✗	✓	✗
RealText-V1 [12]	5397	✓	✗	✓	✓	✓
RealText-V2 (Ours)	**500K+**	✓	✓	✓	✓	✓

verdict and evidence), **Reasoning** (logical deduction from visual clues), and **Completeness** (coverage of manipulated regions and format compliance). This ensures the evaluation aligns with human expert standards.

The final ranking is determined by a weighted sum of the four components, placing significant emphasis on both the detection accuracy and the overall quality of the forensic report:

$$\text{S\_Fin} = 0.3 \times S_{Det} + 0.2 \times S_{Loc} + 0.15 \times S_{Exp} + 0.35 \times S_{Rep} \quad (1)$$

Track 2: (Attack) AIGC Text-Image Editing The objective of this track is to generate high-fidelity, adversarial text forgeries. The evaluation employs the **VIEScore** [6], which balances semantic consistency and perceptual quality.

$$\text{VIEScore} = \sqrt{S_{SC} \times S_{PQ}} \quad (2)$$

- **Semantic Consistency (S_{SC}):** Unlike facial deepfakes that rely on identity similarity, text forgery requires textual precision. We adapt the SC Score as follows:

$$S_{\text{SC}} = \alpha \cdot \text{OCR}(T_{\text{gen}}, T_{\text{orig}}) + \beta \cdot \text{CosSim}(I_{\text{gen}}, I_{\text{orig}}) \quad (3)$$

  where OCR denotes the OCR accuracy between the generated text T_{gen} and the original text T_{orig} in the edited region, measuring attack success; CosSim (computed via CLIP visual encoders) quantifies the cosine similarity between the non-edited regions of the generated image I_{gen} and the original image I_{orig}, ensuring visual consistency.
- **Perceptual Quality (S_{PQ}):** This metric evaluates the realism of the forgery. It aggregates scores from:
  (1) **Visual Fidelity:** Measured by MUSIQ [5] to ensure no visual degradation.
  (2) **Anti-Forensics:** The probability of bypassing baseline detection models (Attack Success Rate against detectors).
  The range for S_{PQ} is normalized to 0-10, where higher scores indicate more realistic and harder-to-detect samples.

4 Teams Size and Logistic Requirements

For our upcoming competitions, we anticipate a minimum of 4,000 participants, forming up to 2,000 global teams, based on the success of previous challenges. We require space for the top team's oral presentations, equipped with a projector and multiple AC outlets. No special logistical arrangements are necessary. Additionally, we would appreciate support for a coffee break and water for the speakers.

5 Datasets

5.1 RealText-V2 Dataset

To support the rigorous requirements of **GenText-Forensics**, we introduce **RealText-V2**, a large-scale, multi-dimensional benchmark specifically curated for text-centric image forensics. As illustrated in **Table 1**, existing datasets often suffer from limited scale, lack of multi-lingual support, or absence of fine-grained reasoning annotations. Building upon the foundation of RealText-V1, V2 represents a significant leap in scale and diversity, expanding the dataset volume to over **500,000 images**. RealText-V2 is the first large-scale benchmark dedicated to diverse text scenarios, ranging from **sparse text** in natural scenes to **dense text** in complex documents. To simulate real-world complexity, the dataset spans **7 different languages** and covers **5 critical scenarios**.

- **For Defense (Track 1):** As shown in **Figure 1**, we provide pixel-level grounding masks and reasoning-oriented natural language explanations for "Hard Samples," enabling the training of explainable forensic agents.
- **For Attack (Track 2):** We incorporate high-fidelity adversarial samples generated by advanced AIGC techniques, such as content replacement and style transfer, as demonstrated in **Figure 2**.

By combining scale, scenario diversity, and rich semantic annotations, RealText-V2 serves as a comprehensive testbed for both defensive analysis and adversarial generation tasks.

References

[1] Frédéric Cayre, Patrick Bas, et al. 2019. The ALASKA Steganalysis Challenge: A First Step Towards Steganalysis "into the wild". In *Proceedings of the 7th ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec '19)*. 1–6. https://doi.org/10.1145/3338533.3364464

[2] CompHub and Alibaba Group. 2023. ICDAR 2023 Competition on Detecting Tampered Text in Images. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR) Workshops*. Organized by Tianchi, Alibaba Group.

[3] Fabrizio Guillaro, Davide Cozzolino, Avneesh Sud, Nicholas Dufour, and Luisa Verdoliva. 2023. TruFor: Leveraging All-Round Clues for Trustworthy Image Forgery Detection and Localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 20606–20615.

[4] Zhiyuan Guo, Yixin Wang, Yang Liu, et al. 2025. Rubrics as Rewards: Reinforcement Learning Beyond Verifiable Domains. *arXiv preprint arXiv:2507.17746* (July 2025).

[5] Jiaxin Ke, Qiang Wang, Yajing Wang, and Jian Yang. 2021. MUSIQ: Multi-scale Image Quality Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 5148–5157.

[6] Ziyi Liu, Yiren Wang, Han Zhang, et al. 2023. VIEScore: A Visual Instruction-Guided Explainable Metric for Evaluating Conditional Image Generation. *arXiv preprint arXiv:2312.14867* (Dec. 2023).

[7] Organizers of the ACM Multimedia Grand Challenge. 2025. (MillionDeepfake) The 2nd One Million Deepfakes Detection Challenge. In *Proceedings of the 33rd ACM International Conference on Multimedia (ACM MM '25) Grand Challenge Track*. Challenge based on the AV-Deepfakes-1M++ dataset. URL: https://deepfakes1m.github.io/2025.

[8] Chenfan Qu, Chongyu Liu, Zhenyu Liu, Chang Zhang, and Lianwen Jin. 2023. Towards Robust Tampered Text Detection in Document Image: New dataset and New Solution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16135–16145.

[9] Chenfan Qu, Yiwu Zhong, Fengjun Guo, and Lianwen Jin. 2024. Revisiting Tampered Scene Text Detection in the Era of Generative AI. *arXiv preprint arXiv:2405.15875* (2024).

[10] Yuxin Wang, Hongtao Xie, Mengting Xing, Jing Wang, Shenggao Zhu, and Yongdong Zhang. 2022. Detecting tampered scene text in the wild. In *European Conference on Computer Vision*. Springer, 215–232.

[11] Yuxin Wang, Boqiang Zhang, Hongtao Xie, and Yongdong Zhang. 2022. Tampered text detection via RGB and frequency relationship modeling. *Chinese Journal of Network and Information Security* 8, 3 (2022), 29–40.

[12] Fanwei Zeng, Changtao Miao, Jing Huang, Zhiya Tan, Shutao Gong, Xiaoming Yu, Yang Wang, Huazhe Tan, Weibin Yao, and Jianshu Li. 2025. LogicLens: Visual-Logical Co-Reasoning for Text-Centric Forgery Analysis. arXiv:2512.21482 [cs.AI] https://arxiv.org/abs/2512.21482

[13] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *International Conference on Learning Representations (ICLR)*.

[14] Yuyang Zhou, Xian Liu, Mauro Barni, et al. 2022. The 2nd Image Forgery Analysis, Detection and Localization Challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 4167–4175. https://doi.org/10.1109/CVPRW56491.2022.00483