Chapter 9
Face Morphing Attack Detection: Privacy and Digital Sovereignty Issues, Methods and Future Direction

Sulaimon Adebayo Bashir
https://orcid.org/0000-0001-8690-3953
Federal University of Technology, Minna, Nigeria

Muhammad Bashir Abdullahi
https://orcid.org/0000-0001-8200-2787
Federal University of Technology Minna, Nigeria

Mary Ogbuka Kenneth
Imperial College London, UK

Abubakar Dahiru
National Information Technology Development Agency (NITDA), Nigeria

Olawale Surajudeen Adebayo
https://orcid.org/0000-0001-7052-4756
National Open University of Nigeria, Jabi, Nigeria

Kehinde H. Lawal
Federal University of Technology Minna, Nigeria

ABSTRACT

This chapter examines the privacy threat facing face recognition and morphing detection systems using the LINDUUN framework as the methodology. It describes various aspects of privacy issues that may hamper safe and secured operation of face recognition and morphing attack detection systems. Digital sovereignty issues affecting the systems are also examined with recommendations on actions to be taken to safeguard privacy and digital sovereignty of the systems. A new taxonomy of the approaches being used for the detection of face morphing attacks is presented. The privacy and digital sovereignty concerns related to the detection methods are elicited with recommendations to alleviate them. The findings reflect that critical

DOI: 10.4018/979-8-3693-9137-2.ch009

privacy and digital sovereignty issues are present in the domain of face recognition and morphing detection systems. Therefore, these expositions will influence policy-makers and researchers to address the issues and also utilized the recommendations for further research.

INTRODUCTION

Face recognition (FR) systems, widely used in various identity verification applications like e-passports and online authentication; they are designed to enhance security by relying on unique facial features of individual. However, these systems are not immune to sophisticated attacks. One of such attacks is the face morphing attack. Face morphing attack involves blending the facial features of two or more individuals to create a single image that can trick a face recognition system into recognizing each contributing person. Figure 1 shows the blending of two faces (Identity A and Identity B) using four different morphing tools- OpenCV scripts, Facemorpher, StyleGAN3 anf MIPGAN-II (Sarkar et al., 2022) to produce four different blended images as shown in Figure 2. Each of the persons that contributed the identity images can use for authentication under a face recognition system. This face morphing attack has serious privacy and security implications around the data being collected, the machine learning models being used for the detection of the attack and also issues of sovereignty of the data being collected.

Figure 1. Two different identity images being blended

Figure 2. Different Morphed Images Generated from 2 identity images from Figure 1a (Face Research Lab London (DeBruine and Jones, 2021) Dataset by Sarkar et al. (2022))

(b) OpenCV (c) FaceMorpher (d) StyleGAN2 (e) MIPGAN-II

In most cases, this attack is prevalent under face recognition system being used for identity verification such e-passport. If an adversary is able to infiltrate a facial biometric recognition system with morphed image during the issuance process such passport can be authenticated successfully for any of the contributing subjects during verification at a border control leading to grave security breaches such as identity theft, fraud, and unauthorized access to personal data.

While the end results of face morphing attack can lead to serious security breach, it also elicits a lot of privacy issues in terms of the data being collected by the facial recognition systems and the models being used under the system for detecting the attack. Therefore, privacy concerns have become worrying due to the scale of personal information being collected and stored by these systems. The information been collected under facial recognition is increasing significantly nowadays, motivated by law enforcement, national security, and economic incentives. The issue of privacy is a common issue affecting many information processing domains such as biometric systems (Melzi et al., 2024; Wang et al., 2024), Internet of Things (Badr et al., 2021), healthcare (Jawad, 2024; López-Martínez et al., 2023), social media networks (Bhattacharya et al., 2023) and information systems (Bu et al., 2020) in general.

The objectives of this chapter are to:

i. examine the digital sovereignty and privacy issues related to face recognition and morphing detection systems.
ii. proffer recommendations on safeguards to resolve the issues of digital sovereignty and privacy issues related to face recognition and morphing detection systems.
iii. analyze the various machine learning approaches being used for the detection of face morphing attacks.
iv. examine the privacy threats that arises from the machine learning models.

Addressing the above objectives provides the following contributions:

i. Taxonomy of face morphing detection approaches is presented.
ii. Elicitation of key digital sovereignty and privacy issues in face recognition and morphing detection systems.
iii. Recommendation on the safeguards for privacy and digital sovereignty in face recognition and morphing detection systems.

The rest of this chapter is organized as follows: section 2 examines privacy and digital sovereignty issues in face recognition and morphing attack detection systems. Section 3 discusses privacy and digital sovereignty issues in face morphing machine learning detection models. Section 4 presents effects of privacy breach in face recognition system under face morphing attacks while section 5 presents the taxonomy of face morphing attack detection methods while section 6 examines the challenges in detecting face morphing attacks. Section 7 suggests future direction of research and policy considerations while section 8 and 9 present the limitations and conclusion respectively.

PRIVACY AND DIGITAL SOVEREIGNTY ISSUES IN FACE RECOGNITION AND MORPHING ATTACK DETECTION SYSTEMS

The concept of privacy has different conational meanings depending on social and cultural issues, study disciplines, stakeholder interests, and application perspective (Deng et al., 2011). However, the general overriding notion of privacy across domains is that, privacy is the fundamental right of individuals to control access to their personal information, autonomy over their bodies and decisions, and freedom from undue surveillance or intrusion. Thus, a key definition of privacy in the context of information technology is given as “the right to be let alone”, focusing on freedom from intrusion, and “the right to informational self-determination”, allowing individuals to “control, edit, manage, and delete information about themselves and decide when, how and to what extent that information is communicated to others” (Hansen, 2008). Another definition of privacy given by Clarke (2018) as “Privacy is the interest that individuals have in sustaining 'personal space', free from interference by other people and organizations.”

Digital Sovereignty on the other hand can be described as the pursuit of processes to protect and safeguard all forms of digital resources and infrastructures including-data, networks, machine learning models of an organization or sovereign nation from external control while minimizing reliance on foreign technologies for the

safe and secure operation of these resources (Dabrock et al., 2021; Jiang and Belli, 2025; Tan et al., 2023). The concept of digital sovereignty has gained significant research attention in recent times because of its potential to support privacy and data protection, enhance national security and promote economic independence through the localized management of digital technology infrastructure. As highlighted by Misra et al. (2025), digital sovereignty is critical in the context of Industry 5.0, enabling sovereign control over essential digital elements such as data, infrastructure, algorithms, and visualization tools. Katsikas (2025) underscores its role in bolstering cybersecurity, while Klare et al. (2025) identify it as a key element to seamless software development process.

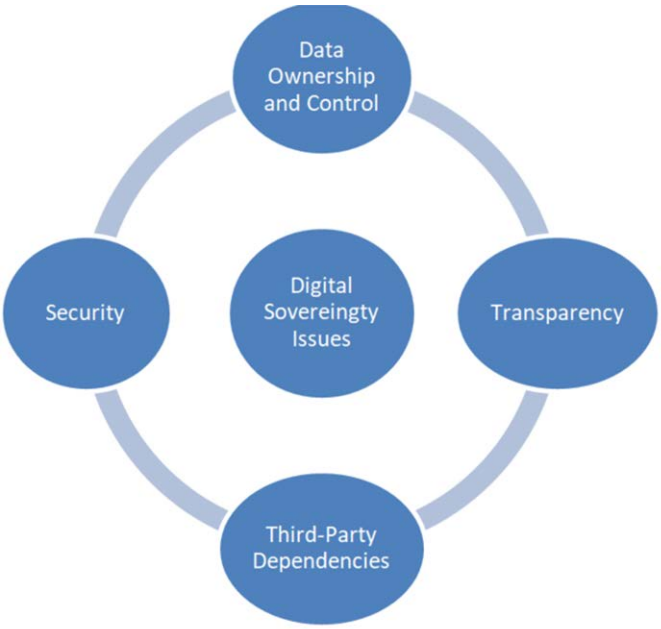
Facial recognition systems, in particular process highly sensitive biometric data, often involving both domestic citizens and foreign nationals undergoing verification—such as at automated border control e-gates. These systems require robust and adequate safeguard to ensure sovereignty of the collected data, the intelligent algorithm employed in the process and the underlying hardware and other supporting technologies.

The rest of this section explores and examines the data sovereignty and privacy issues inherent in the operational processes of face recognition and morphing attack detection systems and diverse approaches to protect and uphold the sovereignty and privacy of such systems' artifacts.

Digital Sovereignty Issues in Face Recognition and Morphing Attack Detection Systems

Face recognition and morphing attack detection systems used for biometric security, present several significant challenges to digital sovereignty—especially in developing countries. There are many issues that have been identified by different researchers that underpin the actualization of digital sovereignty. Misra et al. (2025) identified 8 major aspects of digital sovereignty from the literature in the context modern digital industries. These issues span across data ownership, dependency on external providers, legal and regulation, resilience, goverance, ethics, technology and cybersecurity. However, in the perspective of face recognition and morphing attack detection system, we opine that the key issues in this domain include: data ownership and control, transparency, third-party dependencies, and vulnerability to cyberattacks as depicted in Figure 3.

Figure 3. Digital sovereignty issues in face recognition and morphing detection systems



1. Data Ownership and Control: Facial images and biometric information are inherently sensitive and personal. However, many systems store such data in centralized or cloud environments, often hosted in foreign jurisdictions. This raises serious concerns about who truly owns or controls the data. Without clear data governance, nations risk losing sovereignty over critical personal and national digital assets. Cross-Border Data Flows

The storage and processing of biometric data on servers located in other countries make such data subject to foreign laws. This compromises the ability of national governments to enforce their own data protection laws and exposes citizen data to foreign surveillance and potential misuse. Exacerbating this issue is occasioned by of lack of local capacity for technological infrastructure and skilled manpower in some developing nations that are required to independently develop biometric systems. This gap results in dependence on foreign technologies and limits the country's ability to safeguard its digital autonomy.

2. Transparency: Most morphing detection systems are developed using deep learning models that are typically black-box in nature. The lack of interpretabil-

ity and explainability makes it difficult to understand or audit how decisions are made, thereby weakening accountability and trust, especially in national security or legal contexts. Another aspect of transparency is model bias and discrimination whereby AI models used for facial recognition are trained on non-representative datasets. As a result, these models may perform poorly on African or other underrepresented demographics, leading to errors in identification and potential discrimination. This undermines the sovereignty and fairness of national identity systems.

3. Third-Party Dependencies: The use of third-party tools—including pre-trained models, cloud APIs, or proprietary libraries—introduces external dependencies. These technologies may be governed or maintained by foreign companies, limiting the deploying nation's control and increasing the risk of backdoors, data exfiltration, or sudden service discontinuation.
4. Security: Morphing attacks—where facial images are digitally manipulated to spoof biometric systems—pose a direct threat to face recognition reliability. If detection mechanisms are not sufficiently robust, systems can be easily bypassed, leading to identity theft, fraud, and national-level privacy breaches.

Safeguards for Digital Sovereignty in Face Recognition Systems

Addressing these digital sovereignty issues requires deliberate policy frameworks, technical innovation, and capacity-building initiatives that prioritize national interests.

1. Development of Indigenous Technologies: Governments and institutions should encourage and fund the development of local face recognition and morphing detection solutions. Building these systems on locally curated datasets will help reduce bias and improve performance for native populations while ensuring full control over their functionality and updates.
2. Data Localization and Sovereign Cloud Infrastructures: Biometric and facial data should be stored within national borders under strict data localization policies. Investments should be made in establishing secure, compliant data centres that meet international standards, enabling countries to retain legal and operational control over their digital assets.
3. Auditability and Explainability in AI: AI systems used for biometric verification should include explainability features that allow independent regulators to audit decision-making processes. This is particularly important for legal and high-security applications where transparency is non-negotiable.
4. Strengthening Legal and Regulatory Frameworks: Countries should update or develop comprehensive data protection, cybersecurity, and digital rights legis-

lation that prioritizes digital sovereignty. Regulatory bodies must be empowered to enforce compliance, especially when foreign technologies are used in critical national infrastructure.

5. Cybersecurity and Adversarial Robustness: Ensuring the resilience of face recognition systems against morphing and other adversarial attacks is vital. Techniques like morphing-resistant model training, anomaly detection, and multi-modal authentication can strengthen system security and trustworthiness.
6. Reduction of Third-Party Reliance: Where possible, governments should adopt open-source, auditable tools or work with trusted local partners. Strict vetting procedures should be implemented for any foreign technology integrated into national systems, including contractual agreements on data handling and software updates.
7. Capacity Building and Digital Literacy

To sustain long-term digital sovereignty, there must be strategic investments in local expertise. This includes training AI specialists, cybersecurity professionals, and system auditors who can design, manage, and secure these technologies. Public awareness campaigns are also essential to educate citizens on data privacy and digital rights.

Privacy Issues in Face Recognition System

In this section, we examine privacy issues and threats that are inherent facial recognition systems. Our analysis is conducted using the **LINDDUN framework** (Deng et al., 2011; Wuyts et al., 2014), a structured approach for identifying and mitigating privacy threats. The **LINDDUN framework** categorizes threats into specific types, and maps them to system elements to address vulnerabilities effectively. By applying this framework to **face recognition systems**, it help us to identify privacy concerns at various stages, from data collection to processing and sharing. This framework has also been applied in the work of Beltrán and Calvo (2023) to develop a privacy threat model for facial recognition systems. Given the intertwined and interdependent nature of facial recognition and morphing detection systems, privacy concerns in one system inherently influence the other. This interconnection highlights that privacy issues in these systems are closely related and cannot be addressed in isolation.

Privacy issues in face recognition system can be understood as the specific concerns or challenges that arise due to the design, implementation, operation, or usage of face recognition system which may potentially lead to violations of individuals' privacy rights. These issues often stem from the way data is collected, processed, stored, shared, or used and can affect individuals' autonomy, security, and trust in the

system. Privacy issues are conceptual and systemic problems that, if not addressed, can result in privacy threats or breaches. They highlight areas where personal data or activities may be exposed to unauthorized access, misuse, or exploitation, often due to gaps in system design, lack of regulations, or insufficient user awareness. Some of the privacy issues in face recognition systems include: inadequate user consent, over-collection of data, improper data sharing, insufficient data protection and lack of user control.

1. **Inadequate User Consent** is one of the key issues in FR systems is the lack of adequate mechanisms for obtaining user consent regarding the collection, processing, and storage of facial data especially in the public surveillance FR systems. In many cases, individuals are unaware that their facial data is being captured or how it is subsequently used. For instance, public surveillance cameras equipped with facial recognition capabilities often operate without providing clear notification or obtaining explicit consent from individuals being monitored. Additionally, in many implementations, consent mechanisms are either entirely absent or poorly designed, leaving users uninformed about the extent and purpose of the data collection. This lack of transparency means individuals may not understand how their facial biometric information is processed, stored, or shared, including whether it will be used for purposes beyond the original intent, such as surveillance, profiling, or commercial activities. This issue raises both ethical and legal concerns, particularly as facial biometric data is highly sensitive and irrevocable if misused or compromised. Effective consent mechanisms should not only notify individuals but also provide meaningful choices about their participation in such systems which is mostly not the case.
2. **Over-Collection of Data**: Another pertinent privacy issue in FR system is over-collection of data. Many FR systems collect and store far more information than is necessary for their intended purpose, often due to design choices, operational convenience, or a lack of stringent data minimization practices. For instance, instead of capturing and processing only the specific facial features required for identification or verification, these systems may record and store entire video feeds, including unnecessary background details and bystanders' faces, which are irrelevant to the system's function. This over-collection not only increases the risk of unauthorized access and data breaches but also raises ethical concerns about the unnecessary surveillance of individuals who may not even be the target of the FR system. Furthermore, such practices can inadvertently capture sensitive information about individuals' behaviors, locations, and interactions, potentially creating extensive personal profiles. For example, in public surveillance scenarios, FR cameras may continuously record and store footage of all passersby, even if the system is only designed to track specific individuals for

security purposes. Similarly, in commercial settings like retail stores, FR systems might gather data on every customer's movements and expressions, exceeding the requirements for theft prevention or customer service. This indiscriminate data collection exacerbates privacy risks, as the more extensive the data repository, the greater the potential for misuse, unauthorized sharing, or exploitation by malicious actors.

3. **Improper Sharing of Data**: This critical privacy issue in facial recognition (FR) systems occurs when facial data collected for specific purposes, such as unlocking a device, verifying identity, or enhancing security are shared with third parties without the user's explicit consent or awareness. For instance, biometric data gathered by FR systems in smartphones or apps may be shared with advertisers to create detailed profiles for targeted marketing campaigns, often without providing users with clear information about such practices. In other scenarios, facial data captured by public or private surveillance systems, initially intended for security purposes, could be handed over to law enforcement agencies or private organizations for unrelated investigations or other activities, frequently bypassing proper legal oversight and neglecting to inform the affected individuals. For example, a shopping mall using FR technology to enhance customer experience may also share collected facial data with third-party advertisers to analyze consumer behavior or preferences. Similarly, FR-enabled security cameras installed in urban areas might provide captured facial images to government agencies for unrelated purposes, such as monitoring protests or identifying individuals of interest, raising concerns about potential overreach and the erosion of civil liberties. These practices raise serious ethical and legal concerns, particularly because facial biometric data is highly sensitive and immutable. Facial data, poses unique risks due to its highly sensitive and immutable nature. Facial data is deeply personal and directly tied to an individual's identity, making its mishandling or misuse a significant privacy concern. Unlike passwords or other identifiers, which can be reset or changed if compromised, facial data is permanent—an individual cannot alter their facial features to protect themselves if their data is leaked or improperly used. This makes biometric data a one-time, high-stakes identifier, increasing the potential for long-term harm if it is exposed or shared without consent. The misuse or improper sharing of facial data can lead to severe consequences such as stolen facial data exploited for identity theft, enabling criminals to impersonate individuals in systems that rely on facial recognition. This loss of control over one's data undermines personal autonomy, eroding trust in technology and the organizations that use it.
4. **Insufficient Data Protection**: This privacy is a very significant issue in safeguarding the privacy of individuals in facial recognition systems. When security

measures are weak or encryption protocols are inadequate, sensitive biometric data, such as facial images and associated metadata, becomes vulnerable to breaches or unauthorized access. This lack of robust protection can create opportunities for exploitation by malicious actors, including hackers and insiders with ill intent. For instance, databases storing facial data may be targeted by cybercriminals aiming to steal or misuse the information for purposes such as identity theft, fraud, or even blackmail. Similarly, without proper access controls and monitoring mechanisms, insiders with privileged access could misuse the data for unauthorized purposes, such as selling it to third parties or using it for personal gain. These scenarios highlight the critical role of comprehensive security frameworks in preventing data breaches. Additionally, inadequate data protection measures can undermine user trust in facial recognition systems, as individuals may fear their sensitive information is not being handled responsibly. This is particularly concerning because biometric data is immutable—if compromised, it cannot simply be changed like a password. The consequences of such breaches can be far-reaching, leading to long-term harm for individuals and legal or reputational risks for organizations.

5. **Lack of User Control** A significant privacy issue in facial recognition (FR) systems is the lack of user control over their facial biometric data once it has been captured. Users often have minimal or no options to access, modify, or delete their data stored in such systems. This is especially problematic in the context of public FR systems, where individuals have no control over when, where, or how their facial data is collected, often under the justification of security or law enforcement purposes. For example, surveillance cameras equipped with FR capabilities in public spaces can capture facial images without notifying individuals or seeking their consent. In such scenarios, users have no ability to opt out of the data collection process, nor are they informed about how their data will be used, shared, or retained. This lack of transparency and control creates a power imbalance, leaving individuals vulnerable to privacy violations, misuse of their data, or unauthorized surveillance. Additionally, the inability to manage one's biometric data contradicts fundamental privacy principles, such as the right to be forgotten or data minimization. Without mechanisms to allow users to remove or restrict access to their data, FR systems can perpetuate ongoing risks, including potential misuse by third parties or the government. This concern becomes even more pressing in jurisdictions without robust privacy regulations, where there may be few safeguards against the misuse or overreach of such technologies.

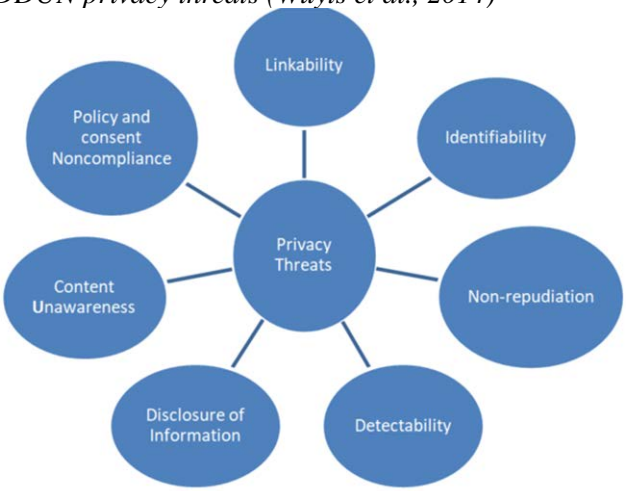
To address all these identified and discussed privacy issues in FR systems, it is essential for organizations to implement robust and user-centric data governance

frameworks while adhering to regulations that govern data privacy within their jurisdictions of operations such as the General Data Protection Regulation (GDPR) in Europe, the California Consumer Privacy Act (CCPA) in America. Transparent consent mechanisms are essential to inform users about how their facial data will be collected, used, shared, and stored, ensuring individuals can opt in knowingly. Systems should adhere to data minimization principles, collecting only the necessary information, such as extracted facial features, rather than retaining full video feeds or unnecessary metadata. Strict policies must regulate data sharing to prevent biometric data collected for one purpose from being repurposed or shared with third parties, such as advertisers or law enforcement, without explicit user consent. To safeguard sensitive information, organizations must employ strong encryption, secure storage protocols, and multi-layered access controls, complemented by regular security audits to prevent unauthorized access or breaches. Furthermore, users must be empowered with tools to access, modify, or delete their biometric data, ensuring they maintain oversight and control. By combining these technical safeguards, compliance with international privacy laws, and organizational accountability, FR systems can be deployed in a manner that upholds privacy, fosters trust, and protects user autonomy.

Privacy Threat Types in Face Recognition Systems

According to Wuyts et al. (2014) the privacy threats in any information processing system can be captured through the LINDDUN framework. Each letter in "LINDDUN" represents a specific privacy threat type, derived by negating a corresponding privacy property that should ideally be upheld in any information processing system. The acronym highlights key areas where privacy can be compromised if these properties are not adequately protected(Deng et al., 2011). The privacy threats outlined in the framework are applied here to identify specific privacy risks associated with facial recognition (FR) systems. These threats are shown in Figure 4 while Table 1 shows the privacy properties and the antagonizing privacy threats.

Figure 4. LINDDUN privacy threats (Wuyts et al., 2014)



Linkability

This is a privacy threat that directly contradicts the desirable privacy property known as unlinkability. Unlinkability, as defined by Pfitzmann and Hansen (2010), refers to the inability of an attacker to determine whether two or more items of interest (IOIs)—such as subjects, messages, actions, or other entities—are related within a system. Specifically, unlinkability ensures that within the system, the attacker cannot sufficiently distinguish whether these IOIs are associated with one another. From this definition, it is clear that unlinkability is a crucial privacy property in face recognition systems, as it prevents attackers from connecting an individual's facial biometric data across multiple contexts or databases. Linkability poses a significant threat in facial recognition systems because it enables the association of facial biometric data with other sources of information. This could involve linking an individual's face data captured in a public setting to private records, such as online profiles, medical histories, or travel itineraries. Such linkages can lead to extensive profiling, surveillance, and potential misuse of personal data, undermining user privacy and autonomy.

Furthermore, this threat increases privacy risks by increasing the potential for unauthorized tracking. For example, in commercial contexts, companies could exploit Linkability to combine shopping behaviors, location data, and facial recognition inputs to create detailed consumer profiles, often without the user's explicit consent. In another scenario in pubic FR, an attacker can cross-referencing facial recognition

data collected at an airport with images from social media or surveillance footage to build a comprehensive profile of the individual.

Identifiability

This is another critical privacy threat that contradicts the desirable privacy properties of anonymity and pseudonymity. According to Pfitzmann and Hansen (2010), anonymity refers to a condition where, from an attacker's perspective, a subject cannot be sufficiently identified within a group of subjects, commonly known as the anonymity set. In the same vein, pseudonymity is said to involve the use of identifiers other than a subject's real names. A subject is considered pseudonymous when an identifier, such as a pseudonym, is used instead of their actual name. These properties aim to ensure that personal identity is obscured or replaced by alternative identifiers to prevent unauthorized recognition.

In the context of facial recognition (FR) systems, identifiability threats encompass the risk of revealing or inferring an individual's identity through their facial biometric data. Unlike anonymity, where individuals remain indistinguishable within a larger group, and pseudonymity, where alternative identifiers mask real identities, FR systems inherently process highly identifiable information, such as facial features. This makes them particularly vulnerable to identifiability threats when data is misused or improperly secured.

For example, identifiability threats in FR systems may occur in public surveillance, where individuals can be uniquely identified in real time or retrospectively across various locations. Such identifications can be linked with other datasets, such as social media profiles or government records, enabling unauthorized profiling and tracking. In retail environments, FR systems might identify individuals to personalize marketing efforts without their consent, further diminishing their anonymity. Similarly, in law enforcement, false positives in identification could lead to unwarranted scrutiny or legal consequences for individuals. Pseudonymity-related threats also arise when the protective pseudonyms assigned to individuals during processing are compromised. Data breaches or linkage attacks could expose the real identities behind these pseudonyms, undermining the privacy protections they were meant to provide. This creates risks such as discrimination, stalking, or exploitation.

Non-Repudiation

This threat directly contradicts the essential and preferred privacy property of *plausible deniability*. Plausible deniability is a privacy safeguard that ensures an attacker cannot prove that a user possesses certain knowledge, has taken a specific action, or has communicated particular information. While non-repudiation is a

key security requirement in many systems—ensuring accountability by providing irrefutable evidence of an action—it was observed that plausible deniability can sometimes take precedence over non-repudiation, depending on the application context (Deng et al., 2011).

For example, in systems designed to protect whistleblowers, users may need the ability to deny sending a message to ensure their safety and maintain anonymity. Similarly, plausible deniability is crucial in scenarios like private, off-the-record conversations, denying the existence of encrypted files, disputing the transmission of files from a specific source, or refuting the association of a database record with an individual (Deng et al., 2011). In such cases, plausible deniability serves as a protective measure against privacy violations and undue scrutiny.

In the context of facial recognition (FR) systems, the threat of non-repudiation arises when the system irrevocably links an individual to specific actions, locations, or behaviors, removing any opportunity for deniability. For example, if an FR system captures an individual's face at a particular event, that person may be unable to dispute their presence, even if being linked to that event poses risks to their safety or reputation. This threat is especially concerning in situations involving protests, sensitive locations, or high-stakes environments, where being identified could result in severe consequences, such as persecution or discrimination. As another scenarios given in (Beltrán and Calvo, 2023), non-repudiation threats in facial recognition (FR) systems can manifest during identity verification processes where the subject provides a facial image. In the first scenario, the subject may genuinely be the source of the data flow but later deny initiating it. For example, if facial images are captured by public cameras or sensors without obfuscation, the event can be visibly recorded, creating a clear link between the subject and the data. Similarly, if the data flow occurs remotely without adequate encryption or concealment, it can be logged in a way that traces the data back to the subject, undermining their ability to plausibly deny involvement. Conversely, in the second scenario non-repudiation threats occurs when an adversary manipulates the system to make it appear as though the subject initiated the data flow when they did not. This can happen through presentation attacks, where an attacker uses a spoofed image or mask to impersonate the subject and successfully pass identity verification. Alternatively, logical impersonation or replay attacks can exploit authentication weaknesses or reuse previously captured biometric data to simulate a legitimate data flow. These threats emphasize the need for robust security measures, including mutual authentication, encryption, and anti-spoofing mechanisms, to protect both the integrity of the system and the privacy of individuals. Without these safeguards, subjects remain vulnerable to irrevocable attribution, whether the data flow originated from them or was falsely attributed to them.

Non-repudiation threat in FR systems requires careful consideration of how facial biometric data is collected, stored, and shared. System designers must implement measures to balance the competing needs for accountability and privacy

Detectability

Detectability refers to the potential for a person's presence or data to be identified or recognized within a system, even when the system is designed to anonymize or exclude individuals. This constitutes a significant threat to privacy, as it directly opposes the corresponding privacy properties of undetectability and unobservability. According to Pfitzmann and Hansen (2010), undetectability is the inability of an adversary to discern whether an item of interest (IOI), such as a subject's facial data, exists within a system. Meanwhile, unobservability combines undetectability with anonymity, ensuring that the IOI remains undetectable to all uninvolved parties while also protecting the anonymity of the associated subjects, even from other involved entities.

Unobservability provides a higher level of privacy assurance by revealing less information than anonymity alone. However, because it is a composite property of undetectability and anonymity, the emphasis in privacy threat frameworks like LINDDUN is placed on undetectability as a standalone property. This threat, therefore, is framed as detectability, highlighting the potential risk of exposing sensitive information or confirming the existence of data.

This threat manifest facial recognition (FR) systems, when the system's intruder is able to identify or confirm the presence of a subject's face in a database, even when anonymization protocols are in place. For instance, an FR system used for law enforcement may inadvertently allow adversaries to detect whether a specific individual is listed in the system, breaching privacy. Similarly, the mere existence of metadata or residual data linked to facial images could enable unauthorized parties to identify subjects indirectly. Detectability threats can occur due to flaws in data encryption, weak system design, or inadequate anonymization techniques, making it a critical challenge for FR systems to address to uphold user privacy.

Disclosure of Information

Discloser of information means exposing personal information to someone who is not eligible to access it. This threat is in direct contradiction to confidentiality privacy property. Confidentiality as a privacy property can be defined as safeguarding of sensitive information to ensure it is accessible only to authorized individuals, protecting it from unauthorized access or disclosure(Woubie et al., 2024). Therefore, confidentiality in FR systems involves facial images or associated metadata, being

exposed to unauthorized parties. For instance, if an FR system's database is not adequately secured, it could be accessed by external attackers, internal malicious actors, or even inadvertently by third parties. Such disclosures could lead to a range of consequences, from identity theft to profiling and surveillance without consent.

A common scenario of disclosure involves the sharing of data between organizations without proper anonymization or encryption. For example, an FR system used for authentication at a corporate facility might share data with third-party vendors for analytics without adequately masking the identities of individuals. Similarly, weak encryption protocols could result in intercepted communications during data transmission, exposing users' sensitive facial data.

The risk of disclosure is further heightened in public FR deployments, such as those used in smart city applications or law enforcement surveillance. In these scenarios, the collection of vast amounts of facial data often lacks robust controls, making it vulnerable to breaches. Without strict adherence to privacy regulations and best practices, this threat undermines users' trust and compromises their fundamental right to privacy.

Context Unawareness

This is a privacy threat that occurs when a user is unaware of the information shared with a system, potentially revealing excessive details that aid attackers or providing inaccurate data, leading to errors in system decisions(Deng et al., 2011). This threat opposes the context awareness privacy property that emphasizes a user's understanding of their own data and the implications of sharing it. Users should be mindful of the potential privacy risks associated with sharing excessive personally identifiable information, as well as the negative outcomes that may arise from providing incomplete or inaccurate data(Deng et al., 2011). In terms FR systems, this threat occurs when users unknowingly share more biometric or contextual data than required. For instance, users may provide facial data captured in environments revealing sensitive personal or location-based details. This data, if improperly stored or analyzed, could inadvertently expose users to surveillance, profiling, or identity theft. Similarly, inaccurate facial data, such as low-quality images or poorly captured angles, can lead to false matches or system errors, potentially implicating individuals in legal matters or denying them access to essential services. Addressing this threat in FR systems requires transparent communication about data usage, explicit consent mechanisms, and the implementation of mechanisms that allow users to control the amount and type of data shared.

Policy and Consent Noncompliance

This the last threat in LIDDUN framework. It entails a situation where a system presents its privacy policies to users but fails to adhere to them in practice. As a result, users' personal data may still be exposed despite the assurance of compliance (Deng et al., 2011). This threat directly contradicts the privacy property of policy and consent compliance, which requires that systems adhere strictly to regulations and uphold the privacy standards they advertise.

With regards to face recognition (FR) systems, Policy and Consent Noncompliance can manifest in various ways. For example, an FR system may present a privacy policy that outlines data retention limits or specifies that biometric data will only be used for authentication purposes. However, the systems may later share or store users' facial data for longer periods or use it for additional purposes, such as surveillance or marketing, without obtaining proper consent. Additionally, users may not have a clear understanding of how their facial data will be processed or shared, and even if they do provide consent, the system might not follow through on its stated policies, leaving users' biometric data vulnerable to misuse or unauthorized access. This lack of compliance with privacy policies could significantly undermine user trust and violate privacy rights, making enforcement of policy compliance essential in FR systems. Table 1 shows the mapping of different privacy issues to the threats they may cause, the system elements that may be involved and the kind of privacy that is affected.

Table 1. Mapping among privacy issues, privacy threat, system elements and privacy types

Privacy Issues	Threat Types	System Elements	Privacy Types
Inadequate User Consent	Unawareness, Non-Compliance	Data Collection, System Policies	Privacy of the Person, Privacy of Behaviour, Privacy of Location and Space
Over-Collection of Data	Detectability, Linkability	Data Collection, Data Storage	Privacy of Data and Image
Improper Data Sharing	Disclosure of Information	Data Sharing, Data Processing	Privacy of Data and Image, Privacy of Association
Insufficient Data Protection	Disclosure of Information,	Data Storage, Data Processing	Privacy of Data and Image
Lack of User Control	Identifiability, Linkability	System Policies, Data Storage	Privacy of Data and Image
Inference of Sensitive Attributes	Identifiability, Non-Repudiation	Data Processing	Privacy of Thought and Feelings, Privacy of Behaviour

Safeguards for Privacy Issues and Threats in Face Recognition Systems

1. Inadequate User Consent
   - Implement clear and explicit consent mechanisms for data collection and processing.
   - Provide comprehensive and accessible privacy notices detailing how data will be used.
2. Over-Collection of Data
   - Enforce data minimization by collecting only the facial features necessary for the system's functionality.
   - Avoid capturing unnecessary data, such as entire video feeds, by implementing specific-use algorithms.
   - Regularly review and update data collection practices to align with privacy regulations.
3. Improper Data Sharing
   - Establish strict data-sharing policies that restrict sharing biometric data without user consent.
   - Encrypt data during transmission and storage to prevent unauthorized access.
   - Use data anonymization techniques to protect user identities when sharing data externally.
4. Insufficient Data Protection
   - Apply strong encryption to secure biometric data in storage and transmission.
   - Conduct regular security audits and vulnerability assessments to identify and address weaknesses.
   - Implement robust access controls to ensure only authorized personnel can access sensitive data.
5. Lack of User Control
   - Provide users with options to modify or delete their biometric data upon request.
   - Implement transparent data management practices and allow users to track how their data is being used.
   - Limit retention periods for facial data and automatically delete data when no longer necessary.
6. Detectability Threat
   - Enforce access controls and ensure database security to prevent unauthorized attempts to detect if data exists.

- Mask or encrypt records to make them indistinguishable from random data.
- Regularly monitor repositories to prevent excessive retention or cross-referencing with external sources.

7. Disclosure of Information
   - Restrict access to sensitive biometric data to only authorized entities.
   - Apply advanced encryption techniques to prevent unauthorized disclosure of personal data.
   - Ensure compliance with confidentiality standards through rigorous audits and policy enforcement.
8. Content Unawareness
   - Educate users on the implications of sharing biometric data through transparent and accessible privacy policies.
   - Limit the information requested during data collection to avoid unintentional over-disclosure.
   - Use feedback mechanisms to inform users about how their data is being used in real-time.
9. Policy and Consent Noncompliance
   - Regularly audit privacy policies to ensure they are accurately implemented in practice.
   - Adopt third-party oversight to verify compliance with advertised privacy policies.
   - Establish penalties and accountability measures for noncompliance to deter violations.

PRIVACY AND DIGITAL SOVEREIGNTY ISSUES IN FACE MORPHING MACHINE LEARNING DETECTION MODELS

In this section we analyze and discuss the causes of privacy breach associated with face morphing detection models. As with many machine learning models that processes training data to build a solution model of the problem, models developed for face morphing attack detection are also exposed to the same threats. Among the major source of privacy threats emanating from face morphing detection models include Model extraction, Data memorization, Model inversion and Membership inference (Zhang et al., 2020).

Model Extraction Attack

Model extraction Attack (Liang et al., 2024; Tramèr et al., 2016; Yan et al., 2023) is an attack on machine learning model where an adversary replicates the model's functionality by systematically querying it and analyzing the outputs. This process involves sending various inputs to the target model, collecting the corresponding outputs—such as class labels or confidence scores—and using this data to train a surrogate model that approximates the original. This attack could lead to a privacy threat whereby the attacker employs the extracted model to uncover sensitive information. In the operation of face morphing detection model as a service, this attack can lead to privacy breach when the attacker uses the extracted model to develop high quality morphed images that can evade detection, potentially compromising the accuracy of the face recognition systems leading to unauthorized access to sensitive biometric data. Also, in some cases, the extracted model can reveal sensitive information about the dataset it was trained on. This can lead to the inadvertent exposure of personal facial biometric data that were used to train the original model.

Data Memorization

This is the phenomenon where machine learning models "remember" aspects of their training data, potentially exposing sensitive information (Carlini et al., 2019; Song et al., 2017). Song et al. (2017) identified this concept by showing how an untrusted third-party algorithm, even when operating in a trusted isolation environment, could encode sensitive training data into a model without significantly affecting its overall accuracy. In the context of **face morphing detection models**, data memorization can lead to specific privacy breaches. For example, if a model is overfitted on a dataset of biometric images, it could "memorize" distinctive facial features or morphing artifacts specific to certain individuals. As a result, an attacker could query the model with new images and, by analyzing the model's response, potentially identify individuals in the original training set or reconstruct features of their faces. This could undermine the security and privacy of face recognition systems, especially in sensitive applications such as identity verification or access control.

Model Inversion Attack

This involves the adversary trying to reverse-engineer the original input based on the model's response to queries. The adversary can do this either through a black-box approach (only having access to model predictions) or a white-box approach (where the attacker has access to the model's internal parameters). The goal is often to recreate sensitive or private data that was part of the model's training set(Nguyen

et al., 2023; Yang et al., 2020; Y. Zhang et al., 2020). For face morphing detection, the attacker aims to invert the model by generating images or data points that are highly likely to be classified as certain identities or morphed faces by the model. This can be achieved by optimizing the input to the model in such a way that the output prediction corresponds to a specific identity or morph, revealing private information about individuals who were part of the training dataset. For instance, using methods like generative adversarial networks (GANs) or variational inference, attackers can reconstruct images from the model's outputs, effectively exposing private information about the faces used during training, even if those faces were not directly visible in the query inputs.

Membership Inversion Attack

This is another type of attack that poses privacy threat to face morphing detection model. **membership inference attack** involves determining whether a specific data sample was part of the model's training data. This is often achieved by analyzing the confidence scores or the output labels provided by the model in response to specific queries. The attacker can build a classifier using these outputs and determine whether a particular sample was trained on by the model(Al-Rubaie and Chang, 2019; Yang et al., 2020). in **face morphing detection models**, an attacker could use this attack to gain insights into whether a specific morphed face was used during training, which could be useful for adversaries seeking to bypass or compromise security systems. Table 2 provides the mapping of these privacy threat attack and the model components under breach with the possible countermeasures to avert the threat.

Table 2. Face morphing detection models' privacy threat, model elements and countermeasures

Attack Type	ML Component	Countermeasures	
Model Extraction	Model (outputs, structure)	Query Limitation, Model Watermarking, Differential Privacy	Limiting the number of queries an attacker can make or adding noise to outputs can prevent unauthorized model replication. Watermarking helps detect model stealing by embedding unique identifiers in the model's responses. Differential privacy ensures that the outputs do not leak sensitive information about the training data.
Data Memorization	Training Data, Model Parameters	Regularization (e.g., weight decay, dropout), Differential Privacy	Regularization prevents memorization of sensitive data by constraining the model. Techniques like weight decay and dropout reduce the model's reliance on specific training instances. Differential privacy adds noise to the model's training process to ensure that no single data point can overly influence the model's performance or reveal private data.
Model Inversion	Model Outputs (confidence scores)	Output Masking, Input Perturbation, Differential Privacy	Output masking or perturbing the input data helps obscure sensitive features that might be reconstructed by an attacker. Differential privacy ensures that the model outputs do not leak too much information about the training data, reducing the likelihood of an inversion attack.
Membership Inference	Model Outputs (labels, probabilities)	Differential Privacy, Confidence Thresholding, Training with Augmented Data	Using differential privacy ensures that the model cannot infer membership of specific data points. Confidence thresholding limits the model's sensitivity to outliers, and augmenting the training data reduces the likelihood of inferring membership based on specific data characteristics. This protects individuals' privacy by preventing attackers from determining if their data was part of the training set.

EFFECTS OF PRIVACY AND DIGITAL SOVEREIGNTY BREACH IN FACE RECOGNITION SYSTEM UNDER FACE MORPHING ATTACKS

Identity Theft and Fraud

The primary concern surrounding face morphing attacks is the danger associated with successful orchestration of such attacks. Attackers can impersonate legitimate users by submitting morphed images to biometric systems, thereby accessing sensitive information or conducting fraudulent activities. This creates the potential for serious privacy breaches, as personal data becomes vulnerable to exploitation. In the same vein this can lead to breach of digital sovereignty of user data if such stolen identities are transferred to a domain outside the governance of the original sovereign owner of the data.

Misuse of Personal Data

Face morphing not only undermines the security of identity verification systems but also facilitates the misuse of personal data. Morphed images, if intercepted, can be used to create false identities or perform illegal activities under the guise of legitimate users. The misuse of such data raises ethical and legal concerns, especially when individuals are unaware of the extent to which their personal information can be manipulated.

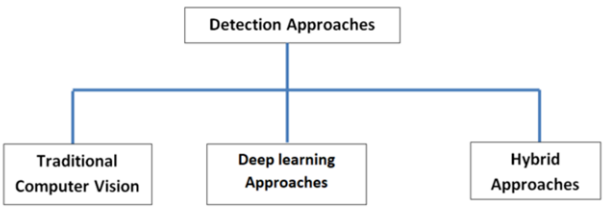
Threats to National Security

At the governmental level, face morphing attacks present significant threats to national security. For example, in border control systems where e-passports are used, the inability to detect morphed images could allow unauthorized individuals to enter a country, posing risks to immigration control and national safety.

METHODS OF FACE MORPHING ATTACK DETECTION

The approaches for solving the problem of face morphing can be broadly categorized into three as depicted in Figure 5.

Figure 5. Taxonomy of face morphing detection methods



Traditional Detection Techniques

Early attempts to detect face morphing attacks relied on traditional feature-based methods. These techniques focus on analyzing facial landmarks or geometric features of the face, such as the distance between the eyes or the shape of the nose. In the literature of image detection with traditional approaches researchers have used image feature extraction algorithms based on texture descriptors, keypoint descriptors, edge-gradient descriptors, frequency and wavelet descriptors, statisti-

cal descriptors among others. These techniques are either used individually or as a fusion of two or more.

The texture descriptors such as Local Binary Patterns (LBP), Local Phase Quantization (LPQ), Gabor Filters, Steerable Filters and BSIF (Binarized Statistical Image Features) are used to extract the surface properties or patterns. Ramachandra et al. (2020) utilized a steerable pyramid which decomposes an image into multiple scales to extract image features from their print scan dataset at different resolutions and allowing analysis on fine details to coarse structures. Kenneth et al. (2021) utilized LBP feature descriptor with decision tree classifier while Scherhag et al. (2020) also utilized a fusion of different configurations of multi-scale block local binary patterns. The same LBP features was also evaluated as part of other methods like SIFT, SURF, LPQ, HOG and BSIF in (Ramachandra et al., 2020; Scherhag et al., 2018b; Wandzik et al., 2018). Similarly, BSIF features with SVM classifier was primarily used by (avendra et altion methods, can Raghavendra et al. (2016)

Keypoint descriptors which are used for detecting and describing significant points in an image have also being used extensively, they include SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), ORB (Oriented FAST and Rotated BRIEF), BRIEF (Binary Robust Independent Elementary Features), FAST (Features from Accelerated Segment Test). SURF and SIFT features descriptors in conjunction with SVM classifier were evaluated as part of other features techniques for morph detection in (Scherhag et al., 2018b) . An ensemble of features across texture descriptors, Keypoint extractors and gradient estimators have also been proposed in (Scherhag et al., 2018a; Venkatesh et al., 2020).

Machine Learning and Deep Learning Approaches

As face morphing attacks have become more advanced, researchers have turned to machine learning and deep learning to develop more robust detection methods. Neural networks and deep learning models are particularly effective in recognizing subtle differences in pixel patterns and textures that are often invisible to the human eye. One popular technique involves the use of convolutional neural networks (CNNs), which are trained on large datasets of morphed and non-morphed images to learn the distinguishing features of morphed faces. Seibold et al. (2017) developed face morphing attacks detection model using deep convolutional neural networks. The use of deep learning feature embedding of images for images pairing selection for morphing and detection of resulted morphed images was investigated in (Kessler et al., 2024). Convolutional Neural network based demorphing network was employed in (Ortega-Delcampo et al., 2020) to discover the initial pictures from two images that has been morphed. By unraveling the constituent's images in a morphed image the method was able to detect morphing attack at ABC gate scenario. In another

research work presented in (Seibold et al., 2020), the use proposed neural network training schemes, which are based on different alternations of the training data, to increase robustness and generality of morphing attack detection model. Siamese network based VGG-16 architecture was introduced in (Soleymani et al., 2021) while deep representation from a pretrained vanilla vision transformer neural network was used with SVM for single image morph detection in(Zhang et al., 2024).

Hybrid Approaches

Hybrid approaches include approaches that combined traditional methods with deep learning methods. Authors in (Aghdaie et al., 2021; Chaudhary et al., 2021) combined traditional approach with deep earning approach by utilizing wavelet sub bands features computed on images as input to spatial attention mechanisms over convolutional and feed forward network to create a morphing detection system. Long et al. (2023) proposed a method utilizing two-stream network with channel attention and residual of multiple color spaces is proposed for face morphing detection. The method first obtains H, S, V, Y, Cb, Cr six color channel image, then use the bilateral filter for filtering the six color channel to get the corresponding residual noise image, then the combined six channel image and the residual noise image as input to the two-stream network for training.

CHALLENGES IN DETECTING FACE MORPHING ATTACKS

Sophistication of Morphing Techniques

One of the main challenges in detecting face morphing attacks is the increasing sophistication of morphing techniques. As morphing software evolves, it becomes more difficult to identify the subtle changes introduced to facial features. Well-crafted morphs can blend features so seamlessly that even advanced detection systems struggle to spot the alterations.

Generalization Across Datasets

Another significant challenge is ensuring that detection systems generalize well across different datasets. Many detection models perform well when tested on the dataset they were trained on but fail when applied to new, unseen data. This problem is compounded by the fact that faces vary significantly across populations in terms of age, gender, ethnicity, and facial expressions, making it difficult to build a one-size-fits-all detection model.

Adversarial Attacks and Countermeasures

Adversarial attacks, where attackers deliberately modify morphed images to bypass detection systems, pose an additional challenge. These attacks exploit weaknesses in detection models, forcing them to misclassify morphed images as legitimate ones. Researchers are working on countermeasures to make detection systems more resilient to adversarial attacks, but this remains an ongoing battle.

FUTURE DIRECTIONS

Improving Detection Algorithms

To combat the growing threat of face morphing attacks, future research must focus on developing more robust detection algorithms. These algorithms need to be capable of detecting even the most subtle morphs while maintaining high accuracy and low computational costs. Researchers are exploring the use of ensemble models, which combine multiple detection techniques to improve performance.

Development of Large-Scale Open-Source Morphed Faces for Research Purpose

Another issue faced in MAD is lack of publicly available MAD algorithm which can be used by researchers for comprehensive experimental evaluation of new and existing MAD algorithms. This situation brings about questions such as how reliable are the current state of-the-art MAD algorithms.

Integration with Multi-Modal Biometrics

Another promising direction for future research is the integration of multi-modal biometrics. By combining facial recognition with other biometric systems such as iris scans or fingerprints, security systems can reduce the risk of morphing attacks. This multi-layered approach provides additional safeguards, making it harder for attackers to fool the system.

Legislative and Policy Measures for Privacy Preservation of Data

Governments and organizations also need to address face morphing attacks through legislation and policy. Research into to frameworks to establish more ro-

bust privacy standards that govern the use of facial biometric data should be given adequate attention. This will help to protect individuals from the misuse of their personal information. Such frameworks should consider international dimension to create consistent policies and guidelines that address this global issue.

LIMITATIONS

This study is conceptual and does not include implementation-level experiments. Also, while we discuss policy recommendations, actual legislative frameworks may vary across countries.

CONCLUSION

Face morphing attacks present a significant challenge to the security and reliability of facial biometric systems. The privacy and digital sovereignty concerns in this domain affect both the face recognition system and the morphing detection models. Various approaches can be used to produce countermeasures to prevent these privacy and digital sovereignty threats. While traditional methods have proven inadequate for detecting face morphing attack, advancements in deep learning offer promising solutions. However, there is still much work to be done in addressing the challenges posed by sophisticated morphing techniques and adversarial attacks. Future research must focus on improving detection accuracy, integrating multi-modal biometrics, and establishing legal frameworks to protect privacy and digital sovereignty of the users' data and other digital infrastructures.

ACKNOWLEDGEMENT

This work was supported by funding provided by the Tertiary Education Trust Fund (TetFund) under the Nigeria Ministry of Education through the National Research Fund (NRF 2021) with research grant no: TETFUND/ES/DR-CE/NRF2021/SETI/ICT/00011/01.

REFERENCES

Aghdaie, P., Chaudhary, B., Soleymani, S., Dawson, J., & Nasrabadi, N. M. (2021). Attention Aware Wavelet-based Detection of Morphed Face Images. *2021 IEEE International Joint Conference on Biometrics (IJCB)*, 1–8. DOI: 10.1109/IJCB52358.2021.9484398

Al-Rubaie, M., & Chang, J. M. (2019). Privacy-Preserving Machine Learning: Threats and Solutions. *IEEE Security and Privacy*, *17*(2), 49–58. DOI: 10.1109/MSEC.2018.2888775

Badr, Y., Zhu, X., & Alraja, M. N. (2021). Security and privacy in the Internet of Things: Threats and challenges. *Service Oriented Computing and Applications*, *15*(4), 257–271. DOI: 10.1007/s11761-021-00327-z

Beltrán, M., & Calvo, M. (2023). A privacy threat model for identity verification based on facial recognition. *Computers & Security*, *132*, 103324. DOI: 10.1016/j.cose.2023.103324

Bhattacharya, M., Roy, S., Chattopadhyay, S., Das, A. K., & Shetty, S. (2023). A comprehensive survey on online social networks security and privacy issues: Threats, machine learning-based solutions, and open challenges. *Security and Privacy*, *6*(1), e275. DOI: 10.1002/spy2.275

Bu, F., Wang, N., Jiang, B., & Liang, H. (2020). "Privacy by Design" implementation: Information system engineers' perspective. *International Journal of Information Management*, *53*, 102124. DOI: 10.1016/j.ijinfomgt.2020.102124

Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., & Song, D. (2019). The secret sharer: Evaluating and testing unintended memorization in neural networks. *28th USENIX Security Symposium (USENIX Security 19)*, 267–284.

Chaudhary, B., Aghdaie, P., Soleymani, S., Dawson, J., & Nasrabadi, N. M. (2021). Differential Morph Face Detection using Discriminative Wavelet Sub-bands. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1425–1434. DOI: 10.1109/CVPRW53098.2021.00158

. Clarke, R. (2018). *Introduction to Dataveillance and Information Privacy, and Definitions of Terms, 2006*.

Dabrock, P., Tretter, M., Braun, M., & Hummel, P. (2021). Data sovereignty: A review. *Big Data & Society*, *8*(1).

DeBruine, L., & Jones, B. (2021). *Face Research Lab London Set*. DOI: 10.6084/m9.figshare.5047666.v5

Deng, M., Wuyts, K., Scandariato, R., Preneel, B., & Joosen, W. (2011). A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, *16*(1), 3–32. DOI: 10.1007/s00766-010-0115-7

. Hansen, M. (2008). Linkage control-integrating the essence of privacy protection into identity management. *Proceedings of eChallenges*, 1585–1592.

Jawad, L. A. (2024). Security and Privacy in Digital Healthcare Systems: Challenges and Mitigation Strategies. *Abhigyan*, *42*(1), 23–31.

Jiang, M., & Belli, L. (Eds.). (2025). *Digital Sovereignty in the BRICS Countries: How the Global South and Emerging Power Alliances Are Reshaping Digital Governance*. Cambridge University Press., DOI: 10.1017/9781009531085

Katsikas, S. K. (2025). Towards a cybersecurity-oriented research agenda for digital sovereignty. *Procedia Computer Science*, *254*, 279–288. DOI: 10.1016/j.procs.2025.02.087

Kenneth, O. M., Bashir, S. A., Abisoye, O. A., & Mohammed, A. D. (2021). Face morphing attack detection in the presence of post-processed image sources using neighborhood component analysis and decision tree classifier. *Information and Communication Technology and Applications: Third International Conference, ICTA 2020, Minna, Nigeria, November 24–27, 2020, Revised Selected Papers 3*, 340–354.

Kessler, R., Raja, K., Tapia, J., & Busch, C. (2024). Towards minimizing efforts for Morphing Attacks—Deep embeddings for morphing pair selection and improved Morphing Attack Detection. *PLoS One*, *19*(5), e0304610. DOI: 10.1371/journal.pone.0304610 PMID: 38820451

Klare, M., Hrestic, R., Stelter, A., & Lechner, U. (2025). Digital Sovereignty and Digital Transformation Practice Recommendation for the Software Life Cycle Process. *Procedia Computer Science*, *254*, 221–229. DOI: 10.1016/j.procs.2025.02.081

Liang, J., Pang, R., Li, C., & Wang, T. (2024). Model extraction attacks revisited. *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*, 1231–1245.

Long, M., Jia, C., & Peng, F. (2023). Face Morphing Detection Based on a Two-Stream Network with Channel Attention and Residual of Multiple Color Spaces. In Xu, Y., Yan, H., Teng, H., Cai, J., & Li, J. (Eds.), *Machine Learning for Cyber Security* (pp. 439–454). Springer Nature Switzerland. DOI: 10.1007/978-3-031-20102-8_34

López-Martínez, A., Gil-Pérez, M., & Ruiz-Martínez, A. (2023). A comprehensive review of the state-of-the-art on security and privacy issues in healthcare. *ACM Computing Surveys*, *55*(12), 1–38. DOI: 10.1145/3571156

. Melzi, P., Rathgeb, C., Tolosana, R., Vera-Rodriguez, R., & Busch, C. (2024). An Overview of Privacy-Enhancing Technologies in Biometric Recognition. *ACM Comput. Surv., 56*(12), 310:1-310:28. DOI: 10.1145/3664596

Misra, S., Barik, K., & Kvalvik, P. (2025). Digital Sovereignty in the Era of Industry 5.0: Challenges and Opportunities. *Procedia Computer Science*, *254*, 108–117. DOI: 10.1016/j.procs.2025.02.069

Nguyen, N.-B., Chandrasegaran, K., Abdollahzadeh, M., & Cheung, N.-M. (2023). Re-thinking model inversion attacks against deep neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16384–16393. DOI: 10.1109/CVPR52729.2023.01572

Ortega-Delcampo, D., Conde, C., Palacios-Alonso, D., & Cabello, E. (2020). Border Control Morphing Attack Detection With a Convolutional Neural Network De-Morphing Approach. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 92301–92313. DOI: 10.1109/ACCESS.2020.2994112

. Pfitzmann, A., & Hansen, M. (2010, August). *A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management*.

Raghavendra, R., Raja, K. B., & Busch, C. (2016). Detecting morphed face images. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 1–7. DOI: 10.1109/BTAS.2016.7791169

Ramachandra, R., Venkatesh, S., Raja, K., & Busch, C. (2020). Detecting Face Morphing Attacks with Collaborative Representation of Steerable Features. In B. B. Chaudhuri, M. Nakagawa, P. Khanna, & S. Kumar (Eds.), *Proceedings of 3rd International Conference on Computer Vision and Image Processing* (pp. 255–265). Springer Singapore. DOI: 10.1007/978-981-32-9088-4_22

Sarkar, E., Korshunov, P., Colbois, L., & Marcel, S. (2022). Are GAN-based morphs threatening face recognition? *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2959–2963.

Scherhag, U., Kunze, J., Rathgeb, C., & Busch, C. (2020). Face morph detection for unknown morphing algorithms and image sources: A multi-scale block local binary pattern fusion approach. *IET Biometrics*, *9*(6), 278–289. DOI: 10.1049/iet-bmt.2019.0206

Scherhag, U., Rathgeb, C., & Busch, C. (2018a). Morph Deterction from Single Face Image: A Multi-Algorithm Fusion Approach. *Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications*, 6–12. DOI: 10.1145/3230820.3230822

Scherhag, U., Rathgeb, C., & Busch, C. (2018b). Performance variation of morphed face image detection algorithms across different datasets. *2018 International Workshop on Biometrics and Forensics (IWBF)*, 1–6. DOI: 10.1109/IWBF.2018.8401562

Seibold, C., Samek, W., Hilsmann, A., & Eisert, P. (2017). Detection of Face Morphing Attacks by Deep Learning. In Kraetzer, C., Shi, Y.-Q., Dittmann, J., & Kim, H. J. (Eds.), *Digital Forensics and Watermarking* (pp. 107–120). Springer International Publishing. DOI: 10.1007/978-3-319-64185-0_9

Seibold, C., Samek, W., Hilsmann, A., & Eisert, P. (2020). Accurate and robust neural networks for face morphing attack detection. *Journal of Information Security and Applications*, *53*, 102526. DOI: 10.1016/j.jisa.2020.102526

Soleymani, S., Chaudhary, B., Dabouei, A., Dawson, J., & Nasrabadi, N. M. (2021). Differential Morphed Face Detection Using Deep Siamese Networks. In Del Bimbo, A., Cucchiara, R., Sclaroff, S., Farinella, G. M., Mei, T., Bertini, M., Escalante, H. J., & Vezzani, R. (Eds.), *Pattern Recognition. ICPR International Workshops and Challenges* (pp. 560–572). Springer International Publishing., DOI: 10.1007/978-3-030-68780-9_44

Song, C., Ristenpart, T., & Shmatikov, V. (2017). Machine learning models that remember too much. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 587–601. DOI: 10.1145/3133956.3134077

Tan, K. L., Chi, C.-H., & Lam, K.-Y. (2023). Survey on digital sovereignty and identity: From digitization to digitalization. *ACM Computing Surveys*, *56*(3), 1–36. DOI: 10.1145/3616400

Tramèr, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016). Stealing Machine Learning Models via Prediction APIs. *25th USENIX Security Symposium (USENIX Security 16)*, 601–618. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/tramer

Venkatesh, S., Ramachandra, R., Raja, K., & Busch, C. (2020). Single Image Face Morphing Attack Detection Using Ensemble of Features. *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 1–6. DOI: 10.23919/FUSION45008.2020.9190629

Wandzik, L., Kaeding, G., & Garcia, R. V. (2018). Morphing Detection Using a General- Purpose Face Recognition System. *2018 26th European Signal Processing Conference (EUSIPCO)*, 1012–1016. DOI: 10.23919/EUSIPCO.2018.8553375

Wang, X., Wu, Y. C., Zhou, M., & Fu, H. (2024). Beyond surveillance: Privacy, ethics, and regulations in face recognition technology. *Frontiers in Big Data*, *7*, 1337465. DOI: 10.3389/fdata.2024.1337465 PMID: 39027377

Woubie, A., Solomon, E., & Attieh, J. (2024). Maintaining Privacy in Face Recognition using Federated Learning Method. *IEEE Access : Practical Innovations, Open Solutions*, *12*, 39603–39613. DOI: 10.1109/ACCESS.2024.3373691

Wuyts, K., Scandariato, R., & Joosen, W. (2014). *LIND (D) UN privacy threat tree catalog*. CW Reports.

Yan, A., Hou, R., Yan, H., & Liu, X. (2023). Explanation-based data-free model extraction attacks. *World Wide Web (Bussum)*, *26*(5), 3081–3092. DOI: 10.1007/s11280-023-01150-6

Yang, Z., Shao, B., Xuan, B., Chang, E.-C., & Zhang, F. (2020). Defending model inversion and membership inference attacks via prediction purification. *arXiv Preprint arXiv:2005.03915*.

Zhang, H., Ramachandra, R., Raja, K., & Busch, C. (2024). Generalized Single-Image-Based Morphing Attack Detection Using Deep Representations from Vision Transformer. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1510–1518. DOI: 10.1109/CVPRW63382.2024.00158

Zhang, J., Li, C., Ye, J., & Qu, G. (2020). Privacy threats and protection in machine learning. *Proceedings of the 2020 on Great Lakes Symposium on VLSI*, 531–536. DOI: 10.1145/3386263.3407599

Zhang, Y., Jia, R., Pei, H., Wang, W., Li, B., & Song, D. (2020). The secret revealer: Generative model-inversion attacks against deep neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 253–261. DOI: 10.1109/CVPR42600.2020.00033

ADDITIONAL READING

Datta, P., Bhardwaj, S., Panda, S. N., Tanwar, S., & Badotra, S. (2020). Survey of Security and Privacy Issues on Biometric System. In Gupta, B., Perez, G., Agrawal, D., & Gupta, D. (Eds.), *Handbook of Computer Networks and Cyber Security*. Springer., DOI: 10.1007/978-3-030-22277-2_30

Jiang, M., & Belli, L. (Eds.). (2025). *Digital Sovereignty in the BRICS Countries: How the Global South and Emerging Power Alliances Are Reshaping Digital Governance*. Cambridge University Press., DOI: 10.1017/9781009531085

Singh, G., Bhardwaj, G., Singh, S. V., & Garg, V. (2021). Biometric Identification System: Security and Privacy Concern. In Awasthi, S., Travieso-González, C. M., Sanyal, G., & Kumar Singh, D. (Eds.), *Artificial Intelligence for a Sustainable Industry 4.0*. Springer., DOI: 10.1007/978-3-030-77070-9_15