

**Human-AI: Using Threat Intelligence to Expose Deepfakes and the
Exploitation of Psychology**

Tiffany Mary Tremont

Capitol Technology University

Chair: Dr. Ian McAndrew, FRAeS

February 2, 2023

Human-AI: Using Threat Intelligence to Expose Deepfakes and the
Exploitation of Psychology

Approved:

Dr. Ian McAndrew, Ph.D. FRAeS, Chair

Dr. Richard Baker, Ph.D., External Examiner

Accepted and Signed:

Richard E. Baker
Richard Baker, Ph.D., FRAeS

February 20, 2023

Date

Ian McAndrew
Ian McAndrew, Ph.D., FRAeS
Dean, Doctoral Programs
Capitol Technology University

20 February 2023

Date

ABSTRACT

This qualitative cyber threat deterrence theory research paper outlined countermeasures used to fight against psychological warfare deepfakes using facial recognition software to aid cybersecurity professionals in decision-making within private industry. Using Threat Intelligence to Expose Deepfakes and the Exploitation of Psychology, this research defined the single population and sampling strategy for this suggested qualitative study on Human artificial intelligence (AI). The sample and population of this research paper assisted in determining the breadth of cross-discipline cybersecurity professional subject matter experts (SMEs) sampling required for a comprehensive, non-biased investigation. Only qualified cybersecurity and threat intelligence professionals in the private industry participated in the study conducted entirely online. Participants were prescreened to ensure the target population's validity and improve data quality. A survey was created using an online survey tool called Survey Monkey. This survey was administered to capture participants' responses on how their tools, tactics, techniques, and procedures (TTP), strategies, and methods are used in defense against AI-enabled deepfake malware exploiting GAN for psychological warfare. Research questions answered this investigation's findings, and the contributions confirmed theories. As a result of these findings, we now understand how AI-based technology and countermeasures help private industry cybersecurity, and threat intelligence professionals defend against attacks. Future qualitative research on this topic may incorporate a case study research design methodology.

Keywords: artificial intelligence, advanced persistent threat, cognition, exploitation, decision-making, deepfake, generative adversarial network, human factors, information warfare, malware, metacognition, psychological warfare, threat intelligence, TTP

DEDICATION

With profound gratitude, I dedicate this research to the most cherished people in my life: my devoted husband, Kelly, and my extraordinary sons, Aiden and Liam. The immeasurable love, unwavering encouragement, and unwavering support that you have conferred upon me have been the guiding lights that have led me to the conclusion of this arduous journey and the attainment of my desired doctorate. Your presence has brought tenderness, inspiration, and motivation into my days, propelling me forward despite adversity. You have been my pillar of fortitude every step of the way, providing me with comfort and the confidence that I could overcome any obstacle. This accomplishment is a result of the profound influence your presence has had on my academic pursuits and personal development. Without you, this accomplishment would have remained an impossible dream. Therefore, I offer this humble tribute to you, my cherished family, for without you, this accomplishment would have been impossible.

ACKNOWLEDGMENT

1 Thessalonians 5:16–18 Rejoice always, pray without ceasing, in everything give thanks; for this is the will of God for you in Christ Jesus.

PREVIEW

Table of Contents

CHAPTER 1: INTRODUCTION	1
Background of the Study	2
AI Attack Tool.....	2
Deepfake Videos.....	3
Problem Statement	3
General Problem	3
Specific Problem.....	4
Purpose of the Study	4
Significance of the Study	4
Theoretical Significance	4
Practical Significance	5
Theoretical and Practical Significance Relationship	5
Psychological Warfare Deepfakes	6
Exploring Countermeasures	7
Observational Study.....	7
Research Method	8
Qualitative Approach Evaluation	9
Quantitative Approach Evaluation	10
Mixed-Method Approach Evaluation	11
Research Design.....	13
Appropriateness.....	14
Research Questions	14

Theoretical Framework	16
Population	18
Sampling	19
Generalizability	20
Definitions	21
Assumptions	23
Hypothesis	23
Bias	24
Scope, Limitations, and Delimitations	24
Scope	24
Limitations	25
Delimitations	26
Chapter Summary	27
CHAPTER 2: LITERATURE REVIEW	29
Research Question	29
Significance	29
Artificial Intelligence Revolution	30
Cognitive Closure and Confirmation Bias	30
Artificial Intelligence and Machine Learning Amplification	31
Pivotal Concepts in Weaponized Artificial Intelligence	32
Structured Analysis of Artificial Intelligence	33
Title Searches, Articles, Research Documents, and Journals Researched	36
Historical Overview	37
Defense-in-depth and Threat Mitigation Techniques	37

Emerging Attack Vectors.....	38
Seminal and Germinal Works.....	40
Artificial Intelligence-enabled Malware	40
Artificial Intelligence-Enabled Attack Dissertations	42
Artificial Intelligence-Enabled Attack Journals	42
Artificial Intelligence as a New Attack Vector	43
Current AI-enabled Deepfake Findings	45
Artificial Intelligence-Enabled Exploitation Studies	45
Prolific Gaps in Artificial Intelligence Affects.....	46
Artificial Intelligence Optimization through Social Media.....	47
Specific AI-enabled Deepfake Findings	48
Ethical Dilemmas	49
An Understanding of Deepfakes.....	51
Benefits of Deepfake and Facial Recognition Technologies	53
Deepfake Manufacturers.....	54
Threats Deepfakes Pose.....	55
Examples of Deepfakes	57
Technological Advances	58
Methods to Combat Deepfakes and Facial Recognition Technologies	61
Legislation and Regulation	62
Policies and Proactive Behaviors.....	65
Cognition, Metacognition, and Intention.....	67
Metacognition	67
Intention	68

	viii
Human Performance	70
Tactics, Techniques, and Procedures	72
Literature Analysis/Themes in Literature.....	75
NIST 800-37 Risk Management Framework	77
Chapter Summary.....	79
Chapter Conclusion	80
CHAPTER 3: RESEARCH METHOD.....	82
Research Methodology and Design	83
Theoretical Synopsis	83
Theory.....	84
Cyber Deterrence and Decision Theory.....	85
Population and Sample.....	85
Instrumentation.....	86
Operational Definitions and Analysis	88
Study Procedures	88
Data Collection and Analysis.....	89
Data Collection Procedures	89
Participants	90
Assumptions	90
Delimitations	91
Population.....	91
Sample	92
Ethical Assurances.....	92
Protection of Participants.....	93

Data Collection	93
Participants Benefits	94
Chapter Conclusion	94
CHAPTER 4: FINDINGS	96
LinkedIn Population	96
Industry Experience Criteria.....	96
Period of the Study	97
Development of the Findings	97
Pilot Study.....	98
Pilot 1 Iteration	98
Pilot 2 and 3 Iterations.....	98
LinkedIn Groups.....	99
Survey Participants' Validity and Reliability of the Data.....	99
Risks to Trustworthiness in Data Validity and Reliability.....	100
Survey Monkey Data Validity and Reliability.....	100
Results	101
Evaluation of the Findings	122
Demographics and Qualifications.....	123
Personal Challenges Deepfake AI-based Malware Vulnerabilities	131
Organizational Challenges Deepfake AI-based Malware Vulnerabilities Present	143
Methods to Combat Deepfake and AI-based Malware.....	163
Future Defensive Posture Against Deepfake and AI-powered Malware.....	175
Chapter Summary.....	183
Chapter Conclusion	184

	x
CHAPTER 5: IMPLICATIONS, RECOMMENDATIONS, AND CONCLUSIONS	185
Artificial Intelligence-enabled Risk.....	185
Artificial Intelligence Awareness	185
Approach	186
Assumptions, Limitations, and Delimitations.....	187
Study’s Importance.....	187
Purpose of the Study.....	188
Observational Study.....	188
Implications	189
Exclusion of Classified Weapons Systems Experts.....	189
Recommendations for Practice.....	190
Recommendation for Practice: Demographics and Qualifications (NIST) Step 1 – Prepare NIST	
800-37 TASK P-1 – Risk Management Roles	191
Recommendation #1.....	191
Recommendation #2.....	191
Recommendation #3.....	192
Recommendation #4.....	192
Recommendation #5.....	193
Opensource Tools.....	193
Recommendation #6.....	194
Respondent Comment 1	195
Respondent Comment 2	195
Respondent Comment 3	196
Recommendation #7.....	196

	xi
Recommendation #8.....	197
NIST 800-37 TASK P-2 – Risk Management Strategy.....	197
Recommendation #9.....	197
Recommendation #10.....	198
Respondent Comment 1	199
Respondent Comment 2	199
Respondent Comment 3	199
NIST 800-37 TASK P-3 – Risk Assessment - Organization.....	200
Recommendation #11.....	200
Recommendation #12.....	201
Recommendation #13.....	201
Recommendation for Practice: Personal Challenges Deepfake Artificial Intelligence-based Malware Vulnerabilities Present (NIST) Step 2 – Categorize	203
NIST 800-37 TASK C-2 – Security Categorization	203
Recommendation #1.....	203
Recommendation #2.....	204
Recommendation #3.....	204
Respondent Comment 1	205
Respondent Comment 2	207
Respondent Comment 3	208
Respondent Comment 4	209
Respondent Comment 5	209
Respondent Comment 6	209
Respondent Comment 7	209

	xii
Psychological Deepfakes as a Disruptor	210
Organizational Challenges Deepfake AI-based Malware Vulnerabilities Present	211
Recommendation #4.....	211
Opensource Tools	212
Respondent Comments 1	212
Recommendation #5.....	213
Recommendation for Practice: Organizational Challenges Deepfake Artificial Intelligence-based Malware Vulnerabilities Present (NIST) Step 3 – Select.....	215
NIST 800-37 TASK S1 – Control Selection	215
Recommendation #1.....	215
Respondent Comment 1	216
Respondent Comment 2	216
Recommendation #2.....	216
Recommendation #3.....	217
Recommendation #4.....	217
Recommendation #5.....	218
TASK S-2 – Control Tailoring	219
Recommendation #6.....	219
Recommendation #7.....	220
NIST 800-37 TASK S-3 - Control Allocation.....	221
Recommendation #1.....	221
Recommendation #2.....	221
Consideration #1	222
Opensource Tools	223

Recommendation for Practice: Methods to Combat Deepfake and Artificial Intelligence-based Malware (NIST) Step 4 – Implement.....	224
NIST 800-37 Step 5 – Assess	224
Recommendation #1.....	224
Opensource Tools	225
Respondent Comment 1	225
Recommendation #2.....	225
Recommendation #3.....	226
Recommendation #4.....	227
Recommendation #5.....	227
Respondent Comment 1	228
Respondent Comment 2	228
Respondent Comment 3	229
Respondent Comment 4	229
Recommendation for Practice: Future Defensive Posture Against Deepfake and Artificial Intelligence-powered Malware (NIST) Step 6 – Authorize.....	230
Recommendation #1.....	230
Recommendation #2.....	230
Respondent Comment 1	231
Recommendation #3.....	231
Respondent Comment 1	232
Recommendation #4.....	232
NIST 800-37 Step 7 - Monitor	233
Recommendation #1.....	234

	xiv
Respondent Comment 1	234
Respondent Comment 2	235
Respondent Comment 3	235
Respondent Comment 4	235
Respondent Comment 5	235
Recommendation #2.....	236
Respondent Comment 1	237
Respondent Comment 2	238
Consideration#1.....	239
Consideration #2	239
Recommendations for Future Research	240
Quantum Computing Threat.....	240
Quantum Computing Cybersecurity Preparedness Act	241
A Call for Quantum Security Mitigation Strategies	242
Contribution to the Body of Knowledge	243
Greatest Contribution to the Body of Knowledge	244
Chapter Summary.....	244
Chapter Conclusion	245
REFERENCES.....	246
APPENDIX A: KEY LITERATURE REVIEW SEARCH TERMS.....	270
APPENDIX B: LITERATURE SEARCH	274
APPENDIX C: LITERATURE MINDMAP	281
APPENDIX D: METHODOLOGY MAP	282
APPENDIX E: CONSENT TO PARTICIPATE IN RESEARCH	283

List of Tables

Table 1	43
---------------	----

List of Figures

Figure 1	20
Figure 2	61
Figure 3	78
Figure 4	85
Figure 5	123
Figure 6	124
Figure 7	125
Figure 8	126
Figure 9	127
Figure 10	128

Figure 11	130
Figure 12	131
Figure 13	132
Figure 14	133
Figure 15	134
Figure 16	135
Figure 17	136
Figure 18	137
Figure 19	140
Figure 20	142
Figure 21	143
Figure 22	145
Figure 23	146
Figure 24	147
Figure 25	148
Figure 26	150
Figure 27	153
Figure 28	154
Figure 29	155
Figure 30	157
Figure 31	159
Figure 32	161
Figure 33	163

Figure 34	164
Figure 35	166
Figure 36	167
Figure 37	169
Figure 38	171
Figure 39	173
Figure 40	175
Figure 41	177
Figure 42	179
Figure 43	181
Figure 44	183
Figure 45	190
Figure 46	202
Figure 47	214
Figure 48	220
Figure 49	223
Figure 50	224
Figure 51	229
Figure 52	233
Figure 53	240

CHAPTER 1: INTRODUCTION

Artificial Intelligence (AI) malware exploiting facial recognition technology is a serious threat to cybersecurity and society (Greengard, 2020; Jones, 2020; Stoecklin, 2018; Vaccari & Chadwick, 2020). As AI advances, more subtle forms of security risks are included in AI and machine learning (ML), contributing to evasive assaults such as Deeplocker—an AI-powered ultra-targeted malware (Stoecklin, 2018). Cognitive dissonance created using Generative Adversarial Networks (GANs) to magnify deepfakes using photographs, videos, and speech to manipulate thinking damages confidence psychologically (Chesney & Citron, 2019; Jones, 2020; Westerlund, 2019). The multipurpose fabrication of misleading content generated to influence a target audience has a powerful effect on people's beliefs, decisions, and perspectives (Carter et al., 2021). Autoencoders and GANs are examples of neural network designs that produce first-rate copies indistinguishable from the original, making it difficult to tell the real from the fake (Rupareliya, 2020).

Deepfake technology came into the public purview in early 2018 with the infamous video depicting United States President Barack Obama voicing opinions and false statements yet convincingly real by all appearances (Greengard, 2020; Jones, 2020; Vaccari & Chadwick, 2020). There are a variety of reasons to oppose the weaponization of deepfakes. Deepfake makers specifically design counterattacks to disrupt through the volume and velocity of disinformation (Jalaluddin, 2020; Jones, 2020; Peters, 2019). As a result, there are no commonly accepted countermeasure guidelines to assist in defense against psychological warfare deepfakes using facial recognition technologies to aid cybersecurity specialists' decision-making within private enterprise (Aiyanyo et al., 2020; Boivin, 2018; Jones, 2020; Westerlund, 2019).

Background of the Study

Malicious actors have made a name for themselves by using malware, threat vectors, and deepfakes to access private data and systems. Particularly worrisome is the possibility that AI could be used as an attack tool or even as an attack surface to help an adversary launch bigger, more autonomous attacks (Liu and Murphy, 2020). Facial recognition software revolutionizes the verification and interaction with mobile phones, computers, social networking, and other biometrics-required applications (apps). Cybersecurity frameworks and standards help firms identify unlawful assaults and patch newly discovered security flaws. As the digital transition proceeds in society, the convincing authenticity of deepfakes will remain a constant. It is common nowadays to underestimate AI technology's complexity; nevertheless, the breadth and scope of technology used to create deepfakes is the most remarkable aspect of AI's effect (Westerlund, 2020; Fletcher, 2018; Anderson, 2018). Therefore, this study discusses the countermeasures and methods used to defend against AI-based malware injection attacks for psychological weaponization using facial recognition software leading to targeted attacks of AI-enabled malware exploiting in the cybersecurity field. Further, through data gathered from cross-functional cybersecurity experts' implications of AI and its impact on the cybersecurity field in the current landscape: AI ethics, understanding deepfakes, benefits of deepfake and facial recognition technologies, who creates deepfakes, threats deepfakes pose, examples of deepfakes, AI legislation and regulation, policies and proactive behaviors, human performance, and methods used to combat deepfakes and facial recognition technologies.

AI Attack Tool

For destructive targeted assaults, AI, specifically deepfakes as an attack tool, is a source of worry (Langa, 2021).

Liu and Murphy, 2020, outline the benefits of applying AI methods to increase malware detection efficiency. To safeguard against evasive threats, an emphasis on the need for AI awareness raises the concern that exploitation of attack profiles and vulnerabilities may exist. This research will look at the advanced tactics and techniques used in AI malware to protect against psychological warfare deepfakes using facial recognition software to equip cybersecurity experts to make better decisions in the commercial sector. The emphasis is the need to synthesize the data to raise awareness of open-source AI tools utilized in defensive cybersecurity tactics, tools, procedures, approaches, tools, strategies, and procedures.

Deepfake Videos

"Deepfake videos are a present-day concern that will only become more challenging as ML technologies develop, increasing the authentic feel of the videos and making detection more difficult" (Peters, 2019, p.10). Face recognition software might produce volume and velocity in hybrid warfare as a contributor to AI-based malware assaults, as deepfake technology offers risks to malign and develop mistrust. (Jones, 2020). The implications of AI-generated deepfake audio and video technologies were investigated in this study (Jones, 2020). Jones (2020) claims that GAN tools allow users to alter video and audio recordings, making it difficult to determine what is a fake, as supported by Jones (2020).

Problem Statement

The research problem is to determine ways to defend against AI-based malware injection attacks for psychological weaponization using facial recognition software leading to targeted attacks (Jalaluddin, 2020; Jones, 2020).

General Problem

The general problem is that generative adversarial network (GAN) is weaponized through

AI malware to exploit facial recognition software (Boivin, 2018; Greengard, 2020; Jalaluddin, 2020; Jones, 2020; Stoecklin, 2018).

Specific Problem

The specific problem is that there are no widely adopted countermeasures identified against AI-enabled malware exploiting GAN for psychological warfare (Greengard, 2020; Jalaluddin, 2020; Jones, 2020; Labrien, 2016). Consequently, it necessitates the development and implementation of additional protective countermeasures.

Purpose of the Study

The purpose of this qualitative survey research design study is to discover the countermeasures used to defend against psychological warfare deepfakes utilizing facial recognition software to assist cybersecurity professionals in decision-making within private industry (Aiyanyo et al., 2020; Boivin, 2018; Jones, 2020; Westerlund, 2019).

Significance of the Study

Theoretical Significance

The theoretical significance of this study will be a guideline to an established single, flexible framework that is based on the seven-step National Institutes of Standards and Technology (NIST) 800-37 Rev. 2 | Risk Management Framework (RMF) for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy, cyber deterrence and decision-making theory, allowing for rigorous continual mitigation countermeasures against AI-enabled malware exploiting GAN for psychological warfare. This qualitative study's findings are important for providing a future defensive countermeasure framework, assisting cross-functional cybersecurity subject matter experts (SMEs), communities, and societies in identifying, assessing, and prioritizing risk and residual risk, and forming a

coordinated and economical application of resources to minimize, monitor, and control the probability and impact of adverse events.

Practical Significance

The practical significance is that if the resulting guideline is implemented, it will holistically benefit the community and society as a whole by developing a new defensive countermeasures guideline to supplement the RMF and serve as a foundation for cybersecurity subject matter expert (SME) decision-making. However, with the growth in cyber capabilities of both state and non-state actors, there is a great urgency to develop an approach that is not only scientifically sound but also operationally feasible. The analytical ML technique of utilizing GANs to create impossibly realistic visuals avoids the issue of facial recognition software. Future generations will be able to use this guideline as a reference, with or without extensive interpretation. Face recognition software used to unlock smartphones and computers, raises the possibility that AI-enabled malware will leverage deepfake technology by replacing a person's face and voice, or a combination of both, as Seker (2020) suggests. Despite this progress, the future of AI science is far from certain. Understanding how GANs generate deepfakes benefits cybersecurity professionals and specialists in detecting and defending against them. Future cybersecurity professionals may use this as a reference to defend against AI-enabled malware, preventing the exploitation of AI deepfake technology.

Theoretical and Practical Significance Relationship

The theoretical significance stems primarily from discovering a solution set of open-source countermeasures for protection and defense against the malicious use of AI facial recognition technologies. Furthermore, the practical significance of the findings affects how GAN-created deepfakes might benefit cybersecurity experts in defense and detection.

Specifically, the theoretical significance will be modelled after the NIST 800-37 Rev. 2 RMF to allow a methodical approach, as reflected by the respondent data and countermeasures provided used in response to AI-enabled malware, threat intelligence, and future deepfakes leveraging psychological cognition and perception bias. Controls are not repairs; rather, controls are automated or manual processes intended to avoid, identify, or remedy malicious activity. The elimination of tangible and intangible inconsistencies, as well as the search for disparate technologies, duplicate controls, or other inefficiencies in existing controls, is conducive to good outcomes in risk mitigation strategies due to the relationship between theoretical and practical significance. The risk treatment outlined in the study's guideline must be proportionate to the estimated target's worth. Adopting the guideline will define and assign control attributes, allowing for the deployment of effective and efficient control schemas that will thwart emerging threats.

Psychological Warfare Deepfakes

It is crucial to recognize and defend against psychological warfare deepfakes utilizing facial recognition software. Deepfake psychological warfare is an emerging weapon that has the potential to wreak havoc on a mass scale. GANS are a biometric standard recognized in the applicable International Organization of Standards (ISO) 30107. Standards provide baseline definitions for identity attributes of a living being such as fingerprints, iris, voice, and face. Seker (2020) agrees that detection methods to identify or circumvent verification procedures to detect deepfakes are a framework suitable for biometric software applications' security.

Since the inception of deepfake, researchers are working on a new framework for techniques to detect GANS by establishing statistical distribution characteristics (Seker, 2020). Software applications need to be made secure against malicious attacks. This research is unique