Exploiting Human Vulnerabilities: A Practical Analysis of Social Engineering Attacks and Countermeasures

Ardi Benusi[1,*,†], Geri Selgjekaj[2,†]

[1] University of Tirana, Faculty of Natural Sciences, Bulevardi Zogu I, Tiranë 1001, Albania
[2] University of Tirana, Faculty of Natural Sciences, Bulevardi Zogu I, Tiranë 1001, Albania

Abstract

Social engineering attacks exploit human vulnerabilities rather than technical flaws, making them one of the most effective methods for breaching security systems. This study provides a practical analysis of social engineering attacks, examining the psychological manipulation techniques used by attackers and their real-world implications. Through case studies and hands-on experimentation, we identify common attack vectors such as phishing, pretexting, and baiting, assessing their success rates and impact. Furthermore, we evaluate existing countermeasures, including awareness training, behavioral interventions, and technical defenses, to determine their effectiveness in mitigating these threats. The findings highlight the urgent need for a multidisciplinary approach that combines cybersecurity measures with human-centered awareness strategies. This research aims to contribute to the development of more resilient defense mechanisms against social engineering attacks.

Keywords

Cybersecurity, Social Engineering, Phishing, Deep Fake, AI in Cybersecurity, Identity Management, Data Privacy, Security

1. Introduction

Social engineering involves manipulating people using various tactics to achieve specific outcomes. By exploiting human psychology, social engineers aim to persuade their targets to act in ways they might not have under normal circumstances. Traditional social engineering techniques like phishing, pretexting, and baiting are well known and extensively documented. However, the rise of deepfake technology, a form of AI-generated synthetic media, has added a new layer of complexity to these attacks. By producing highly realistic audio, video, and text, deepfakes allow attackers to mimic trusted individuals or craft deceptive scenarios with remarkable authenticity.

2. Background

To effectively carry out social engineering, several core principles are often employed. While the specific principles and their labels may differ depending on the source, some of the most frequently cited include:

- Authority is based on the tendency of individuals to comply with figures perceived as having power or control. Social engineers exploit this principle by posing as authoritative figures, such as managers, government representatives, or other roles that command respect or influence within a given context. By assuming such identities, they manipulate targets by following their directives.
- Intimidation operates by instilling fear or applying pressure to coerce an individual into complying with a specific demand. The target, feeling threatened or overwhelmed, is more likely to

*6th International Conference in Recent Trends and Applications in Computer Sciences and Information Technologies

[*] Corresponding author.
[†] These authors contributed equally.

ardi.benusi@fshn.edu.al (A. Benusi)
geri.selgjekaj@fshnstudent.info (G. Selgjekaj)

act in accordance with the social engineer's wishes to avoid perceived negative consequences. This tactic exploits the natural human response to perceived danger or authority.

- Consensus-based exploits the human tendency to conform to group behavior, leveraging the desire to follow what others are doing. In such attacks, the attacker might claim that everyone in a team or department has already performed a specific action, like clicking a link. This principle, sometimes referred to as "social proof," relies on the psychological inclination to align with the perceived actions or opinions of the majority.
- Scarcity is employed in social engineering by creating the perception that a resource, opportunity, or item is in limited supply, thereby increasing its perceived value. For example, a social engineer might claim that an offer is available only for a short time or that there are only a few items left, pressuring the target to act impulsively.
- Familiarity-based attacks exploit the natural human tendency to trust or feel positively toward individuals or organizations we recognize or have an affinity for. A social engineer might impersonate someone you know, a trusted colleague, to increase the likelihood of compliance. By leveraging pre-existing feelings of goodwill or recognition, the attacker makes their request or action seem more legitimate and less suspicious.
- Trust-based techniques rely on establishing a personal or emotional connection with the target to impose a sense of reliability and confidence. Unlike familiarity, which relies on pre-existing recognition or comfort, trust is actively cultivated by the social engineer through rapport-building, empathy, or shared interests. By creating this bond, the manipulator increases the likelihood that the target will comply with their requests, as they perceive the interaction as genuine and trustworthy.
- Urgency is a tactic used in social engineering to create a sense of immediate pressure, compelling the target to act quickly without thorough consideration. By presenting a situation as time-sensitive or critical, such as a limited-time offer, an impending deadline, or a perceived emergency, the manipulator exploits the target's instinct to respond swiftly, often bypassing rational decision-making. This approach increases the likelihood of compliance by inducing stress or fear of missing out.

3. Deepfake Technology

DeepFakes leverage generative adversarial networks (GANs) to produce highly realistic synthetic content. While they first gained attention in entertainment, they have since been misused for harmful activities such as disinformation, fraud, and social engineering. By replicating voices, facial features, and writing styles, deepfakes pose a serious threat to human trust and security.

3.1. Deepfake and Social Engineering Attacks

1. The use of this technology allows attackers to impersonate executives, government officials, or even family members with alarming accuracy. For instance, cybercriminals can generate a deepfake audio clip of a "CEO" instructing an employee to transfer funds or reveal sensitive data. This form of social engineering exploits trust, making it difficult to detect fraud.
2. DeepFakes have the power to create false events or statements, leading to confusion and a breakdown of trust. For example, a digitally altered video showing a political leader making provocative comments could spark social turmoil.
3. DeepFakes can exploit emotional vulnerabilities by creating fake distress calls or messages from loved ones, pressuring victims to act impulsively to protect their loved ones.

3.2. Real-world incident

1. CEO Fraud: A UK-based energy firm lost more than $200,000 after attackers used deepfake audio to impersonate the CEO. This is a real-world incident where AI-generated voice was used to impersonate a CEO and successfully stole in terms of hundreds of thousands.

2. Political Disinformation: A deepfake video of a political candidate went viral during an election campaign, influencing public opinion.
3. A deepfake video of Ukrainian President was circulated, falsely showing him surrendering to enemy forces.
4. Deepfake videos of CEO have been used in cryptocurrency scams, showing him "promoting" fake investment opportunities.

4. Social Engineering Exploitation by DeepFake

4.1. Authority

Many people tend to follow requests from perceived authority figures. Deepfakes create realistic impersonations of leaders or managers. The psychological effect on humans is greatly amplified by the accuracy of impersonation.

Deepfakes exploit a fundamental aspect of human psychology: our tendency to trust and obey authority figures. This phenomenon, rooted in social conditioning and cognitive biases, makes people more susceptible to manipulation when they believe they are interacting with a legitimate leader, manager, or other trusted individuals.

4.2. Urgency and Fear

Deepfakes can create scenarios that evoke urgency or fear, such as a fake emergency requiring immediate action. It can be used to manipulate emotions by creating fabricated emergencies that pressure victims into acting quickly. These scams exploit human psychology when faced with fear or urgency, because people are less likely to verify details and more likely to comply.

4.3. Trust and Familiarity

Deepfakes exploit trust by mimicking familiar voices or faces, making it difficult for victims to be alerted to authenticity. When people hear a familiar voice or see a convincing video of someone they know whether it's a CEO, a government official, or a family member, they instinctively assume it's real. This illusion of authenticity lowers skepticism and makes victims more likely to comply with requests, whether it's transferring money, sharing confidential information, or following misleading instructions.

5. Simulation with DeepFace open source application

To show the potential of image swapping using live streaming applications, a simulation is done using deepfake tools. The program uses machine learning models for tasks like face detection in the frame, facial point detection, and face replacement. These models require intensive computations, which graphical accelerator can handle tens of times faster than on an ordinary processor.

The architecture is based on the Model – View – Controller pattern. Camera source generates the frames and sends it to the next module for processing preserving the final frame per second between slower and faster modules. The final output module outputs the stream to the screen with some delays, which is needed to synchronize the sound.

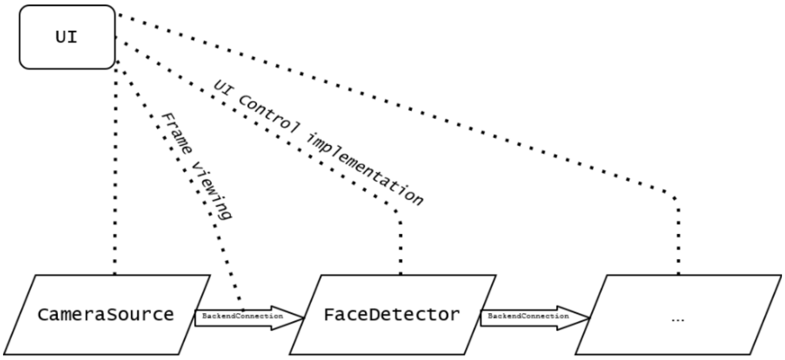

Figure 1: Architecture

A source camera supporting at least 5 frame per second (FPS) is used with a resolution of 1920 x 1080. The source frame looks like in Figure 2, where the real image is blurred to protect the identity of the person behind the camera.

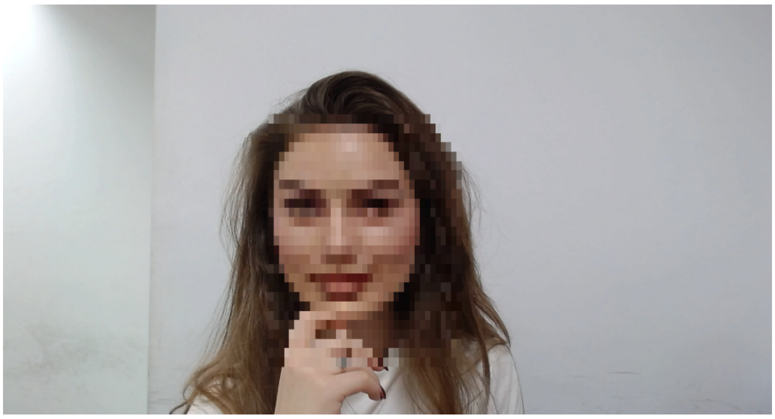

Figure 2: Source Live Streaming

To preserve the quality of the streaming and thus create a more credible picture, a graphic card like GeForce RTX 306 family is used. Face Aligner uses a 180 FPS to align the face before swapping. Face alignment has been in use for quite some time with growing concerns about misusing it for fraud.

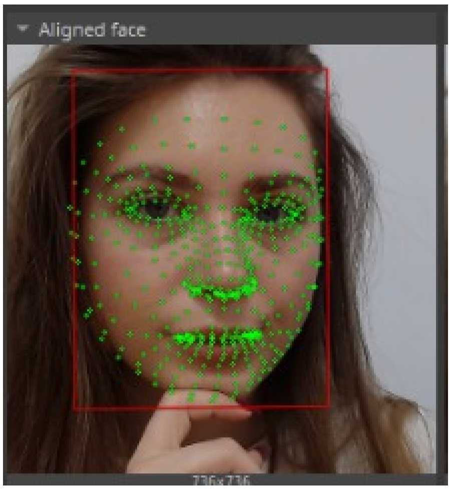

Figure 3: Face alignment

After the face alignment, a face swap like in Figure 4 is done taking a picture of a movie actress, utilizing the InsightFace model for facial recognition and manipulation. With this application, anyone can change faces in images or videos, creating engaging content. If fun was the only interest people would take from the swapping process, then the security issues would be something of the past. As it was described in real-world security incidents, face swapping has been exploited by black hackers and hacktivists for financial or political gains.

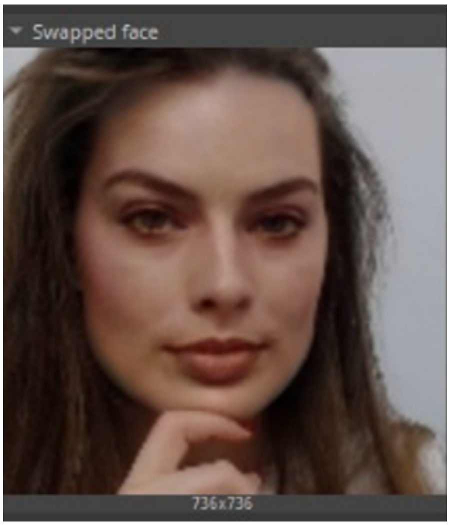

Figure 4: Swapped Face

The next step is integrating the swapped image into the streaming output. Bilinear interpolation with rct color transfer is used with 227 FPS frame adjuster to produce the swapped video as shown in Figure 5. Stream output can be easily uploaded or shared in a local/online web server.

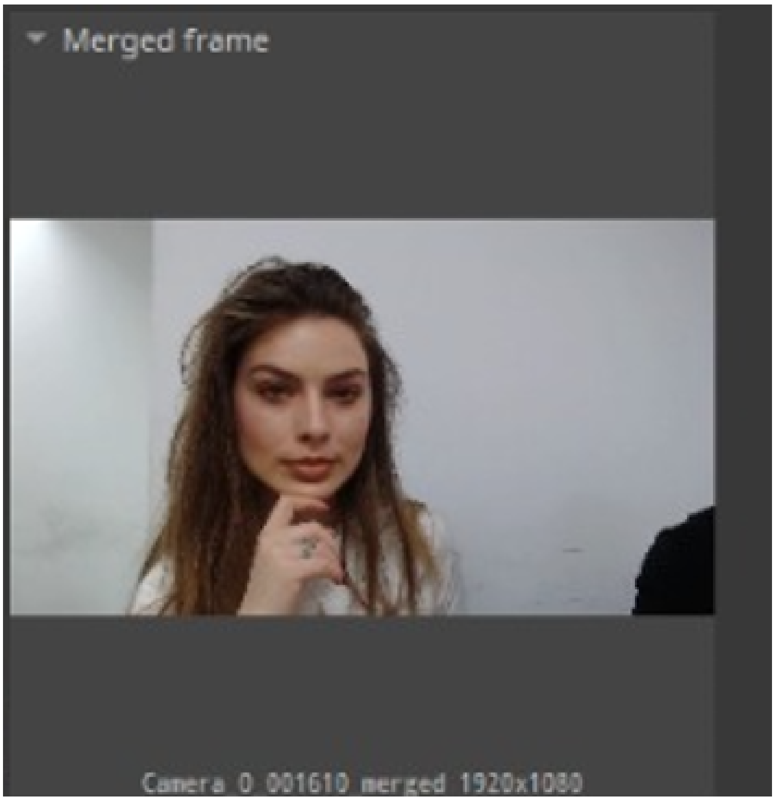

Figure 5: Target Live Streaming

Audio for public individuals can be easily found on the internet. To train the system at max 20 minutes medium quality audio material form the target person may be used to match the video with audio.

5.1. Possible misuse

Using faked or compromised accounts, the content may be injected and appear as legitimate. People following their favorite actress will accept any false content as true and possibly fall victims of scams, or phishing attacks.

The model does not need to be trained. There are public face models that can swap any face without training. However, if a particular face is going to be used for swapping a celebrity, it requires one day to train the model using an RTX GPU. That involves gathering four to five thousand samples of the source face with different lighting, different facial expressions, head direction, eyes direction, being far or closer to the camera. A filtering of not more than two thousand is enough.

Spreading political or religious content using deep fake may influence a lot of people, which can lead to misinformation, fraud, and harm to individuals.

6. Protection

Social engineering combined with phishing creates a powerful attack if used wisely by threat actors. Because psychological methods involve contact between individuals, attackers use a series of techniques to gain trust, such as:

- Not asking frequent and long questions, but short ones to collect basic information from several users to maintain trust.
- The request or question must be reasonable. For example, a question like "Is the director authorized" is more reasonable than the question "Can you check the director's printer to see if he printed the board meeting?"

- Flirting can facilitate the process of gathering information.
- Collecting as much information as possible without arousing suspicion.

6.1. Training and Awareness

Educating individuals about technologies and their potential misuse is critical. Training programs should focus on recognizing alerts and verifying suspicious requests, emails, phone calls, video calls or messages. Information security agencies must frequently publish on their social network channels video training on how to protect against social engineering attacks and especially on new emerging threats. Cybersecurity hygiene must be extended and reachable to wider groups of people who are not necessarily IT.

6.2. Solution using technology

Technology can help in addressing security issues. Detection AI tools can analyze media for signs of manipulation. Protection of user accounts from being compromised using multi-factor authentication could prevent unauthorized and misused access. Cloud and AI companies should advocate responsible AI usage and have implemented policies to prevent the misuse of their tools. A deepfake detecting tool would depend on the final quality of the picture like flickering face, abruptly clipping face mask, irregular colors. In one case, deepware.ai did detect fake pictures, but in a second case a good quality picture was used, where the program did not detect any fake. Training a model with higher face resolution gives a streaming output of better quality and makes it undetected by DeepFake detectors. We cannot depend solely on technology, as adversaries can also exploit it.

6.3. Regulatory Frameworks

Governments agencies and organizations must establish stronger policies to address the ethical and legal implications of DeepFake technology.

7. Conclusion

Deepfake technology marks a major advancement in social engineering attacks, taking advantage of human weaknesses with remarkable accuracy. As these threats grow more advanced, addressing them requires a comprehensive strategy that integrates awareness, technological solutions, and regulatory measures. By examining the connection between AI and human psychology, we can create stronger defenses against this evolving challenge. Security awareness combined with technology can, as in other social engineering attacks prevent these attacks from happening.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Albanian Cyber Academy 4th Edition, 2020. URL: https://aksk.gov.al/en/albanian-cyber-academy-4th-edition/

[2] Chapple, M., Seidl, D., & Greenley, R. (2023). CompTIA Security+ study guide with over 500 practice test questions: Exam SY0-701 (9th ed.). Sybex.

[3] Chesney, Robert and Citron, Danielle Keats, Deesp Fakes: A Looming Challenge for Privacy, Democracy, and National Security (July 14, 2018). 107 California Law Review 1753 (2019), U of Texas Law, Public Law Research Paper No. 692, U of Maryland Legal Studies Research Paper No. 2018-21, Available at SSRN: https://ssrn.com/abstract=3213954 or http://dx.doi.org/10.2139/ssrn.3213954

[4] Deepware. (2023). Deepware Scanner: AI-powered deepfake detection [Computer software]. URL: https://deepware.ai/scanner

[5] European Parliamentary Research Service. (2020). The ethics of artificial intelligence: Issues and initiatives (PE 634.452). European Parliament. URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf

[6] Heavybit. (2023, September 28). AI, data privacy, and security in IAM [Blog post]. Heavybit Library. URL: https://www.heavybit.com/library/article/ai-data-privacy-security-iam

[7] iPerov. (2022). DeepFaceLive: Real-time face swap for PC streaming or video calls [Computer software]. GitHub. https://github.com/iperov/DeepFaceLive

[8] Kim, P. (2018). The hacker playbook 3: Practical guide to penetration testing. Secure Planet LLC.

[9] Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and deepfakes. International Journal of Evidence & Proof, 23(4), 255–273. URL: https://doi.org/10.1177/1365712718807226.

[10] Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys, 54(1), 1–41. https://doi.org/10.1145/3425780.
doi: 10.1145/3425780.

[11] Whitman, M. E., & Mattord, H. J. (2021). Principles of information security (6th ed.). Cengage Learning.