

# Behavior-Based Deepfake Detection: Leveraging Cognitive Response to Visual Face Perception

Hendrik Graupner<sup>1,2</sup> <sup>a</sup>, Mohammad Yeghaneh Abkenar<sup>2,3</sup> <sup>b</sup>, Lisa Schwetlick<sup>4</sup>, Ralf Engbert<sup>5</sup>  
and Christoph Meinel<sup>1</sup>

<sup>1</sup>Hasso Plattner Institute, University of Potsdam, Potsdam, Germany

<sup>2</sup>Innovations Department, Bundesdruckerei GmbH, Berlin, Germany

<sup>3</sup>Department of Linguistics, University of Potsdam, Potsdam, Germany

<sup>4</sup>Psychophysics Lab, Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland

<sup>5</sup>Department of Psychology, University of Potsdam, Potsdam, Germany

**Keywords:** Internet Security, Deepfake Detection, Biometrics, Eye Tracking.

**Abstract:** Face presentation attacks are a propagating issue in an increasingly digitally interconnected world. One of the most recent developments is deepfake impersonation attacks in live video streams. Behavioral biometric analysis is a crucial part of a comprehensive solution to this pressing issue. This paper proposes the application of biological responses to visual self-recognition as a dynamic biometric trait. Self-recognition is a cognitive process that can be leveraged as in-brain identity validation. A sophisticated pre-trained model classifies eye-tracking data to determine the face in the user's current visual focus. One eminent use case is the protection of online video conferences. This paper provides the architecture of a prototypical implementation based on an open-source video conferencing platform. Our work of interdisciplinary research aims to contribute to a holistic solution to protect our modern communication systems and restore trust in digitization.

## 1 INTRODUCTION


Today's society is heavily based on digital systems, ranging from communication to financial services. In this environment, identity theft is a propagating security issue. Sophisticated impersonation, *e.g.*, based on silicone masks, represents a significant challenge (Hernandez-Ortega et al., 2023). More recently, the additional threat of video and audio deepfakes drastically increased the attack potential in online communication systems (Europol, 2023; CNN, 2024). Deepfakes are a face manipulation technique used to change the appearance of a person in a digital video. Without the aid of modern face presentation attack (FPA) detection systems, people can no longer trust their sole perception.


We claim that a comprehensive solution must involve biometrics. Physiological features and human behavior are a rich source of information in a person's video. Including this information in deepfake detection schemes, makes it significantly more difficult to

produce undetectable fakes. Biometric recognition is already an integral part of securing identities in modern computer systems. Our research aims to combine existing knowledge and investigate novel approaches from the field.

This paper utilizes an experimental in-brain identity validation mechanism and aims to employ it as a dynamic biometric trait. The approach based on *visual self-recognition* leverages modern eye-tracking analysis. Due to its success in neuromarketing (Iloka and Anukwe, 2020), eye tracking is an advancing technology of the present day (Grand View Research, 2022). It introduces its application of detecting identity forgery through deepfake attacks in online video-based systems, *i.e.*, video conferencing and video identification.

Section 2 provides a brief overview of related work and concepts on which our approach is based. Former work on the utilization of self-recognition as a biometric trait is summarized in Section 3. Section 4 elaborates on application of the deepfake scenario outlining a detection algorithm and introducing the two use cases of deepfake attacks in video con-

<sup>a</sup>  <https://orcid.org/0000-0002-5305-1624>

<sup>b</sup>  <https://orcid.org/0009-0006-5423-5376>

ferencing and video identification. Implications and limitations are discussed in Section 5 and conclusions are drawn in Section 6.

This paper aims to contribute the first application of visual self-recognition in a real-world scenario and a concrete scheme to detect video deepfakes in real time.

## 2 BACKGROUND

### 2.1 Related Work

Based on a 2022 systematic literature review by Rana *et al.* (Rana *et al.*, 2022), deepfake detection may be grouped into the categories: “deep learning-based techniques, classical machine learning-based methods, statistical techniques, and blockchain-based techniques”. Deep learning techniques, such as convolutional neural networks, outperform classical machine learning and statistics in this task. Hence, there are essentially only two categories of approaches. The authors conclude that deep-learning-based approaches are most common in detecting deepfakes and introduce their work DeepfakeStack (Rana and Sung, 2020) which combines multiple detection algorithms to achieve detection accuracy of up to 99.65%. Machine-learning-based detection does not generalize well beyond the deepfake algorithm it was trained for. Content credentials, *e.g.*, by Hasan *et al.* (Hasan and Salah, 2019), use blockchains to “trace and track the provenance and history of digital content to its original source”.

Recently, researchers have begun to explore the dimension of biometrics as an alternative approach to address the issue from a different perspective. *E.g.*, Agarwal *et al.* (Agarwal *et al.*, 2020) and Cozzolino *et al.* (Cozzolino *et al.*, 2021) make use of deep neural networks (DNNs) that learn biometric features like facial muscle movements. In 2022, a method that covers static analysis of multiple biometric features in the face (*e.g.*, visible blood flow in veins) was published by Intel (Intel Corporation, 2022).

### 2.2 Eye Tracking

The “eye-mind assumption,” proposed by Just and Carpenter (Just and Carpenter, 1980), suggests that the brain processes the information corresponding to the fixation of the eyes. At the same time, interest, mental load, arousal, and implicit processing elicit changes in pupil size, according to Hess and Polt (Hess and Polt, 1960). Therefore, the cognitive reaction to visual stimuli can be measured through

eye tracking to some extent. This work focuses on the time courses of pupil size and microsaccade rates. Microsaccades represent the fastest component of fixational eye movements generated involuntarily during visual fixation (Martinez-Conde *et al.*, 2004).

Eye tracking is an emerging technology as indicated by industry reports, *e.g.* (Grand View Research, 2022). The growing popularity in recent years can be attributed to its success in neuromarketing (Iloka and Anukwe, 2020). We expect that cameras capable of tracking pupils will be present in many end-user devices in the foreseeable future. In 2024, they are already integrated into certain models of smartphones, gaming laptops, and virtual reality (VR) headsets.

### 2.3 Video Deepfakes

Deepfake is a recent face manipulation technique used for changing the appearance of a person in a digital image. It can be used to alter certain attributes, facial expressions, or swap entire faces. The latter is known as identity swap which “[...] consists of replacing the face of one person in a video with the face of another person” (Tolosana *et al.*, 2020).

The same technology used for static deepfake can be used in live streams. In practice, the attacker intercepts their own webcam’s video stream with a live deepfake application, *e.g.*, DeepFaceLive (Github, 2024a). This software uses a pre-trained model to swap the face in real-time. A second software, *e.g.*, OBS Studio (OBS Project, 2024), is required to serve the manipulated video stream as a virtual webcam device to the video conferencing. The financial resources required for this setup are relatively low. Based on our experience, the hardware for acceptable performance can be obtained for around \$3,000.

### 2.4 Video-Ident

Video-Ident is the name of a German video identification method. In Germany and Austria, it is commonly used for remote know-your-customer procedures, *e.g.*, when opening bank accounts. It is implemented by many providers, *e.g.*, POSTIDENT by Deutsche Post (Deutsche Post, 2024).

Video identification builds on a video conferencing application to manually verify a client’s identity. A human agent will request the client to enable their webcam and show their national identity (ID) card. The card’s authenticity is verified by observation of the hologram security features when the card is moved by the holder. The trained agent visually compares the facial features of the ID card and the client to match the claimed and actual identities.

Table 1: Basic cross-validated model performance metrics (including standard deviation) of four different classifiers. The best-performing model (bold) is a pre-trained generic classifier utilizing pupil size and microsaccade rate features.

Model	AUROC	F <sub>1</sub> -Score	EER
Generic, single trials	0.61 ± 0.03	0.59 ± 0.03	0.41 ± 0.03
Generic, aggregated trials (pupil size)	0.74 ± 0.08	0.63 ± 0.06	0.34 ± 0.12
<b>Generic, aggregated trials (pupil size, microsaccade rate)</b>	<b>0.82 ± 0.07</b>	<b>0.76 ± 0.03</b>	<b>0.23 ± 0.07</b>
Individual, single trials	0.62 ± 0.15	–	0.41 ± 0.12

### 3 SELF-RECOGNITION BIOMETRICS

Self-recognition is the cognitive process of recognizing our own face among others. This effect can be measured using eye tracking (see Section 2.2) as shown in a former feasibility analysis (Graupner et al., 2023). Exemplary models were trained on a public data set (Schwetlick et al., 2023a) collected in an empirical study (Schwetlick et al., 2023b) of 116 volunteers in 2022. These models are binary classifiers differentiating between *self* and *non-self-face perception*.

Table 1 shows the results of the four different models that were trained and evaluated in our earlier work (Graupner et al., 2023). The best-performing model (bold) is a generic classifier on aggregated trials using pupil size and microsaccade rate features. In a cross-validated test, it achieves a mean accuracy of 0.82 measured in the area under the receiver operating characteristic curve (AUROC). A more detailed description of the data, feature and model engineering of the underlying machine learning model can be found in our original work.

These results can be leveraged to implement visual self-recognition as a biometric trait. A promising approach is its application as an identity validation mechanism. When a person perceives a face image or video on a screen, the generic classifier can be used to detect the observed face class of self or non-self. This classification result facilitates the decision of whether the face displayed on the screen corresponds to the individual positioned in front of the screen. *I.e.*, the users self-verify their identities to the system. Based on this important piece of information, an identity validation system may detect identity spoofing attacks in certain scenarios.

Other potential applications include continuous authentication based on individually trained models. The system observes the user’s face perception on various occasions and generates long-term identity confidence based on the face recognition results. Due to the high-variance nature of biometric data, fast one-time authentication may not be viable in a real-world scenario.

#### 3.1 Classification Model

The underlying model is a *generic* binary classifier of the classes *self* and *non-self*. It distinguishes between the perception of the observed user’s face and another person’s face. It is trained on a publicly available, high-quality data set on eye-tracking data of face perception (Schwetlick et al., 2023a). The data set is imbalanced by 1 to 10, thus, it was balanced by random undersampling.

The algorithm found to be most effective is *logistic regression* (hyperparameters: solver = lbfgs; max. iter. = 1000; reg. = L2; conv. tol. = 0.0001). The best feature set extracted from the raw eye-tracking data consists of the global maximum of the pupil size, the second local minimum of the microsaccade rate, and the standard deviation of the microsaccade rate. As shown in Table 1 (bold), the model performance is AUROC = 0.82, F<sub>1</sub> = 0.76, and EER = 0.23.

#### 3.2 Application to Real-World Problems

First real-world applications may be aimed at scenarios in which face observation does not require additional effort by the users. This approach makes it relatively uncomplicated to collect data on a larger scale. One scenario that may be apparent is video conferencing. In the course of a video conference participants are often transmitting video streams of their faces. Those real-time face videos are observed by other participants and the transmitting users themselves. Hence, self-recognition can be conducted completely passively without the necessity of additional user interaction.

The technology requires an eye-tracking device to be present at the site of users observed by the system. In our study setup, we use relatively low-cost desktop eye trackers, *e.g.*, the GP3 HD by Gazepoint (Gazepoint, 2024). This sophisticated requirement may currently limit the application of self-recognition technology to special scenarios that require exceptional security. However, in the future, it is very likely that an increasing number of devices already provide built-in functionality to capture eye-tracking data. In 2024, infrared (IR) cameras are already integrated

into certain models of smartphones, gaming laptops, and VR headsets.

Another important aspect is the possibility of expansion to include additional modalities and technologies. Due to the high-variance nature of behavioral biometrics, short-term measurements may only be utilized with a low confidence level. Other solutions may be able to deliver fast results with high confidence but have other limitations. For example, in deepfake detection, a combined approach of deep learning and behavioral biometrics may be more powerful than each of the techniques alone. Deep learning classifiers can produce quick results given video images but are highly limited to the deepfake algorithms they were trained for. Behavioral biometrics need multiple observations and consequently more time, but they do not rely on artifacts of specific deepfake algorithms. A joint solution has the potential to combine the best of both worlds.

### 3.3 User Data and Privacy Considerations

In the scope of this paper, a legally applicable privacy expert opinion is not given. However, we believe that modern security technologies need to consider user privacy already by design.

The approach presented in this paper processes biometric user data which may be used to deduct medical information. Due to the high-sensitive nature of such data, it is of severe importance, that the system is limited to the intended task of self-recognition detection. Consequently, the processes of interest in this solution are local data collection, data transmission to remote services, and remote data processing. Following the paradigm of data minimization, only inevitable data is recorded and transmitted at each step. Under no circumstances, the system processes information about the users (*e.g.* name), to prevent data misuse by giving the opportunity to consolidate medical analysis and other personal information.

#### 3.3.1 Local Data Collection

The biometric data is locally collected by an eye-tracking device. A sophisticated software preprocesses the raw video information recorded by the eye trackers' IR cameras. Hence, the client application only receives the time-series data of *pupil size*, *points of gaze*, and *blink indicators* for the left and right eye, respectively. Based on the gaze data the client application is capable of recognizing the fixation of a user video stream in the video conference. The information gathered is the *time of video stream fixation* and the *binary label of the user (self, non-self)*. Until this

point in the process, no data leaves the local context of the end-user computer.

#### 3.3.2 Data Transmission

The aforementioned time-series data is transmitted to a remote identity validation service that hosts the self-recognition classifier. Data transmission happens immediately after the fixation of a user's video stream plus an observation period of 3,000 ms. Each package transmitted covers the *label of the observed video stream (self, non-self)*, *pupil size*, *points of gaze*, *blink indicators*, and *user session ID*, which is the minimum of required information for classification. Transmission is secured by transport encryption based on the web standard Transport Layer Security (TLS) by default.

#### 3.3.3 Remote Data Processing

A reliable result can only be produced in a controlled environment. Hence, the hosting of the model and inference has to be executed by secure remote service. The aforementioned data transmitted to the service does not contain any user or context information. The labeled eye-tracking data is fed to a generic classifier and the result is interpreted in the context of the security task. Only the *binary result label (suspicious, unsuspecting)* and the *user session ID* are then delivered to an arbitrary monitoring system to display the result or, in a future productive environment, execute subsequent measures (*e.g.*, alerting of operational security personnel). No input or output data is stored by the service.

## 4 DEEPFAKE ATTACK DETECTION BASED ON SELF-RECOGNITION

A recent development of FPAs is image and video deepfakes as introduced in Section 2.3. Our novel approach aims to introduce a sophisticated tool to detect potential face manipulation in video-based systems. Biometric detection offers a way to escape the co-evolutionary dynamic of machine-learning-based deepfake detection by leveraging features and characteristics of humans in addition to digital processing. Especially self-recognition biometrics add a currently unexploited source of information about the relationship between a face in a live video stream and the individuals interacting with it.

The basic principle is to capture a user's involuntary reaction to the video they are streaming, similar to how one might look at themselves in the mirror.



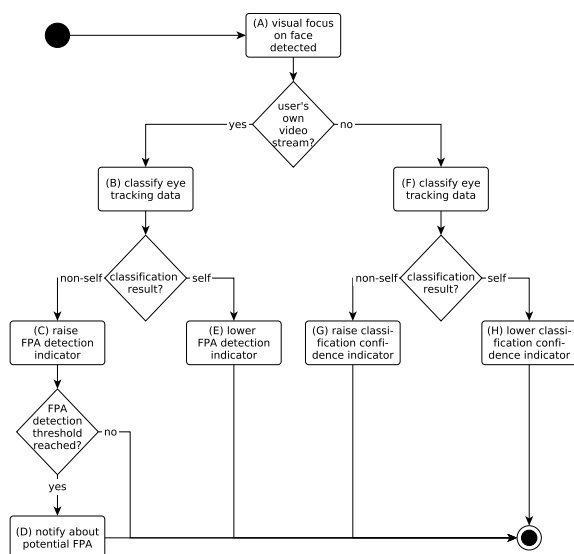


Figure 1: Algorithm used for face presentation attack detection. When a user visually focuses a face in a video stream, the validation system assesses the user’s biometric reaction.

Based on this reaction, the validation system detects a potential discrepancy between the user’s claimed and actual identity. The system classifies the eye-tracking data based on a generic model as described in Section 3.1.

Figure 1 illustrates the algorithm used to detect FPAs in our prototype. Each iteration is triggered when the user visually focuses a face on the screen (A). If the user observes their own webcam’s video stream, confidence in an ongoing FPA can be built by classification of the perceived face (B). In the case of a non-self-face classification, the indicator for FPA detection is raised (C). On reaching a predefined threshold, the system will issue a notification about a potential FPA (D). If the perception is classified as self-face instead, the expected and actual results are equal, hence, the FPA indicator is lowered accordingly (E). If the user views someone else’s face, the system uses the classification (F) result to raise or lower confidence in its classification validity. This mechanism deals with the high natural variance and differences in biological reactions between individuals. If the user perceives the face as non-self, the expected and actual classifications are equal, resulting in increased confidence (G); otherwise, confidence is lowered (H).

This detection algorithm can be leveraged in scenarios where users view their own video streams. Plausible everyday scenarios include online video conferencing and video identification systems. Social engineering attacks backed by impersonation in those contexts yield a high damage potential.

## 4.1 Deepfake Attacks in Video Conferencing

Video conferencing has become an integral part of people’s daily lives. In 2022, 76 % of medium and 93 % of large enterprises conducted online meetings (Eurostat, 2023). This development incentives to take advantage of the vulnerabilities of such systems, and thus, introduce a new attack surface. For example, access from insecure devices is common, and non-authenticated guest users are often allowed. In December 2021, a globally operating gang successfully impersonated the chief executive officer (CEO) of a company to steal an equivalent of \$41 million from French companies by using deepfake technology (Europol, 2023). More recently, a multinational company lost an equivalent of \$25 million in Hong Kong, tricked by the impersonation of their chief financial officer (CFO) in an online video meeting (CNN, 2024).

This deepfake detection approach can be applied to online video conferencing. It aims to protect video conferences with an additional layer of security by detecting FPAs of otherwise unsuspecting users. This method requires users to use an unconcealed device, streaming their face video and eye-tracking data.

The basic attack setup is shown in Figure 2. The attacker intercepts their webcam’s data stream and replaces the original video with a face-swapped video (see Section 2.3). The impersonated user is known to the other conference participants. The conference system is not capable of distinguishing the virtual webcam device from its original source. A remote identity validation service collects the data from local eye-tracking devices and the video conference system, including webcam video streams of all participants. Based on the classification of this data the service validates the users’ identities and potentially notifies another service in case of detection of a potential FPAs according to the algorithm elaborated in Section 3.

In 2023, we implemented a first experimental prototype of this scenario. An initial small data set of twelve volunteers showed promising results. Data collection of a representative data set is ongoing at the time of this publication.

Our prototype extends the open-source video conferencing platform OpenVidu (Github, 2024b). Figure 3 provides an overview of the components of the system. Plug-ins of the client and server component are required to collect information from the video conference, e.g., content in the visual focus of the user. The client application also passes the eye-tracking data to the identity validation service. The latter hosts a pre-trained model and analyzes data during ongoing video calls. Results may be transmitted

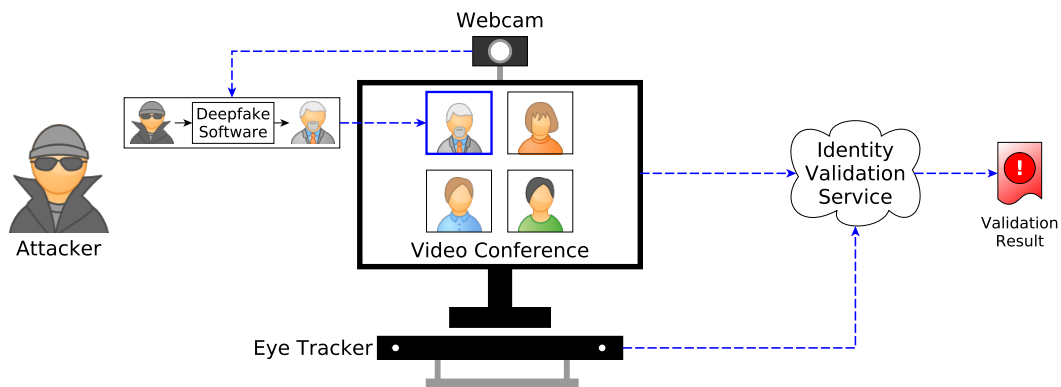


Figure 2: The fundamental attack setup of a deepfake attack in online video conferencing. The attacker intercepts the original video stream of their webcam and presents a virtual device to the video conference, which streams their face-swapped video.

to the server plug-in or an external service. Before inference is possible, training is required and conducted using the standalone component training client. On the client side, the video conferencing application runs in the local browser. Aside from the additional hardware requirement of the eye tracker, the user experience remains seamless.

#### 4.2 Deepfake Attacks in Video-Ident

Like the video conferencing scenario, video identification systems, such as Video-Ident (see Section 2.4), are vulnerable to deepfake-based FPAs. In 2022, the activist group Chaos Computer Club (CCC) hacked six different Video-Ident solutions by presenting a virtual German national ID card even without the use of deepfake technology (Chaos Computer Club, 2022).

Figure 4 illustrates how an attack on Video-Ident works. A client joins an unauthenticated session with a human agent. Authentication is manually performed by matching ID card identity to the face of the person in the video stream. The attacker manipulates their webcam’s raw data like the conference scenario but in addition adds a swap of a real ID card with a fake ID card matching his face swapped identity.

This procedure can not only impersonate an existing person but also invent a non-existing individual. The additional requirement of the client to transmit their eye-tracking data allows the remote identity validation service to verify the client’s perception of the faces displayed on their screen. A potential FPA can be detected, and the agent may be alerted in real time during the live session.

### 5 DISCUSSION

Section 3 summarizes an earlier published demonstration of the feasibility of self-recognition as a biometric trait. Leveraging those concepts, in Section 4, we introduced an FPA detection scheme based on dynamic biometric analysis, which detects deepfake attacks. This method may be a new building block of a multimodal solution to protect users from identity theft in online video-based applications.

A multimodal solution includes state-of-the-art deep learning techniques, static and dynamic biometric analysis, and content credentials. Existing literature (see Section 2.1) shows that the widespread deepfake detection methods based on deep learning are specifically trained for known deepfake algorithms and can achieve very high accuracies on them. On the other hand, biometric approaches like ours are technology-agnostic but involve natural variance. In practice, both can complement each other, and thus, should be combined. A holistic solution can be further enhanced by including static biometric analysis. Content credentials provide additional protection of integrity and authenticity on static content, *e.g.*, images, created by trustworthy devices.

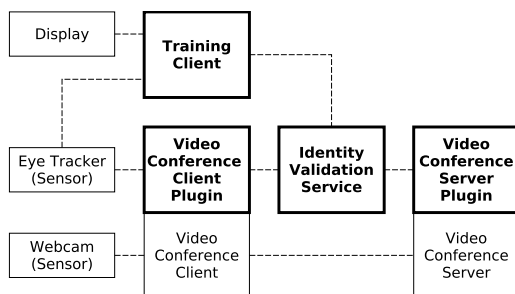


Figure 3: The conferencing system is a client-server architecture communicating with an identity validation service. Four software components (bold) were implemented to add self-recognition-based FPA detection.

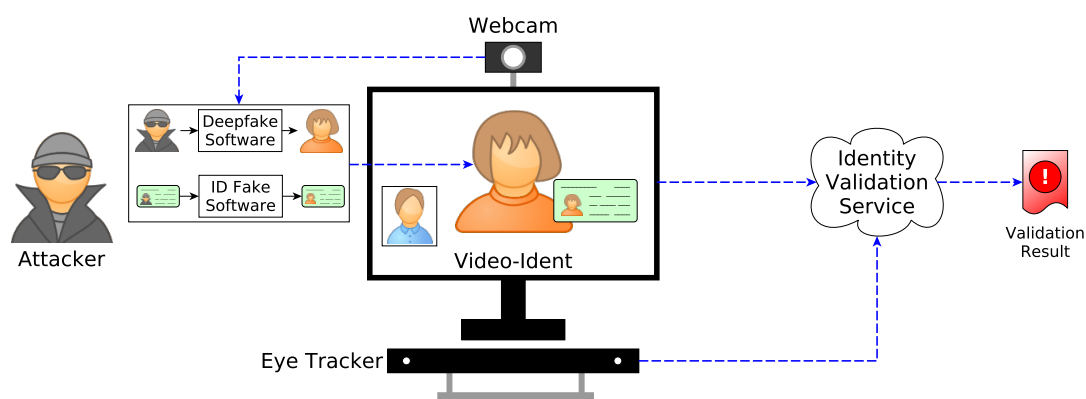


Figure 4: The fundamental attack setup of a deepfake attack in Video-Ident. The attacker intercepts the original video stream of their webcam and presents a virtual device to the Video-Ident system, including a swap of their face and ID card.

### 5.1 Limitations

Eye-tracking data has a high natural variance and significant differences in biological reactions between individuals exist. In addition, the method is limited to eye-trackable individuals who show no significant anomaly in their eye-related reaction to self-perception. This may exclude blind people and those with certain cognitive impairments. To mitigate those limitations, our goal is to combine heterogeneous information sources into a multimodal solution.

Evasion of self-recognition-based deepfake detection may be possible. *E.g.*, attackers could overlay their local video streams with the original webcam, and hence, be undetectable by our identity validation service. This could be prevented by adding visual artifacts to the video stream on the server side, that cause detectable reactions in eye movement. This and other potential bypasses are subject to ongoing research.

### 5.2 Ethical Consideration

It has been shown that classification pipelines perform differently across gender, race, skin color, and other characteristics (Buolamwini and Gebru, 2018). That is why we recognize the importance of ethical considerations and debiasing in our technology.

To mitigate these concerns, we strive to implement more inclusive data collection procedures. By ensuring diverse representation in our data, we aim to address biases related to race, gender, and other identities. This involves carefully tracing how these identities are understood, represented and reflected within the systems and their application contexts.

Additionally, we acknowledge that pre-release trials, independent auditing, and ongoing monitoring are essential to test for bias, discrimination, and other harms in such systems.

## 6 CONCLUSION

This paper describes our approach to using dynamic biometrics for live FPA detection during the online video-based use case of video conferencing (see Section 4.1). We illustrate this exemplarily using deepfake attacks. Moreover, we provide insights into our video conferencing prototype implementation.

Future work to enhance self-recognition-based identification includes the collection of more high-quality data containing deepfake stimuli, *e.g.*, face-swapped or morphed self-face images. Model performance may be improved by investigating more sophisticated features, especially based on microsaccades. Furthermore, the exploration of more dynamic biometric features may be an interesting field for developing smarter FPA detection frameworks. We encourage the community to apply our approach and similar research to upcoming use cases. For example, verification of live avatars in VR scenarios may be interesting. Mobile use cases, *e.g.*, smartphone continuous authentication, are plausible due to the existing availability of eye tracking in many devices.

We aim to build a comprehensive FPA detection framework that includes not only dynamic and static biometric features but also deep-learning-based and content-credential-based protection.

The detection and prevention of online FPAs strengthens the linkage between physical and online identities. It prevents identity fraud and its personal and societal harm. The increasing number of successful frauds exploiting deepfake technology undermines the trust in digital systems such as video conferencing and video identification. This trust must be restored by appropriate countermeasures to not endanger the benefits of today’s technological advances and the potential of tomorrow’s digitization.

## ACKNOWLEDGEMENTS

This research was partially supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under the funded project SENSIBLE-KI (grant number 01MT21005B).

## DISCLOSURE OF INTERESTS

This research was funded by Bundesdruckerei GmbH. At the time of publication, authors H. Graupner and M. Y. Abkenar are employed by Bundesdruckerei.

## REFERENCES

- Agarwal, S., Farid, H., El-Gaaly, T., and Lim, S.-N. (2020). Detecting deep-fake videos from appearance and behavior. In *2020 IEEE international workshop on information forensics and security*, pages 1–6. IEEE.
- Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR.
- Chaos Computer Club (2022). Chaos Computer Club hacks Video-Ident. <https://www.ccc.de/en/updates/2022/chaos-computer-club-hackt-video-ident>. last accessed 2024/06/27.
- CNN (2024). Finance worker pays out \$25 million after video call with deepfake 'chief financial officer'. <https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html>. last accessed 2024/06/27.
- Cozzolino, D., Rössler, A., Thies, J., Nießner, M., and Verdoliva, L. (2021). Id-reveal: Identity-aware deepfake video detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15108–15117.
- Deutsche Post (2024). How does the identity check by video chat work? <https://www.deutschepost.de/en/p/postident/privatkunden/identifikation-per-videochat.html>. last accessed 2024/06/27.
- Europol (2023). Franco-israeli gang behind eur 38 million ceo fraud busted. <https://www.europol.europa.eu/media-press/newsroom/news/franco-israeli-gang-behind-eur-38-million-ceo-fraud-busted>. last accessed 2024/06/27.
- Eurostat (2023). Online meetings and remote access to enterprise resources - statistics. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Online\\_meetings\\_and\\_remote\\_access\\_to\\_enterprise\\_resources\\_-\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Online_meetings_and_remote_access_to_enterprise_resources_-_statistics). last accessed 2024/06/27.
- Gazept (2024). GP3 HD Eye Tracker 150Hz. <https://www.gazept.com/product/gp3hd>. last accessed 2024/06/27.
- Github (2024a). Deepfacelive. <https://github.com/iperov/DeepFaceLive>. last accessed 2024/06/27.
- Github (2024b). OpenVidu. <https://github.com/OpenVidu/openvidu>. last accessed 2024/06/27.
- Grand View Research (2022). Eye tracking market size, share & trends analysis report. <https://www.grandviewresearch.com/industry-analysis/eye-tracking-market>. last accessed 2024/06/27.
- Graupner, H., Schwetlick, L., Engbert, R., and Meinel, C. (2023). Unconventional biometrics: Exploring the feasibility of a cognitive trait based on visual self-recognition. In *2023 IEEE International Joint Conference on Biometrics*, pages 1–10. IEEE.
- Hasan, H. R. and Salah, K. (2019). Combating deepfake videos using blockchain and smart contracts. *IEEE Access*, 7:41596–41606.
- Hernandez-Ortega, J., Fierrez, J., Morales, A., and Galbally, J. (2023). Introduction to presentation attack detection in face biometrics and recent advances. *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pages 203–230.
- Hess, E. H. and Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, 132(3423):349–350.
- Iloka, B. C. and Anukwe, G. I. (2020). Review of eye-tracking: A neuromarketing technique. *Neuroscience Research Notes*, 3(4):29–34.
- Intel Corporation (2022). Intel introduces real-time deepfake detector. <https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html>. last accessed 2024/06/27.
- Just, M. A. and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*, 87(4):329.
- Martinez-Conde, S., Macknik, S. L., and Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature reviews neuroscience*, 5(3):229–240.
- OBS Project (2024). Obs studio. <https://obsproject.com>. last accessed 2024/06/27.
- Rana, M. S., Nobi, M. N., Murali, B., and Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE access*, 10:25494–25513.
- Rana, M. S. and Sung, A. H. (2020). Deepfakestack: A deep ensemble-based learning technique for deepfake detection. In *2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, pages 70–75. IEEE.
- Schwetlick, L., Engbert, R., and Graupner, H. (2023a). Data Set: Face- and Self-Recognition Effects on Pupil Size and Microsaccade Rate. *Open Science Framework*.
- Schwetlick, L., Graupner, H., Dimigen, O., and Engbert, R. (2023b). Self-recognition generates characteristic responses in pupil dynamics and microsaccade rate. *arXiv*.
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., and Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64:131–148.