DeepAuthVerify—A Modular Framework for Deepfake Detection in Facial Authentication Systems

Domenico Di Palma, Alexander Lawall, Kristina Schaaff
IU International University of Applied Sciences
Erfurt, Germany
{alexander.lawall | kristina.schaaff}@iu.org

***Abstract*—The rise of deepfake technologies poses a significant threat to biometric authentication systems, especially those based on facial recognition. In our study, we investigate the reliability of commercial facial recognition systems when exposed to deepfake attacks and propose a modular authentication solution (*DeepAuthVerify*) that integrates deepfake detection into the verification process. We developed *DeepAuthVerify* as a two-layered system combining deep learning-based face recognition and feature extraction with the semantic interpretability of a Large Language Model (LLM) for decision-making. Despite achieving lower accuracy (66.89%) compared to commercial solutions (OpenCV: 91.43%, Amazon Rekognition: 93.80%), *DeepAuthVerify* demonstrates the potential as a complementary layer for deepfake detection, enhancing transparency and modularity. The results indicate that commercial systems, when properly configured, offer robust protection against deepfake attacks. However, their black-box nature limits adaptability and auditability. Our proposed system provides a novel, extensible architecture that fosters explainability and integration into existing authentication environments. In addition to the evaluation, we publicly release the evaluation pipeline to allow reproducibility and comparability of future research.**

***Keywords*-Deepfake Detection; Facial Recognition; Authentication Systems; Large Language Models.**

I. Introduction

The rise of deepfake technologies, synthetically generated or manipulated images and videos created using deep learning techniques, poses a growing threat to biometric authentication systems [1], particularly those based on facial recognition. While these systems are increasingly adopted in Multi-Factor Authentication (MFA) due to their convenience and user acceptance [2], their vulnerability to sophisticated impersonation attacks remains a critical security concern.

Recent advancements in generative models, such as Generative Adversarial Networks (GANs) [3] and Variational Autoencoders (VAEs) [4], have enabled the realistic creation of fake identities that can evade the detection by recognition algorithms. Simultaneously, user-friendly deepfake tools like Reface [5] and DeepFaceLab [6] have reduced the technical barrier for attackers. As reported by the Entrust Cybersecurity Institute, deepfake attacks in identity verification contexts are increasing significantly, with one attempt occurring approximately every five minutes as of 2024 [7].

In our study, we propose *DeepAuthVerify*, a novel, modular authentication framework that augments traditional recognition with a deepfake detection layer using deep learning and semantic analysis through a Large Language Model (LLM). Moreover, we systematically evaluate the resilience of facial recognition systems against deepfake attacks using a data corpus created based on the *Celeb-DF* dataset [8].

Our key contributions are:

1) A hybrid system combining deep learning-based face recognition, facial feature extraction, and LLM-based decision logic.
2) An approach that enhances modularity, interpretability, and integrability in authentication contexts.
3) A novel integration of structured semantic explanations to support transparent deepfake classification and human oversight.
4) Public release of implementation code and test setup to foster reproducibility and future research [9].

The remainder of the paper is structured as follows. Section II outlines the theoretical background of biometric authentication, deepfake generation, and detection technologies, with emphasis on LLM-based reasoning. Section III reviews related work on deepfake detection methods, highlighting their limitations in transparency and integration. Section IV introduces the design of the DeepAuthVerify framework, detailing its requirements, modular architecture, and two-phase verification process. Section V presents the evaluation methodology, datasets, and empirical results compared to commercial systems. Section VI explores system integration options, including standalone, parallel, and pre-filtering deployment modes. Section VII concludes with key findings, limitations, and directions for future research on explainable and adaptive authentication systems.

II. Theoretical Background

This section outlines the (i) biometric authentication and the technologies that enable facial recognition, and (ii) machine learning foundations underlying deepfake generation and detection, including recent advances in LLMs.

A. Biometric Authentication and Facial Recognition

MFA enhances system security by combining independent factors: knowledge (e.g., passwords), possession (e.g., smartphones, tokens), and biometrics (e.g., fingerprints or facial features) [2]. Among these, facial recognition has emerged as one of the most widely adopted due to its usability and low user friction [10]. However, it introduces new attack vectors, such as presentation attacks and digital identity forgery, especially in remote settings.

Facial recognition systems typically follow a pipeline with face detection, feature extraction, and identity verification [11]. Early approaches include feature-based and template-matching methods [12], as well as holistic techniques like the Eigenfaces algorithm [13]. Recent advances use three-dimensional face modeling like in Apple's Face ID [14] and Google's Face Mesh [15]. Despite high accuracy, commercial Application Programming Interfaces (APIs), such as Amazon Rekognition and OpenCV, are typically closed-source and provide limited interpretability, making them hard to audit in sensitive applications.

B. Generative AI, Deepfake Detection, and LLM Reasoning

Modern face recognition and manipulation systems rely on deep learning. Convolutional Neural Networks (CNNs) are particularly effective at extracting facial features and achieving high classification performance under varying conditions [16]. Models like FaceNet [17] have demonstrated their strength in person identification and clustering tasks.

In parallel, generative models, such as GANs and VAEs, have enabled the creation of highly realistic synthetic face images, known as deepfakes. Techniques include face swapping, reenactment, and full-face synthesis [18], often implemented in publicly available tools, such as DeepFaceLab or StyleGAN [19]. The accessibility of such tools raises security concerns, as even low-skilled attackers can generate high-quality forgeries.

Detection models have been proposed, mostly using CNN-based binary classifiers trained on labeled datasets, such as *Celeb-DF*, to counteract these threats. While effective, these systems often operate as non-transparent detectors with limited reasoning capacity.

LLMs, such as GPT [20] and BERT [21], have shown promising results in bridging this gap by offering contextual understanding and semantic interpretation. Built on transformer architectures [22], LLMs can synthesize structured input into human-readable justifications. Their applications in anomaly detection, adversarial reasoning, and decision support have triggered increasing interest in the cybersecurity domain [23], where transparency and interpretability are relevant.

C. Implications for Authentication Systems

The convergence of these technologies enables both advanced biometric verification and new attack vectors. While commercial systems achieve high accuracy under ideal conditions, their susceptibility to deepfake manipulation and lack of interpretability raise critical concerns [24]. Integrating deepfake detection components into authentication pipelines, particularly those combining deep learning with LLM reasoning, offers a possibility for enhanced robustness and transparency.

III. Related Work

The detection of deepfakes has become an active research area due to their increasing misuse in identity fraud, misinformation, and biometric spoofing. Numerous approaches have emerged that influence advances in computer vision, signal analysis, and adversarial learning to distinguish manipulated content from authentic input.

Deepfake detectors often use CNNs to capture subtle spatial or temporal inconsistencies in face-swapped or synthesized videos. MesoNet [25], XceptionNet [26], and Capsule Networks [27] are among the architectures that have shown promising results on benchmark datasets, such as FaceForensics++ [26], ForgeryNet [28], and Celeb-DF [8]. These models often achieve better performance when trained on large-scale datasets that include both authentic and manipulated samples.

The vulnerability of facial recognition systems to presentation attacks and deepfakes has been widely studied [29], [30]. Even state-of-the-art face recognition APIs can be deceived by high-quality synthetic content [24]. As a countermeasure, ensemble classifiers [31] and temporal analysis models [32] have been proposed to improve robustness in real-time verification systems. Nevertheless, these methods often suffer from lack of transparency and limited interoperability with commercial authentication workflows.

Recent work has explored combining deep learning-based feature extraction with interpretable or modular architectures. For example, [33] discuss the integration of transformer-based language models in security incident response systems, highlighting the role of contextual reasoning. These approaches point toward a new generation of hybrid systems that prioritize transparency and human-aligned decision-making; yet their application in biometric authentication remains limited.

While prior research has explored detection accuracy and network architectures, our work contributes a novel perspective by integrating a deepfake detection module with a semantic reasoning layer into a facial authentication pipeline. Unlike black-box detectors, our system emphasizes modularity, transparency, and interpretability. We aim to bridge the gap between detection research and deployable security systems.

IV. System Design

In this paper, we propose *DeepAuthVerify*, a modular, two-layered authentication system that integrates deepfake detection and explainable decision-making into the verification process to address the increasing threat posed by deepfakes in facial recognition-based authentication. This section details the functional and non-functional requirements of the system, followed by an architectural overview of its layered design and operational flow.

A. Design Requirements

The design of *DeepAuthVerify* is driven by several core requirements, which reflect both practical integration and current research challenges in biometric security:

- **(R1) Compatibility:** The system must be compatible with existing biometric MFA infrastructures, particularly those using commercial face recognition APIs (e.g., Amazon Rekognition, OpenCV).
- **(R2) Robustness against Deepfakes:** The system must reliably detect synthetic facial images generated through deep learning methods (GANs, VAEs, etc.).

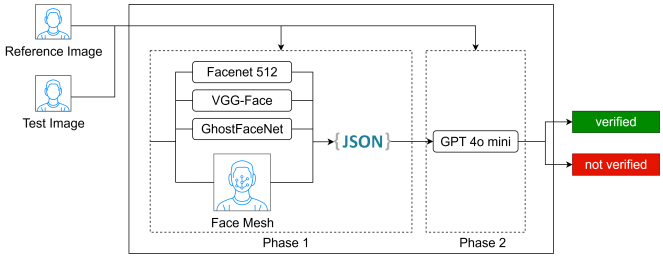


Figure 1. High-level Architecture of *DeepAuthVerify*

- **(R3) Explainability:** Decisions should be transparent and accompanied by interpretable reasoning to improve auditability and trust, especially in ambiguous or borderline cases. This is especially important in regulatory-sensitive environments to improve auditability and trust.
- **(R4) Modularity and Extensibility:** System components (e.g., detection modules, APIs, reasoning layer) must be loosely coupled to support individual updates and enhancements.

B. Design Rationale & Advantages

The modular architecture of *DeepAuthVerify* ensures the separation of concerns and simplifies both testing and future extension. For instance, the deepfake detection layer can be updated with new deep learning models without affecting the face verification logic. Similarly, the LLM layer can be replaced by task-specific models or rule-based systems, depending on deployment requirements and privacy constraints.

A key innovation lies in the semantic reasoning layer. Rather than producing a binary decision alone, the system generates an explanatory narrative that enables human reviewers to understand the rationale behind the verdict. This supports informed escalation in edge cases and fulfills demands for AI transparency in critical identity verification workflows.

Additionally, the architecture supports deployment flexibility. Components can be containerized and orchestrated via microservices, making the system suitable for both cloud-based and on-premise environments.

C. Architectural Overview

DeepAuthVerify follows a modular and multi-phase architecture that integrates deep learning-based face recognition and facial feature extraction with an LLM to verify the authenticity of facial input data. This section outlines the design and implementation of each architectural phase in detail.

Figure 1 illustrates the system's high-level architecture. The input to the system are two facial images, which are processed through each stage sequentially. Each module is independently operable and can be adapted or replaced based on specific authentication requirements. This means the system exposes methods that manage facial image input, feature extraction, and classification. The modular structure allows for reusability and supports the replacement of individual modules, such as the landmark extraction pipeline or the classification algorithm.

The verification process follows two phases: The classification by a deep learning model with facial embedding extraction and the semantic validation.

1) Phase 1—Classification and Embedding Extraction: In the first phase, facial landmarks are extracted using the Face Mesh technology from the MediaPipe framework [15]. This technique identifies three-dimensional facial landmarks, which are then normalized to ensure scale and pose invariance. These normalized landmarks serve as the basis for constructing embedding vectors used in similarity comparisons.

The system integrates three pretrained deep learning models for facial feature extraction to enhance representational power and robustness:

- FaceNet 512 [17]: generates compact embeddings using triplet loss
- VGG-Face [34]: CNN model trained on a large-scale face dataset
- GhostFaceNets [35]: lightweight model optimized for fast and efficient inference. Each model produces feature vectors that are compared to reference embeddings using cosine similarity.

Each model produces feature vectors that are compared to reference embeddings using cosine similarity. Additionally, key facial landmarks are extracted using the Face Mesh library and included in the result. The final output is returned as a structured JSON object containing the model name, verification result, threshold, cosine distance, detector backend, and the extracted facial landmarks. The system architecture allows for flexible integration, enabling these models to be exchanged or extended at any time without major adjustments.

2) Phase 2—Semantic Validation via LLM: The extracted facial features and metadata are analyzed using an LLM. We implemented the LLM integration using OpenAI's GPT API with custom prompting logic. The prompt is designed to guide the LLM in semantically interpreting the context, such as inconsistencies in facial symmetry, unnatural artifacts, or landmark misalignments. The transmitted JSON serves as support. The LLM acts as a semantic validator, assessing the likelihood that a given image is synthetically generated. Additionally, the output includes a detailed explanation of the classification result. For example, in the case of an input image classified as manipulated, the explanatory output may look as follows:

> "Discrepancies detected in left jawline contour and reflection inconsistency in the left eye region. Landmarks appear overly symmetric compared to the reference face, suggesting GAN-based synthesis."

This explanation enables reviewers to understand the rationale behind a rejection, rather than relying solely on a similarity score or binary decision. Such interpretability is crucial in edge cases or escalated verification workflows.

3) Integration Challenges: While each building block of *DeepAuthVerify*, such as face embeddings, landmark extraction, and the LLM-based semantic validator, has been studied in isolation, their combination introduces non-trivial challenges. These include synchronizing heterogeneous outputs

across modules, preventing error propagation between the embedding similarity layer and the LLM interpretation, managing additional latency overhead, and maintaining consistent thresholds across components. We emphasize these integration issues as part of the motivation for adopting a modular design that allows individual components to be improved or replaced without destabilizing the overall framework.

V. System Evaluation

This section presents the evaluation of *DeepAuthVerify* and the comparison with commercial systems.

A. Evaluation Setup

We conducted a structured evaluation to assess the effectiveness and robustness of *DeepAuthVerify*. In total, we used a set of 1,178 image pairs based on the *Celeb-DF* data set [8]. As the *Celeb-DF* data set contains videos only, for our tests, we generated image pairs from the video data. The image pairs are distributed across the following test variants:

- Test set 0—Control group (342 image pairs): two unaltered images of the same person, variations in facial expression, lighting, or angle.
- Test set 1—Deepfake with preserved context (218 image pairs): One of the two images has been manipulated using deepfake techniques to replace the face, while background, pose, and clothing remain unchanged. Only the face is synthetic; the rest of the image context is identical.
- Test set 2—Deepfake with altered context (618 image pairs): The manipulated image includes both a deepfaked face and altered context (e.g., background, pose, facial expression). Both the face and the surrounding scene are synthetically modified.

The test cases consist of image comparisons distributed across four different gender categories. The dataset is composed of male (54.92%) and female (41.94%) subjects, with non-binary (1.87%) and unknown (1.27%) gender entries. A test is successful if, in test set 0, the system correctly identifies the images as matching, and in test sets 1 and 2, it correctly identifies the images as non-matching.

B. Evaluation Results

We tested our final system using the three test variants presented in Section V-A. The aim was to assess the accuracy and robustness of the optimized system under identical conditions. To determine the most suitable threshold for the selected test dataset, we performed a Receiver Operating Characteristic (ROC) analysis, allowing us to identify the optimal decision threshold that balances sensitivity and specificity.

After optimization, we were able to achieve an accuracy of 66.89%, showing a general classification performance despite the complexity of the task. The F1-score of 47.68%, shows a balance between detection sensitivity and false positive control. Our system achieved a precision of 44.09%, indicating that nearly half of the images identified as manipulated were correctly classified. Moreover, the system yielded a recall of 52.34%, which reflects its ability to detect more than half of all correct images correctly.

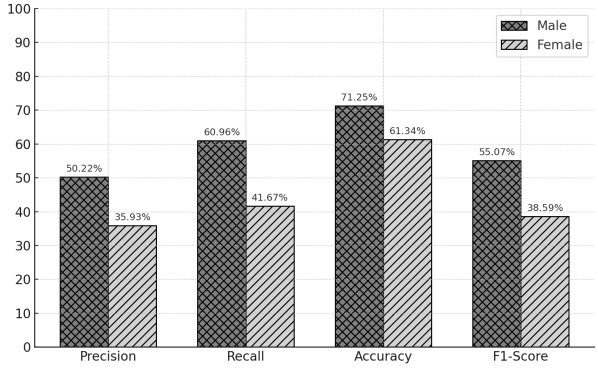


Figure 2. Gender-Specific Performance of *DeepAuthVerify*

In addition to the overall evaluation, we analyzed the performance of the optimized system with respect to gender-specific differences (cf. Figure 2). The test dataset was approximately balanced across male and female subjects to ensure a fair comparison. The results revealed a discrepancy in detection performance between the two groups.

For male subjects, the system achieved consistently higher scores across all metrics, including precision (50.22%), recall (60.96%), accuracy (71.25%), and F1-score (55.07%). In contrast, the performance for female subjects was substantially lower, particularly in recall (41.67%) and F1-score (38.59%), indicating that the system was less effective at detecting manipulated images in this group. The lower precision and accuracy for female subjects suggest a higher rate of false positives and an increased overall classification error. These findings point to a gender-related performance disparity in the optimized model.

This performance gap may be attributed to differences in facial structure, image variability, or bias introduced during the model's training phase. It emphasizes the importance of addressing demographic fairness in biometric verification systems.

C. Comparison with Commercial Systems

To assess the effectiveness of *DeepAuthVerify*, we conducted a comparative evaluation against two established commercial facial recognition solutions: Amazon Rekognition and OpenCV. All three systems were tested using the same balanced dataset, which included both authentic and manipulated facial images. This ensured a consistent evaluation environment across all systems. We focused on four key performance metrics: precision, recall, accuracy, and F1-score, providing the system's strengths and limitations in detecting manipulated identities.

DeepAuthVerify shows lower performance across all core classification metrics compared to the commercial systems Amazon Rekognition and OpenCV. *DeepAuthVerify* achieves a precision of 44.09%, whereas Amazon Rekognition reaches 85.87% and OpenCV 80.82%, indicating that *DeepAuthVerify*

generates significantly more false positives when identifying manipulated images. The difference in recall is even more pronounced: *DeepAuthVerify* identifies only 52.34% of all manipulated samples correctly, while Amazon Rekognition and OpenCV reach 94.15% and 92.40%, respectively. This gap reveals a limited sensitivity of *DeepAuthVerify* to actual deepfakes. In terms of overall accuracy, *DeepAuthVerify* achieves 66.89%, which is lower than Amazon Rekognition (93.80%) and OpenCV (91.43%). This also reflects the combined weaknesses in both precision and recall. The F1-score, which balances precision and recall, further illustrates the disparity: 47.68% for *DeepAuthVerify* versus 89.82% for Amazon Rekognition and 86.22% for OpenCV. This confirms that *DeepAuthVerify* currently lacks robustness and reliability under real-world conditions, though it demonstrates the conceptual feasibility of a hybrid LLM-integrated verification pipeline.

Although *DeepAuthVerify* underperforms in accuracy due to limited data and complexity, it adds interpretable semantics and decision reasoning that are absent in commercial APIs. Moreover, the evaluation highlights that hybrid systems combining visual detection with semantic interpretation can support human decision-making in ambiguous or adversarial input scenarios.

D. Discussion

The evaluation results reveal a trade-off between raw detection accuracy and system transparency. While Amazon Rekognition and OpenCV achieve higher recognition rates, they operate as black-box models with no interpretability or context-aware feedback. While our evaluation demonstrates that commercial services provide higher accuracy in controlled settings, these are limited in terms of transparency and interpretability. Our approach trades a portion of accuracy for improved explainability and modular design. This trade-off is particularly relevant in high-risk identity verification contexts, such as remote onboarding or digital voting, where system outputs must be auditable and justifiable. This raises concerns in domains where traceability, user trust, and regulatory compliance (e.g., General Data Protection Regulation, AI Act) are essential.

DeepAuthVerify prioritizes explainability through a layered architecture that incorporates a semantic reasoning component. Nevertheless, qualitative analysis indicates that the added interpretability can enhance human-in-the-loop decision making, especially in edge cases. Another strength of *DeepAuthVerify* lies in its modularity. Each layer (face verification, detection, reasoning) can be independently updated or replaced without altering the core logic. This design enables quick adaptation to emerging deepfake techniques or new recognition APIs and supports potential integration with other biometric modalities, such as voice or gait.

VI. System Integration Options

DeepAuthVerify is designed for seamless integration into existing authentication systems. In our evaluation, we assessed the system as a stand-alone module, fully replacing the traditional facial recognition pipeline. This allowed for an isolated analysis of its detection capabilities and semantic reasoning logic.

However, one of the core strengths of *DeepAuthVerify* lies in its modular architecture, which supports flexible deployment strategies beyond full replacement. Therefore, we propose the following integration variants, which are illustrated in Figure 3. In particular, integration variants 2 and 3 offer promising options for real-world use cases:

- **Variant 1: Full Replacement**—*DeepAuthVerify* replaces the system layer entirely as in our evaluation.
- **Variant 2: Parallel Evaluation (Veto Layer)**—The system operates alongside commercial recognition APIs. Its classification output or explanation can override or verify existing decisions, enabling more transparent decision pipelines.
- **Variant 3: Pre-filtering Stage**—*DeepAuthVerify* works as a deepfake screening layer prior to conventional recognition, filtering out manipulated inputs before further processing.

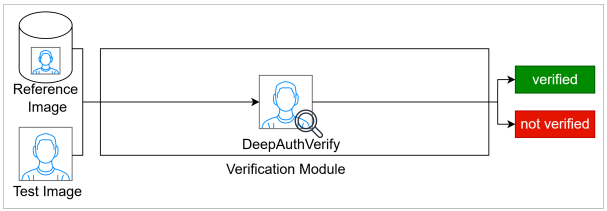
These hybrid integration modes highlight a major advantage of our approach: commercial systems can be extended with a semantic explanation component without altering their internal architecture. This enables organizations to enhance the auditability and trustworthiness of their authentication workflows by adding explainable, AI-assisted reasoning without reducing the performance benefits of mature commercial solutions.

VII. Conclusion & Future Work

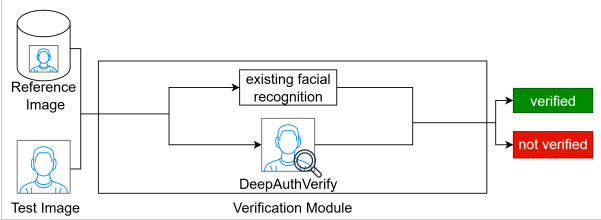
In our paper, we presented *DeepAuthVerify*, a novel, modular authentication system that augments commercial face recognition with deepfake detection and semantic reasoning. It was systematically evaluated on a subset of the *Celeb-DF* dataset that commercial APIs remain effective under clean conditions. Under default threshold settings, they detect deepfakes reasonably well. However, such optimizations are often tailored to the specific test dataset and may not generalize well to unseen data.

Our approach addresses this gap by using deep learning models for robust facial feature extraction, combined with an LLM to enable transparent and interpretable decision-making. This architecture provides a solid foundation for future authentication systems that are explainable, secure, and adaptive, even in adversarial settings, while the system's performance can be enhanced. *DeepAuthVerify* introduces an explainability-first design. The inclusion of LLM-driven semantic feedback empowers system operators to trace, understand, and document decisions in high-stakes environments, offering an advantage over commercial black-box solutions. As regulatory frameworks (e.g., European Union AI Act) increasingly require transparent AI reasoning, our system is positioned as a compliant and auditable alternative.

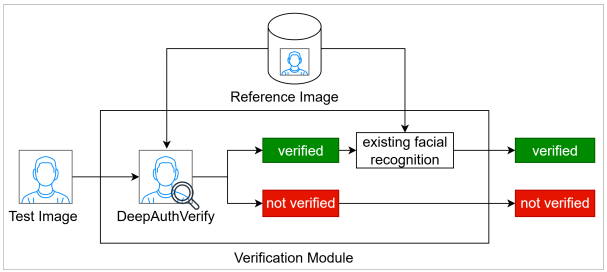
Valuable insights aside, our study has its limitations. The evaluation was conducted under controlled conditions using a predefined test dataset, which may limit the extent to which



(a) Variant 1: Full Replacement



(b) Variant 2: Parallel Evaluation (Veto Layer)



(c) Variant 3: Pre-filtering Stage

Figure 3. System Integration Options of *DeepAuthVerify*

findings can be generalized to real-world scenarios. The deep learning models in the first phase and the configuration of the large language model were based on standard parameters, providing a solid baseline but leaving room for further optimization. The similarity score generated by the LLM is based on internal mechanisms that are not fully transparent, which may pose challenges for interpretability in specific cases. Additionally, aspects, such as potential bias effects, e.g., related to gender, were explored and would benefit from a more comprehensive analysis in future research.

Future research includes advancements in prompt engineering for the LLM, the adoption of newer, higher-performing GPT architectures, and the calibration of model-specific threshold values. Moreover, the fine-tuning or selection of more effective deep learning models for the initial processing phase could yield further improvements. In the long term, extending the system architecture to support video-based analysis represents a promising direction. The integration of traditional image processing methods with explainable AI techniques may also contribute to the development of transparent and reliable security solutions for practical deployment scenarios.

Continuous advancement of deepfake detection remains necessary, as generative models are constantly evolving. Without regular updates to the detectors, their effectiveness against new types of manipulation may decline, potentially impairing the reliability of facial recognition [36]. Additionally, the system's design as an API-compatible component facilitates its integration into existing authentication workflows. Deployment on cloud platforms, such as Amazon Web Services, would enable modular connectivity with diverse systems, thereby enhancing scalability and adaptability.

REFERENCES

[1] A. Lawall and P. Beenken, "Subject-and Process-Oriented Comparison of Multi-factor Authentication Methods," in *Subject-Oriented Business Process Management. Models for Designing Digital Transformations*, M. Elstermann and M. Lederer, Eds., Springer. Cham: Springer Nature Switzerland, 2024, pp. 153–159.

[2] Bundesamt für Sicherheit in der Informationstechnik, "Technische Richtlinie TR-03107: Multi-Faktor-Authentifizierung [Technical Guideline TR-03107: Multi-Factor Authentication]," BSI, Tech. Rep., 2023, [retrieved: September 2025].

[3] I. Goodfellow *et al.*, "Generative adversarial networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, 2014, pp. 2672–2680.

[4] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013, presented at ICLR 2014.

[5] Inc. NeoCortext, "Reface: Face Swap AI Generator," https://apps.apple.com/us/app/reface-face-swap-ai-generator/id1488782587, 2025, [retrieved: September 2025].

[6] I. Perov *et al.*, "DeepFaceLab: A Simple, Flexible and Extensive Face Swapping Framework," https://github.com/iperov/DeepFaceLab, 2020, [retrieved: May 2025].

[7] Entrust Cybersecurity Institute, "2025 identity fraud report," Entrust Corporation, Tech. Rep., 2025, available from https://www.entrust.com, [retrieved: September 2025].

[8] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A large-scale challenging dataset for deepfake forensics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3207–3216.

[9] D. D. Palma, "Face Recognition With Deepfake Detection," https://github.com/domdipa/FaceRecognitionWithDeepFakeDetection, 2025, [retrieved: June 2025].

[10] M. Liao, D. Agnihotri, and X. Zhong, ""paying with my face"–understanding users' adoption and privacy concerns of facial recognition payment," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2022, pp. 731–735.

[11] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

[12] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993.

[13] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[14] Apple Inc., "Informationen zur fortschrittlichen Technologie von Face ID [Information about the advanced technology of Face ID]," https://support.apple.com/de-de/102381, Dec. 2024, [retrieved: September 2025].

[15] google-ai-edge, "MediaPipe Face Mesh," Nov. 2024, [retrieved: September 2025]. [Online]. Available: https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/face_mesh.md#face-landmark-model

[16] C. M. Bishop and H. Bishop, *Deep Learning: Foundations and Concepts*. Cham: Springer International Publishing, 2024.

[17] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, Jun. 2015, pp. 815–823.

[18] Z. Akhtar, "Deepfakes Generation and Detection: A Short Survey," *Journal of Imaging*, vol. 9, no. 1, p. 18, Jan. 2023.

[19] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, Jun. 2020, pp. 8107–8116.

[20] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI, Technical Report, Jun. 2018, preprint. [Online]. Available: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
[21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.
[22] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[23] M. Hasan, E. Rundensteiner, and E. Agu, "Emotex: Detecting Emotions in Twitter Messages," *Academy of Science and Engineering (ASE), USA,© ASE 2014*, 2014.
[24] S. Tariq, S. Jeon, and S. S. Woo, "Am I a Real or Fake Celebrity? Measuring Commercial Face Recognition Web APIs under Deepfake Impersonation Attack," 2021.
[25] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," in *Proceedings of the IEEE Workshop on Information Forensics and Security (WIFS)*, Sep. 2018, presented at WIFS 2018.
[26] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in *International Conference on Computer Vision (ICCV)*, 2019.
[27] H. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-forensics: Using capsule networks to detect forged images and videos," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 2307–2311.
[28] Y. He *et al.*, "ForgeryNet: A Versatile Benchmark for Comprehensive Forgery Analysis," *arXiv preprint arXiv:2103.05630*, 2021.
[29] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *Information Fusion*, vol. 64, pp. 131–148, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253520303110
[30] S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, "Protecting world leaders against deep fakes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
[31] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. Jain, "On the detection of digital face manipulation," 06 2020, pp. 5780–5789.
[32] D. Guera and E. Delp, "Deepfake video detection using recurrent neural networks," 11 2018, pp. 1–6.
[33] I. Hasanov, S. Virtanen, A. Hakkala, and J. Isoaho, "Application of Large Language Models in Cybersecurity: A Systematic Literature Review," *IEEE Access*, vol. 12, pp. 176 751–176 778, Jan. 2024.
[34] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.
[35] M. Alansari, O. A. Hay, S. Javed, A. Shoufan, Y. Zweiri, and N. Werghi, "GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations," *IEEE Access*, vol. 11, pp. 35 429–35 446, 2023.
[36] F. Tassone, L. Maiano, and I. Amerini, "Continuous fake media detection: Adapting deepfake detectors to new generative techniques," *Computer Vision and Image Understanding*, vol. 249, p. 104143, Dec. 2024.