

FEDERAL BUREAU OF INVESTIGATION
FOI/PA
DELETED PAGE INFORMATION SHEET
FOI/PA# 1416188-001

Total Deleted Page(s) = 84
Page 140 ~ b1; b3; b6; b7C; b7E;
Page 141 ~ b1; b3; b7E;
Page 142 ~ Referral/Consult;
Page 143 ~ b1; b3; b6; b7C; b7E;
Page 144 ~ b1; b3; b7E;
Page 145 ~ b7E;
Page 146 ~ b7E;
Page 147 ~ b7E;
Page 148 ~ b7E;
Page 149 ~ b7E;
Page 150 ~ Referral/Consult;
Page 151 ~ Referral/Consult;
Page 152 ~ Referral/Consult;
Page 219 ~ b1; b3; b6; b7C; b7E;
Page 221 ~ b1; b3; b6; b7C; b7E;
Page 235 ~ Duplicate;
Page 236 ~ Duplicate;
Page 237 ~ Referral/Consult;
Page 238 ~ Referral/Consult;
Page 258 ~ Duplicate;
Page 259 ~ Duplicate;
Page 260 ~ Duplicate;
Page 261 ~ Duplicate;
Page 262 ~ Duplicate;
Page 263 ~ Duplicate;
Page 264 ~ Duplicate;
Page 265 ~ Duplicate;
Page 266 ~ Duplicate;
Page 267 ~ Duplicate;
Page 268 ~ Duplicate;
Page 269 ~ Duplicate;
Page 270 ~ Duplicate;
Page 271 ~ Duplicate;
Page 273 ~ Duplicate;
Page 274 ~ Duplicate;
Page 275 ~ Duplicate;
Page 276 ~ Duplicate;
Page 277 ~ Duplicate;
Page 278 ~ Duplicate;
Page 279 ~ Duplicate;
Page 280 ~ Duplicate;
Page 281 ~ Duplicate;
Page 282 ~ Duplicate;
Page 283 ~ Duplicate;
Page 284 ~ Duplicate;
Page 285 ~ Duplicate;
Page 286 ~ Duplicate;
Page 287 ~ Duplicate;
Page 288 ~ Duplicate;
Page 289 ~ Duplicate;
Page 290 ~ Duplicate;
Page 293 ~ Duplicate;
Page 294 ~ Duplicate;
Page 295 ~ Duplicate;
Page 296 ~ Duplicate;
Page 297 ~ Duplicate;
Page 298 ~ Duplicate;
Page 299 ~ Duplicate;
Page 300 ~ Duplicate;
Page 301 ~ Duplicate;
Page 302 ~ Duplicate;
Page 303 ~ Duplicate;
Page 304 ~ Duplicate;
Page 305 ~ Duplicate;
Page 306 ~ Duplicate;
Page 307 ~ Duplicate;
Page 308 ~ Duplicate;

Page 309 ~ Duplicate;
Page 310 ~ Duplicate;
Page 320 ~ Referral/Consult;
Page 321 ~ Referral/Consult;
Page 327 ~ Referral/Consult;
Page 328 ~ Referral/Consult;
Page 329 ~ Referral/Consult;
Page 330 ~ Referral/Consult;
Page 331 ~ Referral/Consult;
Page 332 ~ Referral/Consult;
Page 333 ~ Referral/Consult;
Page 334 ~ Referral/Consult;
Page 335 ~ Referral/Consult;
Page 336 ~ Referral/Consult;
Page 337 ~ Referral/Consult;
Page 338 ~ Duplicate;
Page 339 ~ Duplicate;

XXXXXXXXXXXXXXXXXXXXXXXXXXXX
X Deleted Page(s) X
X No Duplication Fee X
X For this Page X
XXXXXXXXXXXXXXXXXXXXXXXXXXXX

Congress of the United States
Washington, DC 20515

September 13, 2018

The Honorable Daniel R. Coats
Director of National Intelligence
Office of the Director of National Intelligence
Washington, DC 20511

Dear Director Coats:

We request that the Intelligence Community report to Congress and the public about the implications of new technologies that allow malicious actors to fabricate audio, video and still images.

Hyper-realistic digital forgeries — popularly referred to as “deep fakes” — use sophisticated machine learning techniques to produce convincing depictions of individuals doing or saying things they never did, without their consent or knowledge. By blurring the line between fact and fiction, deep fake technology could undermine public trust in recorded images and videos as objective depictions of reality.

You have repeatedly raised the alarm about disinformation campaigns in our elections and other efforts to exacerbate political and social divisions in our society to weaken our nation. We are deeply concerned that deep fake technology could soon be deployed by malicious foreign actors.

Forged videos, images or audio could be used to target individuals for blackmail or for other nefarious purposes. Of greater concern for national security, they could also be used by foreign or domestic actors to spread misinformation. As deep fake technology becomes more advanced and more accessible, it could pose a threat to United States public discourse and national security, with broad and concerning implications for offensive active measures campaigns targeting the United States.

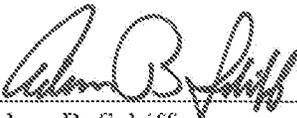
Given the significant implications of these technologies and their rapid advancement, we believe that a thorough review by the Intelligence Community is appropriate, including an assessment of possible counter-measures and recommendations to Congress. Therefore, we request that you consult with the heads of the appropriate elements of the Intelligence Community to prepare a report to Congress, including an unclassified version, that includes:

- (a) An assessment of how foreign governments, foreign intelligence services or foreign individuals could use deep fake technology to harm United States national security interests;
- (b) A description of any confirmed or suspected use of deep fake technology by foreign governments or foreign individuals aimed at the United States that has already occurred to date;
- (c) An identification of technological counter-measures that have been or could be developed and deployed by the United States Government or by the private sector to deter and detect the use of deep fakes, as well as analysis of the benefits, limitations and drawbacks, including privacy concerns, of such counter-technologies;

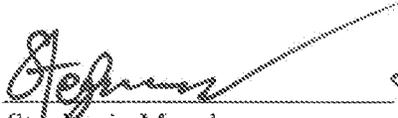
- (d) An identification of the elements of the Intelligence Community that have, or should have, lead responsibility for monitoring the development of, use of and response to deep fake technology;
- (e) Recommendations regarding whether the Intelligence Community requires additional legal authorities or financial resources to address the threat posed by deep fake technology;
- (f) Recommendations to Congress regarding other actions we may take to counter the malicious use of deep fake technologies; and
- (g) Any other information you believe appropriate.

We would appreciate your cooperation in producing this report as soon as feasible, but no later than December 14, 2018. Thank you for your assistance.

Sincerely,



Adam B. Schiff
MEMBER OF CONGRESS



Stephanie Murphy
MEMBER OF CONGRESS



Carlos Corbeo
MEMBER OF CONGRESS

[Redacted]

(IMD) (CON)

From: [Redacted] (CYD) (FBI)
Sent: Monday, August 07, 2017 12:22 PM
To: [Redacted] (CYD) (FBI); [Redacted] (CYD) (FBI)
Subject: AI NatSec - final.pdf
Attachments: AI NatSec - final.pdf

b6
b7C

For Thursday.

[Redacted]

(IMD) (CON)

b6
b7C

From: [Redacted] (CYD) (FBI)
Sent: Monday, July 23, 2018 3:05 PM
To: [Redacted] (CYD) (FBI); [Redacted] (CYD) (FBI)
Subject: RE: AEP topics

In case you want to aggregate for the whole Unit before we respond to [Redacted]

- [Redacted]

10 and 44 pose the same technical question. The difference is in the application.

- [Redacted]
- [Redacted]
- [Redacted]

b7E

Regards,

[Redacted]

FBI Cyber Division

b6
b7C
b7E

[Redacted] (desk)
[Redacted] (mobile)

-----Original Message-----

From: [Redacted] (CYD) (FBI)
Sent: Monday, July 23, 2018 2:41 PM
To: [Redacted]

b6
b7C
b7E

[Redacted]; Karl, Larry D. (CYD) (FBI); [Redacted]

Subject: AEP topics

Afternoon everyone. I will be serving as one of the reviewing officials determining the topics for next year's DHS analyst exchange program. The selection committee will likely pick 6-10 topics and we have the opportunity to modify or combine proposed topics. Wanted to solicit your thoughts and feedback on the proposed topics. Pls let me know if you have any questions. Thanks. Have a great day.

[Redacted]

FBI Senior National Intelligence Officer for Cyber MR 410: [Redacted] FBIHQ 11816: [Redacted]

b6
b7C
b7E

Cell: [Redacted]

[Redacted]

BELFER CENTER STUDY

Artificial Intelligence and National Security

Greg Allen

Taniel Chan

A study on behalf of Dr. Jason Matheny, Director of the U.S.
Intelligence Advanced Research Projects Activity (IARPA)



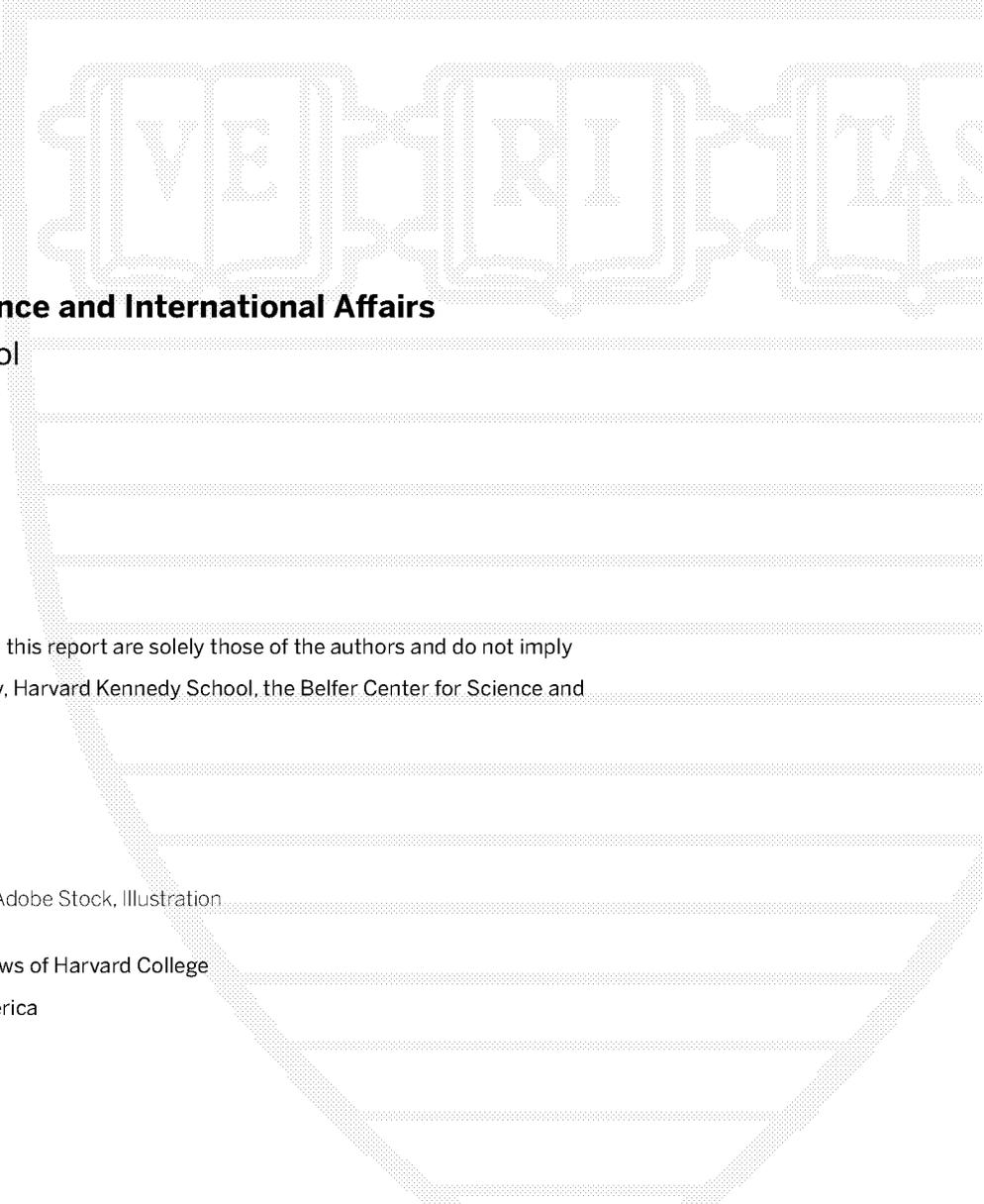
HARVARD Kennedy School

BELFER CENTER

for Science and International Affairs

STUDY

JULY 2017



Belfer Center for Science and International Affairs

Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138

www.belfercenter.org

Statements and views expressed in this report are solely those of the authors and do not imply endorsement by Harvard University, Harvard Kennedy School, the Belfer Center for Science and International Affairs, or IARPA.

Design & Layout by Andrew Facini

Cover photo and opposite page 1: Adobe Stock, Illustration

Copyright 2017, President and Fellows of Harvard College

Printed in the United States of America

Artificial Intelligence and National Security

Greg Allen

Taniel Chan

A study on behalf of Dr. Jason Matheny, Director of the U.S.
Intelligence Advanced Research Projects Activity (IARPA)



HARVARD Kennedy School

BELFER CENTER

for Science and International Affairs

STUDY
JULY 2017

Acknowledgments

We would like to thank our advisors at Harvard Kennedy School and Harvard Business School, Dr. Joseph Nye Jr. and Dr. Gautam Mukunda, respectively, for their insight and dedication to the success of this project. We also thank Dr. Matthew Bunn and Dr. John Park, who provided timely feedback at multiple stages of this effort.

We are grateful to the Belfer Center for Science and International Affairs, the Mossavar-Rahmani Center for Business and Government, and the Social Enterprise Initiative for their generous financial support, without which this research would not have been possible.

We would also like to thank our client, Dr. Jason Matheny, who went out of his way to support this effort at every step.

We thank the dozens of experts across government, industry, and academia who shared their time and expertise in being interviewed. We would especially like to thank Matt Daniels at the Department of Defense's Office of Net Assessment. Matt's brilliant insights can be found throughout this document.

Finally, we would like to thank those individuals who took the time to review and provide feedback on early drafts of this document, including Dr. Edward Felten, Dr. Richard Danzig, Dr. Lynne Parker, Ambassador Richard Norland, and Dr. Randy Bryant. Without their assistance, this document would have been far weaker. Any remaining mistakes are ours alone.

Project Overview

Partially autonomous and intelligent systems have been used in military technology since at least the Second World War, but advances in machine learning and Artificial Intelligence (AI) represent a turning point in the use of automation in warfare. Though the United States military and intelligence communities are planning for expanded use of AI across their portfolios, many of the most transformative applications of AI have not yet been addressed.

In this piece, we propose three goals for developing future policy on AI and national security: preserving U.S. technological leadership, supporting peaceful and commercial use, and mitigating catastrophic risk. By looking at four prior cases of transformative military technology—nuclear, aerospace, cyber, and biotech—we develop lessons learned and recommendations for national security policy toward AI.

About the Authors

Greg Allen is an Adjunct Fellow at the Center for A New American Security in Technology and National Security Program. Mr. Allen focuses on the intersection of Artificial Intelligence, cybersecurity, robotics, and national security. His writing and analysis has appeared in *WIRED*, *Vox*, and *The Hill*. Mr. Allen holds a joint MPP/MBA degree from the Harvard Kennedy School of Government and the Harvard Business School. Find him on twitter @gregory_c_allen

Taniel Chan was the Associate Director of Strategy and Analytics for the NYC Department of Education and a financial and economic analyst at Goldman Sachs. Mr. Chan also worked for the White House National Economic Council on technology industry policy and workforce development. This fall, he will be join Bain & Company in London. Mr. Chan holds a joint MPP/MBA degree from the Harvard Kennedy School of Government and the Harvard Business School.

Table of Contents

Executive Summary	1
Introduction & Project Approach.....	7
Part 1: The Transformative Potential of Artificial Intelligence	12
Implications for Military Superiority.....	12
Implications for Information Superiority.....	27
Implications for Economic Superiority.....	35
Part 2: Learning from Prior Transformative Technology Cases	42
Key Technology Management Aspects.....	42
Government Technology Management Approach.....	44
Government Management Approach “Scorecard”.....	45
AI Technology Profile: A Worst-case Scenario?	46
Lessons Learned	48
Part 3: Recommendations for Artificial Intelligence and National Security	58
Preserving U.S. Technological Leadership.....	58
Supporting Peaceful Use of AI Technology	64
Mitigating Catastrophic Rsk	67
Conclusion	70
Appendix: Transformative National Security Technology Case Studies.....	71
Case Study #1: Nuclear Technology	71
Case Study #2: Aerospace Technology.....	82
Case Study #3 Internet and Cyber Technology.....	91
Case Study #4 Biotechnology	100
 Citations.....	 111





Executive Summary

- **Researchers in the field of Artificial Intelligence (AI) have demonstrated significant technical progress over the past five years, much faster than was previously anticipated.**
 - Most of this progress is due to advances in the AI sub-field of machine learning.
 - Most experts believe this rapid progress will continue and even accelerate.
- **Most AI research advances are occurring in the private sector and academia.**
 - Private sector funding for AI dwarfs that of the United States Government.
- **Existing capabilities in AI have significant potential for national security.**
 - For example, existing machine learning technology could enable high degrees of automation in labor-intensive activities such as satellite imagery analysis and cyber defense.
- **Future progress in AI has the potential to be a transformative national security technology, on a par with nuclear weapons, aircraft, computers, and biotech.**
 - Each of these technologies led to significant changes in the strategy, organization, priorities, and allocated resources of the U.S. national security community.
 - We argue future progress in AI will be at least equally impactful.

- **Advances in AI will affect national security by driving change in three areas: military superiority, information superiority, and economic superiority.**
 - For military superiority, progress in AI will both enable new capabilities and make existing capabilities affordable to a broader range of actors.
 - For example, commercially available, AI-enabled technology (such as long-range drone package delivery) may give weak states and non-state actors access to a type of long-range precision strike capability.
 - In the cyber domain, activities that currently require lots of high-skill labor, such as Advanced Persistent Threat operations, may in the future be largely automated and easily available on the black market.
 - For information superiority, AI will dramatically enhance capabilities for the collection and analysis of data, and also the creation of data.
 - In intelligence operations, this will mean that there are more sources than ever from which to discern the truth. However, it will also be much easier to lie persuasively.
 - AI-enhanced forgery of audio and video media is rapidly improving in quality and decreasing in cost. In the future, AI-generated forgeries will challenge the basis of trust across many institutions.
 - For economic superiority, we find that advances in AI could result in a new industrial revolution.
 - Former U.S. Treasury Secretary Larry Summers has predicted that advances in AI and related technologies will lead to a dramatic decline in demand for labor such that the United States “may have a third of men between the ages of 25 and 54 not working by the end of this half century.”

- Like the first industrial revolution, this will reshape the relationship between capital and labor in economies around the world. Growing levels of labor automation might lead developed countries to experience a scenario similar to the “resource curse.”
 - Also like the first industrial revolution, population size will become less important for national power. Small countries that develop a significant edge in AI technology will punch far above their weight.
- **We analyzed four prior cases of transformative military technologies—nuclear, aerospace, cyber, and biotech—and generated “lessons learned” for AI.**
 - **Lesson #1:** Radical technological change begets radical government policy ideas.
 - As with prior transformative military technologies, the national security implications of AI will be revolutionary, not merely different.
 - Governments around the world will consider, and some will enact, extraordinary policy measures in response, perhaps as radical as those considered in the early decades of nuclear weapons.
 - **Lesson #2:** Arms races are sometimes unavoidable, but they can be managed.
 - In 1899, Fears of aerial bombing led to an international treaty banning the use of weaponized aircraft, but voluntary restraint was quickly abandoned and did not stop air war in WWI.
 - The applications of AI to warfare and espionage are likely to be as irresistible as aircraft. Preventing expanded military use of AI is likely impossible.

- Though outright bans of AI applications in the national security sector are unrealistic, the more modest goal of safe and effective technology management must be pursued.
- **Lesson #3:** Government must both promote and restrain commercial activity.
 - Failure to recognize the inherent dual-use nature of technology can cost lives, as the example of the Rolls-Royce Nene jet engine shows.
 - Having the largest and most advanced digital technology industry is an enormous advantage for the United States. However, the relationship between the government and some leading AI research institutions is fraught with tension.
 - AI Policymakers must effectively support the interests of both constituencies.
- **Lesson #4:** Government must formalize goals for technology safety and provide adequate resources.
 - In each of the four cases, national security policymakers faced tradeoffs between safety and performance, but the government was more likely to respond appropriately to some risks than to others.
 - Across all cases, safety outcomes improved when the government created formal organizations tasked with improving the safety of their respective technology domains and appropriated the needed resources.
 - These resources include not only funding and materials, but talented human capital and the authority and access to win bureaucratic fights.
 - The United States should consider standing up formal research and development organizations tasked with investigating and promoting AI safety across the entire government and commercial AI portfolio.

- **Lesson #5:** As technology changes, so does the United States’ National Interest.
 - The declining cost and complexity of bioweapons led the United States to change their bioweapons strategy from aggressive development to voluntary restraint.
 - More generally, the United States has a strategic interest in shaping the cost, complexity, and offense/defense balance profiles of national security technologies.
 - As the case of stealth aircraft shows, targeted investments can sometimes allow the United States to affect the offense/defense balance in a domain and build a long-lasting technological edge.
 - The United States should consider how it can shape the technological profile of military and intelligence applications of AI.
- **Taking a “whole of government” frame, we provide three goals for U.S. national security policy toward AI technology and provide 11 recommendations.**
 - Preserve U.S. technological leadership
 - Recommendation #1: The DOD should conduct AI-focused war-games to identify potential disruptive military innovations.
 - Recommendation #2: The DOD should fund diverse, long-term-focused strategic analyses on AI technology and its implications.
 - Recommendation #3: The DOD should prioritize AI R&D spending areas that can provide sustainable advantages and mitigate key risks.
 - Recommendation #4: The U.S. defense and intelligence community should invest heavily in “counter-AI” capabilities for both offense and defense.

- Support peaceful use of the technology
 - Recommendation #5: DARPA, IARPA, the Office of Naval Research, and the National Science Foundation should be given increased funding for AI-related basic research.
 - Recommendation #6: The Department of Defense should release a Request for Information (RFI) on Dual-Use AI capabilities.
 - Recommendation #7: In-Q-Tel should be given additional resources to promote collaboration between the national security community and the commercial AI industry.

- Manage catastrophic risks
 - Recommendation #8: The National Security Council, the Defense Department, and the State Department should study what AI applications the United States should seek to restrict with treaties.
 - Recommendation #9: The Department of Defense and Intelligence Community should establish dedicated AI-safety organizations.
 - Recommendation #10: DARPA should fund research on fail-safe and safety-for-performance technology for AI systems.
 - Recommendation #11: NIST and the NSA should explore technological options for countering AI-enabled forgery.

Introduction & Project Approach

Over the past five years, researchers have achieved key milestones in Artificial Intelligence (AI) technology significantly earlier than prior expert projections.

Go is a board game with exponentially greater mathematical and strategic depth than chess. In 2014, the computer expert who had designed the world's best Go-playing program estimated that it would be ten more years until a computer system beat a human Go champion.¹ Instead, researchers at DeepMind achieved that goal one year later.² Other researchers have since achieved new milestones in diverse AI applications. These include beating professional poker players,³ reliable voice recognition,⁴ image recognition superior to human performance,⁵ and defeating a former U.S. Air Force Pilot in an air combat simulator.⁶

There are four key drivers behind the rapid progress in AI technology:

1. Decades of exponential growth in computing performance
2. Increased availability of large datasets upon which to train machine learning systems
3. Advances in the implementation of machine learning techniques
4. Significant and rapidly increasing commercial investment

Combined, these trends appear poised to continue delivering rapid progress for at least another decade.^A Leading commercial technology companies report that they are “remaking themselves around AI.”

Most of the recent and near-future progress falls within the field of *Narrow AI* and machine learning, specifically. *General AI*, meaning AI with the scale and fluidity of a human brain, is assumed by most researchers to be at least several decades away.⁷

There are strong reasons to believe—as many senior U.S. defense and intelligence leaders do—that rapid progress in AI is likely to impact national security.

Deputy Secretary of Defense Robert Work, a leader in developing and implementing the Department of Defense’s “Third Offset” strategy, supported this view in a speech at the Reagan Defense forum: “To a person, every single person on the [Defense Science Board Summer Study] said, we can’t prove it, but we believe we are at an inflection point in Artificial Intelligence and autonomy.”⁸ Such statements indicate national security leaders are confident that rapid progress in AI technology will continue and will have impact a significant impact on their mission.

A Of these trends, exponential growth in computational power and storage appears the most vulnerable due to the recent slowdown in the pace of shrinking silicon transistors. However, there are many proposed paths for achieving continued improvements in computing hardware performance, including the use of processors custom designed for implementation of neural networks and machine learning. For a discussion of these issues, see “The Beast from Below—How Changes in the Hardware Ecosystem Will Disrupt Computer Science” by Doug Burger of Microsoft Research.

The U.S. government has recently sponsored several significant studies on the future of AI and its implications for governance and national security.^B

These studies are generally concerned with the near-term future of AI and are especially concerned with increased utilization of Deep Learning techniques.

Apart from the Office of Net Assessment's Summer Study,⁹ work to date generally does not focus on the long-term, more transformative implications of AI. This work is intended to assist in closing that gap.

Our Approach—Part 1: Analyzing possible technology development scenarios related to AI and exploring how these might transform national security

In this report, we supplement work to date with greater consideration across three dimensions:

- Greater diversity in potential applications of advances in AI
- Greater analysis of the implications of AI advances beyond what is currently possible or expected to be possible in the next five years
- Greater consideration of what technology management paradigms are best suited for AI and evaluating these in historical context

^B See for example:

- June 2016—Defense Science Board: "Summer Study on Autonomy"
- July 2016—Department of Defense Office of Net Assessment: "Summer Study: (Artificial) intelligence: What questions should DoD be asking"
- October 2016—National Science and Technology Council: "The National Artificial Intelligence Research and Development Strategic Plan"
- October 2015—National Science and Technology Council: "Preparing for the Future of Artificial Intelligence"
- December 2016—Executive Office of the President: "Artificial Intelligence, Automation, and the Economy"
- January 2017—JASON: "Perspectives on Research in Artificial Intelligence and Artificial General intelligence Relevant to DoD"

Our Approach—Part 2: Evaluating prior transformative military technologies in order to generate “lessons learned” for designing responses to the emergence of an important field of technology such as AI

We argue that AI technology is likely to be a transformative military technology, on a par with the invention of aircraft and nuclear weapons. Governments have long competed for leadership over rivals in driving and harnessing technological progress. Though machine learning and AI technology are comparatively young, human and organizational responses to the new technology are often echoes of prior experiences.¹⁰ We believe learning from the past offers significant wisdom with which to guide a future course of action with respect to AI.

Accordingly, we investigate four prior cases of transformative technologies which we believe to be especially instructive and relevant for AI. These are listed in Figure 1.

Figure 1: Four case studies of transformative military technologies

Nuclear	Cyber
Aerospace	Biotech

Our Approach—Part 3: Providing AI-related policy recommendations to preserve U.S. technological leadership, support peaceful AI use, and mitigate catastrophic risk

For each case, we focus on the early decades of these technologies after they began to see military application. During this period, responsible agencies had to develop technology management strategies under significant uncertainty. We examine the nature of the technology, how the government sought to manage its evolution and utilization, and evaluate

the results of those efforts through the lens of achieving the following three goals:

- 1: Preserve U.S. technological leadership**
Underwrite continued military and intelligence capability superiority
- 2: Support peaceful use of the technology**
Help civil/commercial sectors reap benefits of tech. applications
- 3: Manage catastrophic risks**
Prevent and mitigate dangers from accidental and adversarial use

These goals are not always necessarily in alignment and may conflict. Nevertheless, they capture what the national security community should seek. Finally, we provide policy recommendations^c for how the United States national security community should respond to the opportunities and threats presented by AI, including achieving the three goals.

C This analysis was specifically developed on behalf of Jason Matheny, Director of the Intelligence Advanced Research Projects Activity, who requested a "whole of government" approach to findings and recommendations.

Part 1: The Transformative Potential of Artificial Intelligence

In a modified version of the framework laid out in the Office of Net Assessment AI Summer Study,¹¹ we analyze AI's potentially transformative implications across three dimensions: **military superiority**, **information superiority**, and **economic superiority**.

In these we take note of existing technological capabilities and trends and then examine how further improvements in capability and/or decreases in cost might transform national security. We then lay out specific hypotheses for how these trends might interact to produce a transformative scenario.

As an overarching frame, consider this statement from the 2016 White House report on AI: "AI's central economic effect in the short term will be the automation of tasks that could not be automated before."¹² The same is true for military affairs. AI will make military and intelligence activities that currently require the efforts of many people achievable with fewer people or without people.

Implications for Military Superiority

In this section, we examine trends in AI that are likely to impact the future of military superiority. In particular, we analyze how future progress in AI technology will affect capabilities in robotics & autonomy and cybersecurity. After establishing key trends and themes, we conclude by laying out scenarios where these capability improvements would result in transformative implications for the future of military superiority.

Robotics & Autonomy

One of the prime uses of robots is to do things that are too dangerous for humans, and fighting wars is about as dangerous as it gets.

–Pedro Domingos, *The Master Algorithm* ¹³

Autonomous systems have been used in warfare since at least WWII. Delegation of human control to such systems has increased alongside improvement in enabling technologies.

Very simple systems that use a sensor to trigger an automatic military action, such as land mines, have been in use for centuries. In recent decades, computers have since taken on more responsibility in the use of force.¹⁴ With the invention of the Norden Bombsight^D and V-1 buzz bomb in World War II, computer systems were first linked to sensors involved in the dynamic control and application of lethal force.¹⁵ So-called “fire-and-forget” missiles, for example, allow the onboard sensors and computer to guide a missile to its target without further operator communications following initial target selection and fire authorization.¹⁶ The U.S. military has developed directives restricting development and use of systems with certain autonomous capabilities. Chief among these is that humans are to be always “in the loop” and directly make the decisions for all uses of lethal force.¹⁷ ^E

The market size for both commercial and military robotics is increasing exponentially, and unit prices are falling significantly.

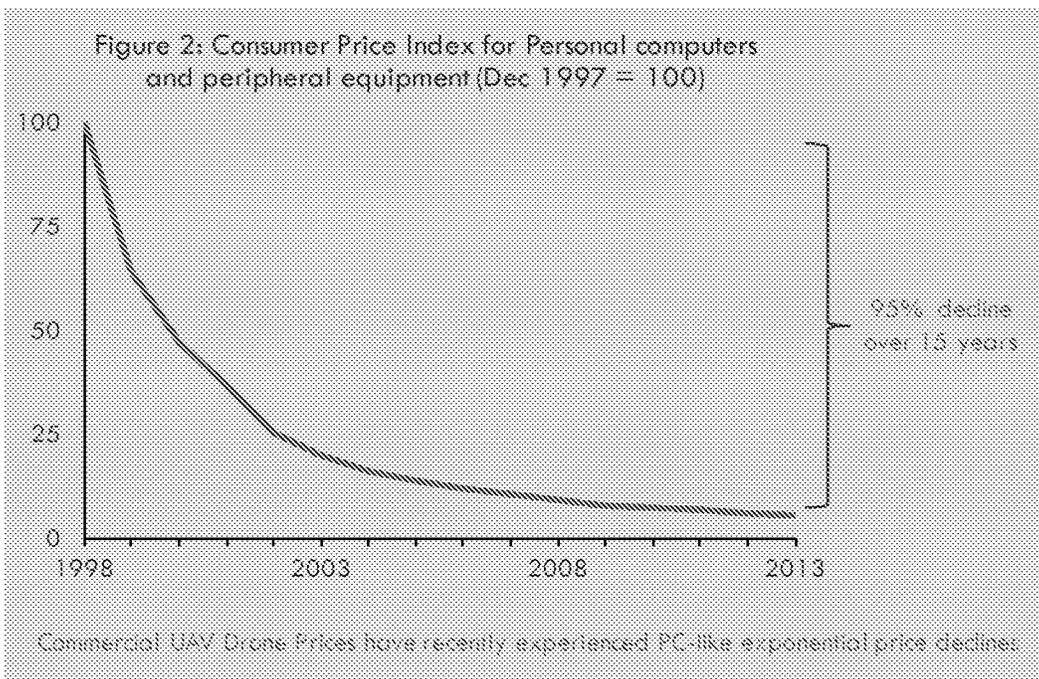
According to the Boston Consulting Group, between 2000 and 2015, the worldwide spending on military robotics (narrowly defined as only

D As Allied bombers approached their targets, the pilot and bombardier would turn over control of the aircraft to the computerized autopilot which would autonomously fly the aircraft to the optimal location based on wind speed and other automatically measured conditions, and thereafter release the bomb payloads over the target. Although the onboard bombardier programmed the autopilot, it was the latter’s computer that determined—based on sensor data—when and where to open the bomb bay doors and release the bombs.

E Interestingly, the directive explicitly “Does not apply to autonomous or semi-autonomous cyberspace systems for cyberspace operations.” As physical assets are increasingly connected to the internet, the ability to use cyber capabilities for the delivery of lethal force grows e.g. hacking a plane’s avionics system and causing it to crash.

unmanned vehicles) tripled from \$2.4 billion to \$7.5 billion and is expected to more than double again to \$16.5 billion by the year 2025.¹⁸ Even this rapid growth may understate the true impact of increased adoption due to falling unit prices and the increasing overlap between commercial and military systems.

One type of robot, the Unmanned Aerial Vehicle,^F otherwise known as a drone, has seen major commercial price declines over just the past few years.¹⁹ Bill Gates has argued that robotics is poised for the same reinforcing cycle of rapid price declines and adoption growth that personal computers experienced.²⁰ As shown in Figure 2, in the 15 years from 1998 to 2013, the average price of a personal computer fell by 95%.²¹ If a high-quality drone that costs \$1,000 today were available for only \$50 in the future, especially if that drone possessed improved autonomous capabilities, it would transform the cost curve for all sorts of military activity. As Paul Scharre has written, “Ultra-cheap 3D-printed mini-drones could allow the United States to field billions—yes, billions—of tiny, insect-like drones.”²²



^F Not all unmanned systems are autonomous. Some are merely remotely operated. However, the autonomous capabilities of commercial UAVs have increased significantly in recent years.

Expanded use of machine learning, combined with market growth and price declines, will greatly expand robotic systems' impact on national security.

We argue that the use of robotic and autonomous systems in both warfare and the commercial sector is poised to increase dramatically. We concur with Gill Pratt, former DARPA Program Manager and leader of the DARPA Robotics Challenge, who argues that, technological and economic trends are converging to deliver a “Cambrian Explosion” of new robotic systems.²³ The robotic “Cambrian Explosion” is an analogy to the history of life on Earth, specifically the period roughly 500 million years ago in which the pace of evolutionary change, for both diversity and complexity of life forms, increased significantly. Pratt points to several trends, but of particular importance are the improvements in the utilization of machine learning techniques and the ability for these techniques to allow robots to intelligently make decisions based on sensor data. Humans have been able to build self-driving automobiles for as long as they have been able to make automobiles, but they would invariably crash. Only recently has the technology been available to produce autonomous cars that can safely and reliably operate in the real world. The same is true for an incredibly diverse array of robotic systems.

Like the impact of cyber, increased utilization of robotics and autonomous systems will augment the power of both non-state actors and nation states.

The introduction of the cyber domain had benefits for all types of actors. Major states built powerful cyber weapons, conducted extensive cyber-espionage, and enhanced existing military operations with digital networking.

Since cyber capabilities were far cheaper than their non-cyber equivalents,²⁴ smaller states with less powerful militaries also made use of cyber. Ethiopia and many other governments, for example, used cyber tools to monitor political dissidents abroad.²⁵ Likewise, hostile non-state actors, including both criminals and terrorists, have made effective use of cyber tools for geographically dispersed activities that would be much more

difficult to execute in the physical domain.²⁶ In the near term, the Cambrian Explosion of robotics and autonomy is likely to have similar impacts for power diffusion as the rise of national security operations in the cyber domain did.

In the short term, advances in AI will likely allow more autonomous robotic support to warfighters, and accelerate the shift from manned to unmanned combat missions.

Initially, technological progress will deliver the greatest advantages to large, well-funded, and technologically sophisticated militaries, just as Unmanned Aerial Vehicles (UAVs) and Unmanned Ground Vehicles (UGVs) did in U.S. military operations in Iraq and Afghanistan. As prices fall, states with budget-constrained and less technologically advanced militaries will adopt the technology, as will non-state actors. This pattern is observable today: ISIS is making noteworthy use of remotely-controlled aerial drones in its military operations.²⁷ In the future they or other terrorist groups will likely make increasing use of autonomous vehicles. Though advances in robotics and autonomy will increase the absolute power of all types of actors, the relative power balance may or may not shift away from leading nation states.

The size, weight, and power constraints that currently limit advanced autonomy will eventually be overcome, just as smartphones today deliver what used to be supercomputer performance.

Automobile manufacturers expect to be selling fully autonomous vehicles by the year 2021.²⁸ These cars will have large, expensive, and power-hungry computers onboard, but over time prices will fall, and sizes will shrink. A modern smartphone, which costs \$700 and fits in a pocket, is more powerful than the world's fastest supercomputer from the early 1990s.²⁹ The processors that will power upcoming autonomous vehicles are much, much closer to those of current phones than they are to current supercomputers (which require their own power plants).

Over the medium to long-term, robotic and autonomous systems are likely to match an increasing set of the technological capabilities that have been proven possible by nature.

We especially like this “Cambrian Explosion” biological analogy because biology is full of intelligent autonomous systems. An “existence proof” is when one acquires the knowledge that a specific technology is possible because one observes it in action. For instance, many militaries around the world first learned that precision-guided-missile (PGM) technology was possible when they saw the technology successfully used by the United States military during the Gulf War in 1991. Most militaries could not themselves build PGMs, but suddenly they knew that PGMs were technologically achievable.

Similarly, the natural world of biology can be considered a set of technological existence proofs for robotics and autonomy. Every type of animal, whether insect, fish, bird, or mammal has a suite of sensors, tools for interacting with its environment, and a high-speed data processing and decision-making center. Humans do not yet know how to replicate all the technologies and capabilities of nature, but the fact that these capabilities exist in nature proves that they are indeed possible. Consider the common city pigeon: the pigeon has significantly more flight maneuverability, better sensors, faster data processing capability, and greater power efficiency than any comparable drone. The combination of a pigeon’s brain, eyes, and ears is also superior at navigation and collision avoidance than any autonomous car, despite requiring less than one watt of power to function.³⁰ Humans do not know what the ultimate technological performance limit for autonomous robotics is, but the ultimate limit can be no lower than the very high level of performance that nature has proven possible with the pigeon, the goose, the mouse, the mosquito, the dolphin, etc.

Over the long term, these capabilities will transform military power and warfare.

Autonomous robots are unlikely to match all the technology and performance of nature in the next decade or two. Nevertheless, the robotic systems that are possible will be capable enough to transform military

power. Human-developed technology can do things that nature's engineering approach cannot, such as adapting capabilities from one system to another. A hypothetical robotic "bird" could also possess night vision or a needle for injecting venom. Even the most advanced robots are far from achieving this combination of capabilities and performance today, but given that these technologies exist in nature, there is no reason in principle why advanced military robots could not possess these capabilities. Robots can also make use of technologies that do not exist in nature, such as radar, explosives, ballistics, and digital telecommunications.

Cybersecurity & Cyberwar

Top U.S. national security officials believe that AI and machine learning will have transformative implications for cybersecurity and cyberwar.

In response to a question from the authors of this report, Admiral Mike Rogers, the Director of the National Security Agency and Commander of U.S. Cyber command, said "Artificial Intelligence and machine learning—I would argue—is foundational to the future of cybersecurity [...] We have got to work our way through how we're going to deal with this. It is not the *if*, it's only the *when* to me."³¹ We agree.

As with all automation, AI and machine learning will decrease the number of humans needed to perform specific tasks in the cyber domain.

The advent of cyber tools dramatically increased the productivity of individuals engaged in espionage. As Bruce Schneier of Harvard University points out, "the exceptionally paranoid East German government had 102,000 Stasi surveilling a population of 17 million: that's one spy for every 166 citizens."³² By comparison, using digital surveillance, governments and corporations can surveil the digital activities of billions of individuals with only a few thousand staff. Increased adoption of AI in the cyber domain will further augment the power of those individuals operating and supervising these tools and systems.

AI will be useful in bolstering cyber defense, since probing for weaknesses and monitoring systems can be enhanced with intelligent automation.

DARPA is currently working on systems that will bring AI into cyber defense. These include techniques for automatically detecting software code vulnerabilities prior to release and using machine learning to detect deviations from normal network activity.³³ Cyber defense is currently quite labor intensive and skilled cyber labor is in short supply. Additionally, AI will enable new paradigms for cyber defense. Most cyber defense systems today are based on *a priori* knowledge assumptions, in which the defender has optimized their system to address known threats, and is less well protected against unknown threats. AI and machine learning might allow systems to not only learn from past vulnerabilities, but also observe anomalous behavior to detect and respond to unknown threats.³⁴

However, the same logic suggests AI advances will enable improvements in cyber offense.

For cybersecurity, advances in AI pose an important challenge in that attack approaches today that are labor-and-talent constrained may—in a future with highly-capable AI—be merely capital-constrained. The most challenging type of cyberattack, for most organizations and individuals to deal with, is the Advanced Persistent Threat (APT). With an APT, the attacker is actively hunting for weaknesses in the defender’s security and patiently waiting for the defender to make a mistake. This is a labor-intensive activity and generally requires highly-skilled labor. With the growing capabilities in machine learning and AI, this “hunting for weaknesses” activity will be automated to a degree that is not currently possible and perhaps occur faster than human-controlled defenses could effectively operate. This would mean that future APTs will be capital-constrained rather than labor-and-talent constrained. In other words, any actor with the financial resources to buy an AI APT system could gain access to tremendous offensive cyber capability, even if that actor is very ignorant of internet security technology. Given that the cost of replicating software can be nearly zero, that may hardly present any constraint at all.

Near term, bringing AI technology applications into the cyber domain will benefit powerful nation-state actors. Over the long term, power balance outcomes are unclear, as is the long-term balance between cyber offense and defense.

To some extent there is already a market for the services of skilled cyber criminals. However, there are many people who refuse to serve as hitmen but are willing to sell guns. We should therefore be concerned about AI advances making cyber “guns” much more capable and autonomous. Developing cyber weapons includes the difficult steps of weaponizing undetected vulnerabilities, customizing software to have the desired effects, and engineering the weapons to avoid defenses. As AI-related cyber techniques improve, a greater and greater portion of the operations may be amenable to automation.³⁵ If true, the Stuxnet of the future may not require tens or hundreds of millions of dollars to develop and launch but merely hundreds or thousands of dollars as the steps requiring high-skill human cyber operator customization are reduced or eliminated through AI. At that point, most software can be reproduced at near-zero marginal cost.

Applications of AI therefore have exceptional abilities to strengthen the cyber capabilities of powerful nation-states, small states, and non-state actors. There is no obvious, stable outcome in terms of state vs. non-state power or offense vs. defense cyber advantage. It will depend on the balance of research and development investments by all actors, the pace of technological process, and underlying limitations in economics and technology.

Potential Transformative Scenarios

The trends and themes described above could combine to create a military power landscape very different from what exists today. Below, we provide ten scenarios by which the growing capabilities of AI could transform military power. These are not meant as firm predictions. Rather, they are intended to be provocative and to demonstrate how wide the range of possible outcomes is—given current trends. Moreover, they are not mutually exclusive alternatives. More than one or several could potentially happen simultaneously.

1. Lethal autonomous weapons form the bulk of military forces.

For nearly eight decades, as automatic and autonomous systems have become more capable, militaries have become more willing to delegate authority to them.³⁶ Given that an AI-based pilot running on a \$35 computer has already demonstrated the ability to beat a U.S. Air Force-trained fighter pilot in a combat simulator,³⁷ many actors will face increasing temptation to delegate greater levels of authority to a machine, or else face defeat. The Russian Military Industrial Committee has approved an aggressive plan that would have 30% of Russian combat power consist of entirely remote-controlled and autonomous robotic platforms by 2030.³⁸ ³⁹ G Other countries facing demographic and security challenges are likely to set similar goals. For example, Japan and Israel, which have highly advanced technology sectors and unique demographic challenges, may find lethal autonomous weapons especially appealing. The United States Department of Defense has enacted restrictions on the use of autonomous and semi-autonomous systems wielding lethal force. Other countries and non-state actors may not exercise such self-restraint.

2. Disruptive swarming technologies render some military platforms obsolete.

As of 2013, The United States possessed 14,776 military aircraft, some of which cost more than \$100 million per unit.⁴⁰ A high-quality quadcopter UAV currently costs roughly \$1,000, meaning that for the price of a single high-end aircraft, a military could acquire one million

G The clear majority of the publicly announced systems that Russia is developing are remotely operated or only partially autonomous.

drones. If the robotics market sustains current price decline trends, in the future that figure might become closer to one billion. In such a scenario, drones would be even cheaper than some ballistic munitions are today, e.g. ~\$150 per 155mm shell.

Commercial drones currently face significant range and payload limitations but become cheaper and more capable with each passing year. Imagine a low-cost drone with the range of a Canada Goose, a bird which can cover 1,500 miles in under 24 hours at an average speed of 60 miles per hour.⁴¹ How would an aircraft carrier battlegroup respond to an attack from millions of aerial kamikaze explosive drones? Some of the major platforms and strategies upon which U.S. national security currently relies might be rendered obsolete.

3. Robotic assassination is common and difficult to attribute.

The low-cost of cyber has given offense the edge for targeted digital attacks. Widespread availability of low-cost, highly-capable, lethal, and autonomous robots could make targeted assassination more widespread and more difficult to attribute. A small, autonomous robot could infiltrate a target's home, inject the target with a lethal dose of poison, and leave undetected. Alternatively, automatic sniping robots could assassinate targets from afar.

4. Mobile-robotic-IEDs give low-cost, PGM-like capabilities to terrorists.

Improvised Explosive Devices (IED) posed a significant challenge to U.S. forces in Iraq because they were low-cost, easily manufactured, and could cause significant damage. As commercial robotic and autonomous vehicle technology becomes widespread, some groups will leverage this to make more advanced IED technology. For example, the technological capability to rapidly deliver explosives to a precise target from many miles away is currently restricted to powerful nation states who sometimes spend millions of dollars for each Precision Guided Munition (PGM). If long distance package delivery by drone becomes a reality, the cost of precisely delivering explosives from afar would fall from millions of dollars to thousands or even hundreds. Similarly, self-driving cars could make suicide car bombs more frequent and devastating since they no longer require a suicidal driver.

5. Military power grows disconnected from population size and economic strength.

The CIA World Factbook still counts the number of combat-age males in a country as one of the elements for determining a country's military potential. In the future, however, even countries with small, elderly, and declining populations may be able to use robotics and autonomy to possess robotic "manpower" far beyond their human population size. Consider South Korea: after Google DeepMind's AlphaGo system defeated the South Korean Go Champion Lee Sedoul, South Korea's government announced that it would spend nearly \$1 billion over the next five years on AI research and development.⁴² Including government in-kind contributions and reprogrammed funds, South Korea's annual AI R&D spending may reach \$1 billion within the next year or two.⁴³

If South Korea does reach such a figure, it would match the 2015 AI R&D budget of the United States, a country with a nearly fifteen-fold larger economy. Though such a scenario is speculative, it is possible that a technologically advanced country with a smaller population, such as South Korea, could build a significant advantage in AI based military systems and thereby field greater numbers of more capable robotic "warfighters" than some more populous adversaries.

6. Cyberweapons are frequently used to kill.

The linkage of digital and physical systems will expand the number of possibilities for killing with cyberweapons. A self-driving car could be hacked and made to crash on the highway.⁴⁴ While lethal cyberattacks are possible without AI, AI will change the situation in two ways: First, capabilities might make it possible or even easy to execute such attacks at scale and possible for well-funded actors with limited cyber expertise to perpetrate. Second, the growth of AI applications will help bring more hackable things into the physical world.

7. Most actors in cyber space will have no choice but to enable relatively high levels of autonomy, or else risk being outcompeted by “machine-speed” adversaries.

There are some sectors of military power where high levels of autonomy are a pre-requisite for success. Missile defense, for instance, cannot always wait for human operators to individually target and approve the launching of each counter-missile. Similarly, AI cyber defense will have to be given high levels of autonomy to respond to high speed cyberattacks or else risk being overwhelmed. In recent years, some attackers of government networks have attempted to maintain their presence even after discovery, actively fighting with the United States for control.⁴⁵ Machine-speed AI defenders or attackers would likely have an advantage in this sort of virtual “hand to hand combat”⁴⁶ since they operate at gigahertz speed. As with missile defense, those defenders unwilling to turn over control to AI, will simply lose out to attackers who are more willing to do so.⁴⁷

8. Unexpected interactions of autonomous systems cause occasional “flash crashes.”

Autonomous systems can make decisions incredibly rapidly, much faster than humans can monitor and restrain them without the aid of machines. Because of autonomous systems’ high speed, unexpected interactions and errors can spiral out of control rapidly. One ominous example is the stock market “Flash Crash” of May 2010, which the U.S. Securities and Exchange Commission reported was enabled and exacerbated by use of autonomous financial trading systems.⁴⁸ In the Flash Crash, one trillion dollars of stock market value was wiped out within minutes because of unintended machine interactions (emergent effects). One must consider the cybersecurity or autonomous vehicle equivalent of a flash crash.

The system verification and validation process for autonomous systems that leverage machine learning is still in its relative infancy, and the flash crash suggests that even systems which perform better than humans for 99%+ of their operations may occasionally have catastrophic, unexpected failures. This is especially worrisome given the adversarial nature of warfare and espionage. Pedestrians and

other drivers want autonomous vehicles to be successful and safe. The military adversaries of robotic systems, like those in financial markets, will be less kind.

9. Involving machine learning in military systems will create new types of vulnerabilities and new types of cyberattacks that target the training data of machine learning systems.

Since machine learning systems rely upon high-quality datasets to train their algorithms, injecting so-called “poisoned” data into those training sets could lead AI systems to perform in undesired ways. For instance, researchers have proven that an adversary with access to a deep neural network image classifier’s training data, could expose it to data that the classifier would systematically miscategorize.⁴⁹ One could imagine a more extreme data poisoning attack that would lead a sensor to falsely recognize friend as foe or foe as not present at all. Such manipulations are possible with existing cyber systems, but as we increase use of machine learning, the nature of the attack will change. Given rising levels of autonomy, the impact of an attack might also increase significantly.

10. Theft and replication of military and intelligence AI systems will result in AI cyberweapons falling in the wrong hands.

In aerospace or other technologies, stealing the blueprints for a weapon does not actually give the thief access to the weapon or even a guaranteed ability to develop one. As one of us wrote in a previous article⁵⁰ for Vox:

When China stole the blueprints and R&D data for America’s F-35 fighter aircraft, for example, it likely shaved years off the development timeline for a Chinese F-35 competitor. But, China didn’t actually acquire a modern jet fighter or the immediate capability to make one. That’s because aerospace manufacturing is incredibly difficult, and China can’t yet match US competence in this area.⁵¹ But when a country steals the code for a cyberweapons, it has stolen not only the blueprints, but also the tool itself — and it can reproduce that tool at near zero-marginal cost.

In the cyber domain, groups have reportedly stolen access to U.S. government cyber tools and used them to infect hundreds of thousands of computers for criminal purposes.⁵² Cyber tools utilizing AI may also share this property, and the result—especially if offense-dominance remains the case—would be that highly-destructive AI cyberweapons could be widely available and difficult to control.

Hacking of robotic systems might also pose a serious risk. Paul Scharre has pointed out that autonomous weapons “pose a novel risk of mass fratricide, with large numbers of weapons turning on friendly forces [...] This could be because of hacking, enemy behavioral manipulation, unexpected interactions with the environment, or simple malfunctions or software errors.”⁵³

Implications for Information Superiority

If World War III will be over in seconds, as one side takes control of the other's systems, we'd better have the smarter, faster, more resilient network.

—Pedro Domingos, *The Master Algorithm*⁵⁴

In this section, we examine trends in Artificial Intelligence that are likely to impact the future of information superiority. In particular, we analyze how future progress in AI technology will affect capabilities of intelligence collection and analysis of data, and the creation of data and media. We believe the latter set of capabilities will have significant impacts on the future of propaganda, strategic deception, and social engineering. After establishing the key trends and themes, we conclude by laying out scenarios where these capability improvements would result in transformative implications for the future of information superiority.

Collection & Analysis of Data

U.S. Intelligence agencies are awash in far more potentially useful raw intelligence data than they can analyze.

According to a study by EMC Corporation, the amount of data stored on Earth doubles every two years, meaning that as much data will be created over the next 24 months as over the entire prior history of humanity.⁵⁵ Most of this new data is unstructured sensor or text data and stored across unintegrated databases. For intelligence agencies, this creates both an opportunity and a challenge: there is more data to analyze and draw useful conclusions from, but finding the needle in so much hay is tougher. The Intelligence Agencies of the United States each day collect more raw intelligence data than their entire workforce could effectively analyze in their combined lifetimes.⁵⁶

Computer-assisted intelligence analysis, leveraging machine learning, will soon deliver remarkable capabilities, such as photographing and analyzing the entire Earth's surface every day.

Analysts must prioritize and triage which collected information to analyze, and they leverage computer search and databases to increase the amount of information that they can manage. Some datasets that were previously only analyzable by human staff, such as photos, are newly amenable to automated analysis based on machine learning. In 2015, image recognition systems developed by Microsoft and Google outperformed human competitors at the ImageNet challenge.⁵⁷ These machine learning-based techniques are already being adapted by U.S. intelligence agencies to automatically analyze satellite reconnaissance photographs,⁵⁸ which may make it possible for the United States to image and automatically analyze every square meter of the Earth's surface every single day.⁵⁹ Since machine learning is useful in processing most types of unstructured sensor data, applications will likely extend to most types of sensor-based intelligence, such as Signals Intelligence (SIGINT) and Electronic Intelligence (ELINT). Machine learning-based analysis is also useful for analyzing and deriving meaning from unstructured text.

Creation of Data and Media

AI applications can be used not only to analyze data, but also to produce it, including automatically-generated photographs, video, and text.

Researchers have demonstrated rapid progress in the ability of AI to generate content. Existing AI-related capabilities include but are not limited to the following:

- Realistically changing the facial expressions and speech-related mouth movements of an individual on video in real-time, using only a retail-consumer webcam^H 60
- Generating a realistic-sounding, synthetic voice recording of any individual for whom there is sufficient training data, so-called “Photoshop for Audio”^I 61
- Producing realistic, fake images based only on a text description^J 62
- Producing written news articles based on structured data such as political polls, election results, financial reports and sports game statistics^K 63
- Creating a 3D representation of an object (such as a face) based on one or more 2D images^L 64
- Automatically producing realistic sound effects to accompany a silent video^M 65

In the near future, it will be possible even for amateurs to generate photo-realistic HD video, audio, and document forgeries—at scale.

Today, many of these AI-forgery capabilities are real enough that they can sometimes fool the untrained eye and ear. In the near future, they will be good enough to fool at least some types of forensic analysis. Moreover, these tools will be available not only to advanced computer scientists, but to

H See <https://www.youtube.com/watch?v=ohmajJTcpNk> for a demonstration of this capability.

I See <https://lyrebird.ai/demo> for a demonstration of this capability.

J See <https://www.youtube.com/watch?v=qX8A1RsFmTA> for a demonstration of this capability.

K See <https://www.youtube.com/watch?v=OFW99AQmMc8> for a demonstration of this capability.

anyone, unless the government effectively restricts their availability.^L When tools for producing fake-video at higher quality than today's Hollywood Computer-Generated Imagery (CGI) are available to untrained amateurs, these forgeries might comprise a large part of the information ecosystem.

The existence of widespread AI forgery capabilities will erode social trust, as previously reliable evidence becomes highly uncertain.

Since the invention of the photographic camera in the mid-1900s, the technology for capturing highly reliable evidence has been significantly cheaper and more available than the technology for producing convincing forgeries. Today, every individual with a smartphone can record HD video of events to which they bear witness. Moreover, most people can today also generally (though not always) tell when a video they are looking at is fake. Currently, producing high-quality fake video is extremely expensive. Hollywood movies spend tens of millions of dollars to produce believable CGI, and still many fans occasionally complain that the images look fake.⁶⁶ This will change. As one of us wrote in an article for WIRED,^M

Today, when people see a video of a politician taking a bribe, a soldier perpetrating a war crime, or a celebrity starring in a sex tape, viewers can safely assume that the depicted events have actually occurred, provided, of course, that the video is of a certain quality and not obviously edited.

But that world of truth—where seeing is believing—is about to be upended by artificial intelligence technologies [...]

When tools for producing fake video perform at higher quality than today's CGI and are simultaneously available to untrained amateurs, these forgeries might comprise a large part of the information ecosystem. The growth in this technology will transform the meaning of evidence and truth in domains across journalism,

L Governments do attempt to restrict some types of forgery related technology, with mixed results. Most photocopiers automatically detect attempts to copy or scan money and refuse the request. In 2015, France passed an anti-anorexia law that restricts the use of image-editing software in fashion magazines.

M Allen, Greg. "Artificial Intelligence Will Make Forging Anything Entirely Too Easy." *Wired*. June 30, 2017. Accessed July 06, 2017. <https://www.wired.com/story/ai-will-make-forging-anything-entirely-too-easy/>

government communications, testimony in criminal justice, and, of course, national security.

A future where fakes are cheap, widely available, and indistinguishable from reality would reshape the relationship of individuals to truth and evidence. This will have profound implications for domains across journalism, government communications, testimony in criminal justice, and of course national security. Today, when someone sees a leaked video of a terrorist perpetrating a massacre or a politician admitting to taking a bribe, (assuming the video is of a certain quality and not obviously edited), the person can safely assume that the depicted events actually occurred. In the future, people will be constantly confronted with realistic-looking fakes.

We will struggle to know what to trust. Using cryptography and secure communication channels, it may still be possible to, in some circumstances, prove the authenticity of evidence. But, the “seeing is believing” aspect of evidence that dominates today—one where the human eye or ear is almost always good enough—will be compromised.

Potential Transformative Scenarios

As the above analysis indicates, AI is useful both for using data to arrive at conclusions and for generating data to induce false conclusions. In other words, AI can assist intelligence agencies in determining the truth, but it also makes it easier for adversaries to lie convincingly. Which of these two features predominates is likely to shift back and forth with specific technological advances. Below, we outline six possible scenarios for how AI capabilities could transform the future of information superiority. We acknowledge that some of these are mutually exclusive. Our aim is to show how wide the range of possible transformative outcomes is, not to flawlessly forecast the future.

1. Supercharged surveillance brings about the end of guerilla warfare.

There is a plausible “winner-take-all” aspect to the future of AI and surveillance, especially for nation-states. Terrorist and guerrilla

organizations will struggle to plan and execute operations without leaving dots that nation-states can collect and connect. Imagine, for instance, if the United States could have placed low-cost digital cameras with facial recognition and the robotic equivalent of a bomb-sniffing dog's nose⁶⁷ every 200 yards on every road in Iraq during the height of U.S. operations. If robotics and data processing continue their current exponential price declines and capability growth, this sort of AI-enhanced threat detection system might be possible. If it did exist, guerilla warfare and insurgency as we know it today might be impossible.

2. A country with a significant advantage in AI-based intelligence analysis achieves decisive strategic advantage decision-making and shaping.

Over the longer term, AI offers the potential to effectively fuse and integrate the analysis of many different types of sensor data sources into a more unified source of decision support. The Office of Net Assessment Summer Study astutely compared the potential of AI intelligence support to the advantage that the United Kingdom and its allies possessed during World War II once they had decrypted the Axis Enigma and Purple codes.⁶⁸

3. Propaganda for authoritarian and illiberal regimes increasingly becomes indistinguishable from the truth.

Given the ease of producing forgeries using AI, regimes that control official media will be able to produce high quality forgeries to shape public perceptions to a degree even greater than today. Supposedly “leaked” videos could be produced of hostile foreign leaders shouting offensive phrases or ordering atrocities. Though forged media will also be produced against authoritarian regimes, state control of media and social media censorship might limit its ability to be disseminated.

4. Democratic and free press difficulty with fake news gets dramatically worse.

The primary problem with fake news today is that it fools individual

citizens and voters. In the future, even high-quality journalist institutions and governments will face persistent difficulty in separating fake news from reality. Because of a flood of high-quality forgeries, even the best news organizations will sometimes report hoaxes as real and fail to report real news because they are tricked into believing that it is fake.

5. Command and Control organizations face persistent social engineering threats.

Widely available AI-generated forgeries will pose a challenge for Command and Control organizations. Those giving and receiving orders will struggle to know which communications (written, video, audio) are authentic. Social engineering hacks,^N which are analogous to digital hacking but target people instead of computers, might be a much greater problem in the future. Allowing an individual in a video or audio phone call to assume the likeness and voice of someone they are impersonating adds another significant layer of difficulty to validating communications. One can imagine an adversary impersonating a military or intelligence officer and ordering the sharing of sensitive information or taking some action that would expose forces to vulnerability. AI could be used to produce counterfeit versions of DOD Directives and statements of policy and to disseminate them widely across the internet. Adversaries of a military could use these technologies to produce large quantities of forged evidence purporting to show that the military has engaged in war crimes.

6. Combined with cyberattacks and social media bot networks, AI-enabled forged media threatens the stability of an economy or government regime.

On April 23, 2013, hackers took control of the Associated Press' official Twitter account and tweeted "BREAKING: Two Explosions in the White House and Barack Obama is injured" to the account's nearly two million followers.⁶⁹ In the two minutes following the tweet, the U.S. stock market lost nearly \$136 billion in value until the hack was revealed.⁷⁰ With AI-enabled forgery, one could imagine a future,

^N A simple example is when a criminal calls a person's credit card company (from a masked phone number) and persuades the human operator to add the criminal to the account.

more devastating hack: Hackers would take control of an official news organization website or social media account being used to spread not only false text, but also false video and audio. A network of social media bots could then be used to spread the fake messaging rapidly and influence a broad number of individuals. Exactly this sort of social media botnet influencing approach was reportedly used by Russia in its attempt to influence the outcome of the 2016 U.S. presidential election.⁷¹

To some extent this problem is not new. For instance, in 2014, some of the images, circulated widely on social media, that claimed to depict Israel's airstrikes on Gaza in 2014 were photographs of the more extensive violence from conflicts in Syria and Iraq.⁷² However, if forged evidence were sufficiently compelling and effectively disseminated, it might result in stock market crashes, riots, or worse. One way this might be executed by an adversary would be to acquire thousands of real (and sensitive) documents through cyber-espionage and then leak the real documents alongside a few well executed forgeries which could then be supported by "leaked" forged audio and video. Even if the government offered widespread denials and produced contradicting evidence, still it would struggle to squash the false understanding in a population that such an operation could bring about. The government would also face major difficulty in limiting and remediating the potentially significant consequences of that false understanding.

Implications for Economic Superiority

In the same way that a bank without databases can't compete with a bank that has them, a company without machine learning can't keep up with one that uses it [...] It's about as fair as spears against machine guns. Machine learning is a cool new technology, but that's not why businesses embrace it. They embrace it because they have no choice.

—Pedro Domingos, *The Master Algorithm*⁷³

In this section, we examine trends in Artificial Intelligence that are likely to impact the future of economic superiority. In particular, we analyze how future progress in AI technology will affect the speed of technological innovation, and how increases in automation will affect employment. After establishing key trends and themes, we conclude by laying out scenarios where these capability improvements would result in transformative implications for the future of economic superiority.

Innovation Supercharger

Artificial Intelligence might be a uniquely transformative economic technology, since it has the potential to dramatically accelerate the pace of innovation and productivity growth.

Many advancements in the domain of AI have the character of general purpose technologies, meaning that they enhance productivity across a broad swath of different industries. AI applications can do more, however. They can accelerate the pace of inventing and innovation itself. Consider three examples:

- 1. Automation of scientific experiments:** researchers developed a robotic system that can autonomously develop scientific genomic hypotheses, conduct scientific biology experiments to test the hypotheses, and then reach conclusions about the hypothesis that informs the next generation of hypothesis formation.⁷⁴

2. **Synthesizing findings in thousands of scientific papers:** A partnership between the Barrow Neurological Institute and IBM resulted in an AI system that used language processing algorithms to analyze thousands of peer-reviewed research articles related to a neurodegenerative disease and then correctly predicted five previously unknown genes related to the disease.⁷⁵
3. **Automatically generating and optimizing engineering designs:** machine learning algorithms supported by advanced mechanical simulation have proven useful in developing new designs for mechanical equipment, including car engines.⁷⁶

These examples show that developing a leading technological position in conducting AI research will likely deliver benefits to the pace of research and development progress in many fields, including AI. AI applications can therefore act as an “innovation supercharger.”

Automation and Unemployment

The 2016 White House Report on Artificial Intelligence, Automation, and the Economy found that increasing automation will threaten millions of jobs⁷⁷ and that future labor disruptions might be more permanent than previous cases.

Automation has always led to the destruction of jobs. After the invention of the mechanized tractor, for example, agricultural labor in the United States began a permanent decline. Farming work today is performed by only 1% of the American population (3.2 million). In 1920, farming labor comprised 30% of the population (32 million).⁷⁸

What is different today, according to the White House report, is the speed of the economic disruption. Economic theory suggests that the increased productivity through automation should ultimately also decrease prices and provide consumers more disposable income with which to generate demand for other goods, services and the workers that provide them.⁷⁹ This

price effect can be slow, however, especially in comparison to the pace of job loss and the length of time required to retrain displaced workers.

It may be the case, however, that large populations of workers lose their jobs due to automation and thereafter face a dearth of new job opportunities. Former U.S. Treasury Secretary Larry Summers has indicated credence for this view: “This question of technology leading to a reduction in demand for labor is not some hypothetical prospect ... It’s one of the defining trends that has shaped the economy and society for the last 40 years” he said in a June 2017 interview. More worryingly, however, Summers went on to posit the following dire scenario:

“I suspect that if current trends continue, we may have a third of men between the ages of 25 and 54 not working by the end of this half century, because this is a trend that shows no sign of decelerating. And that’s before we have ... seen a single driver replaced [by self-driving vehicles] ..., not a trucker, not a taxicab driver, not a delivery person. ... And yet that is surely something that is en route.”⁸⁰

Notably, the one-third unemployment rate that Summers’ predicts is higher than either the United States or Germany faced at the height of the Great Depression.⁸¹ If Summers’ scenario comes to pass, the political stability and national security consequences could be dire.

One worst case scenario, which is not included in the White House report but is taken seriously by some economists and computer scientists, is that the next wave of automation will leave many workers around the world in the same position that horses faced during the mechanized agriculture and transportation revolutions⁸²—unable to remain economically competitive with machines at any price and unable to acquire new, economically useful skills. Human farm laborers successfully retrained to work in other industries when the need for farm labor declined. Horses could not. In 1900, there were 21 million horses and mules in the United States, mostly for animal labor. By 1960, there were fewer than 3 million.⁸³ If artificial intelligence significantly and permanently reduces demand for human unskilled labor, and if significant portions of the unskilled labor workforce struggle

to retrain for economically valuable skills, the economic and social impacts would be devastating.

If AI does lead to permanent worker displacement, technologically advanced countries may face the “Resource Curse” problem, whereby the owners of productive capital are highly concentrated, and economics and politics become unstable.

The Resource Curse problem refers to a diverse and robust set of economic analyses that show countries where natural resources comprise a large portion of the economy tend to be less developed and more unstable than countries with more diversified economies. For instance, one extensive study of the topic found that “between 1960 and 1990, the per capita incomes of resource-poor countries grew two to three times faster than those of resource-abundant countries.”⁸⁴ The main mechanisms for the Resource Curse (as it applies to natural resource wealth) are summarized below:

- **The composition of extractive industries promotes inequality and poor governance:** Extractive industries, such as mining, are capital-intensive and labor-light relative to their scale in the economy. These characteristics imply that a small number of people reap outsized benefits of resource exports.
- **Redistribution of resource revenues risks government corruption:** By taxing extractive industries, the government raises significant revenues which it can then use to provide public goods such as infrastructure and services. Though potentially beneficial, this allocative model of wealth encourages corruption and weak institutions since those with power will be tempted to allocate capital based on political imperatives rather than in accordance with long term economic goals.
- **Inequality promotes political and civil conflict:** The outsized concentration of national wealth in relatively few areas encourages conflict over who will control those resources rather than collaboration over how to promote sustainable economic growth overall.

For example, Sierra Leone's decades of war prior to 2000 were fueled by conflict over which faction would control the country's diamond mines.

- **Success in the natural resource export sectors harms other industries:** Increased demand for a country's natural resource exports causes pressure on its currency to appreciate. The more valuable domestic currency in turn makes other export sectors—such as manufacturing and agriculture—more expensive and less competitive. Domestic-focused producers are also harmed as the stronger domestic currency makes imports cheaper.

Though there are interesting parallels between the resource curse and how automation might enable consolidation of control over the economy, there are also important differences. Most notably, production and consumption of the natural resources typically associated with the resource curse (e.g. oil) is relatively inelastic, meaning large change in the price of a good might only result in a modest change in production or consumption. Further study is needed on this issue.

Potential Transformative Scenarios

1. **Automation-induced “Resource Curse” plagues technologically developed economies.**

Though speculative, some have argued that Resource Curse mechanisms would operate in a country where the owners of automation capital (in both manufacturing and service sectors) were concentrated among elites and labor was comparatively weak in its bargaining power. To illustrate, consider the trajectory of the first industrial revolution. At the beginning, the productivity of both labor and capital increased significantly, but worker wages remained low, and most of the returns went to the owners of capital. Only by organizing into groups that had economic power (the ability to go on strike and halt production) and political power (the ability to influence the state's regulation and

enforcement behavior) were workers able to secure a greater share of the economic returns of industrialization. In resource curse economies, only a small number of well-compensated workers are required to sustain the main economic drivers, and the non-resource industry workers generally lack economic bargaining power. The owners of capital therefore need only be limited by political concerns, which lead them to redistribute the minimum amount of resource wealth required to establish sustainable political or military governing constituencies. If automation could perform a significant portion of current jobs at higher quality and lower cost, and if the displaced labor population lacked skills and the ability to retrain for any newly created job demand, a similar operative mechanism to the resource curse theory is plausible for heavily automated economies.

If true, advanced economies, including the United States and many of its allies, will face significant future challenges in maintaining good governance and political stability. Increasing instability among OECD countries could result in a wave of illiberalism and corruption among democracies. In the worst case, such a scenario might threaten the US-led system of democratic alliances and U.S. national security.

2. A country with a significant lead in AI-enabled innovation technology develops a self-reinforcing technological and economic edge.

AI's role as innovation-supercharger can deliver a strategic (and perhaps permanent) economic and military advantage to a country that develops a significant lead in exploiting AI applications. Because of this recursive-improvement property, and because AI applications also facilitate the automation of labor, it is possible to imagine a breakaway economic and innovation growth scenario, whereby a country develops a significant lead in developing certain AI applications, which then guarantee it will be the first to discover the next generation of innovations, and so on. In the most extreme scenario, one could imagine a small, technologically advanced country like Singapore developing an accelerating technological edge that facilitates extreme economic growth, far beyond what would normally be expected of a country with only five million people. This may sound implausible, but consider the

fact that in 1900, Great Britain, a country of only 40 million people, came to control an empire with dominion over nearly 25% of the Earth's land and population. Being the first to exploit a technological revolution can have outsized consequences. Likewise, this AI-enabled recursive-improvement scenario might result in one country acquiring radically superior military technology, especially in the domain of cyberweapons, where experiments and simulations can be run at digital speeds.

3. AI-enabled economic sabotage emerges as a new type of weapon.

As described herein the Information Superiority section, the 2015 AP twitter account hack led to major, though extremely brief, implications for the U.S. stock market. A more extreme version of this capability could be harnessed into a generalized economic weapon, intended to crash stock or other trading markets, or to disrupt the major digitally-connected means of production in an economy.⁸⁵ To some extent, this threat exists today due to cyberattacks, but AI capabilities might allow much smaller teams of non-nation state actors to launch such an attack and might also increase the scale of such an attack. In 2001, Enron, a corrupt energy company, deliberately shut down a power plant in California on false pretenses to raise energy prices and generate billions in excess profits. The crisis resulted in waves of blackouts across California.⁸⁶ An economic terrorist or nation-state adversary using AI-enhanced cyberweapons might replicate this sort of attack for either strategic military advantage or even just to make a profit by making calibrated investments ahead of time.

Part 2: Learning from Prior Transformative Technology Cases

Having summarized the mechanisms by which Artificial Intelligence might prove to be a transformative field for military technology, this section will summarize our analysis of prior transformative military technologies—Nuclear, Aerospace, Cyber, and Biotech—and thereafter generate lessons learned that apply to the management of AI technology. Our full analysis of these prior cases is included in the Appendix, but Part 2 will summarize this analysis and the lessons learned that we propose.

Key Technology Management Aspects

Though each of these technology cases were transformational for U.S. national security, they had different underlying scientific and economic conditions, which affected the optimal approach for the U.S. government to manage them. We evaluated each case across five different dimensions:

1. **Destructive potential:** Using the technology, how much destruction can weapons cause? How easy is it to demonstrate the destructive potential? How assured is the destruction?
2. **Cost profile:** What resources, and at what price, are required to develop the technology? What is the marginal cost of weapons production at scale? Does production require large fixed assets?
3. **Complexity Profile:** What types of technical expertise are required to develop the technology? To use it after acquisition? Is this expertise primarily dependent on formal knowledge (e.g. mathematics) or tacit knowledge (e.g. manufacturing excellence)?
4. **Military/Civil dual-use potential:** Does experience with commercial versions of the technology imply easy transitions to the military version? Do companies that produce in one sphere tend to also produce

for the other? Do workers with skills from the commercial sector have relevant skills for the military sector?

5. **Difficulty of espionage and monitoring:** Is it easy for adversaries to monitor the progress of a military development program? Is it easy for developers to hide their development, or portray it as commercially intended? Is the technology easily replicated or reverse-engineered?

Again, detailed justification for our technology management aspects is provided in the Appendix. Our summary of the technology profile for each case is presented in Table 1:

Table 1: Key Technology Aspects

	Destructive potential	Cost profile	Complexity	Military/ Civil dual-use potential	Difficulty of espionage/ monitoring
Nuclear	Destructive power is immense, assured, and easily demonstrated	Dev. required share of GDP for first 5 nuclear states, still expensive now	Dev. of nuclear tech. required advanced scientific and engineering	Nuclear power and medicine both carry significant proliferation risk	Aerospace ISR (1955); signals intel. and radioactive tracing allow decent monitoring
Aerospace	Only in vast quantities can aircraft threaten state's existence; attacks can be defended	In 1945, fighter aircraft were roughly 50 times as expensive as a new civilian car	By WW2, only sophisticated orgs could match state of the art in aerospace tech.	One of the first passenger airlines used reconfigured WW1 bombers	Factories appear similar to other industry and can be concealed
Cyber	Cyber can damage physical infrastructure and steal key info, but less assured	Even terrorists and criminals can afford quite useful capabilities	Low-end attacks require minimal expertise; high-end reserved for states	Commercial IT systems can be used for attacks; similar skills in demand for civil/military	Even sensitive national security systems are routinely infiltrated without detection
Biotech	Natural pandemics have killed tens of millions; bioweapons could also	Equipment is cheap, though expertise can be expensive	Though different now, at first relatively few people had needed expertise	Biopharma and medical industries need similar equipment and expertise as bioweapons	Weaponization facilities difficult to distinguish from commercial
	Low		Moderate		High

Government Technology Management Approach

In what is admittedly (and necessarily) a partial oversimplification, we have classified the U.S. government’s management paradigm for each of the four technologies. Our goal here is to clarify how government viewed the nature of the challenge—especially in its early decades—and characterize what approach they ultimately took to meet it. A more detailed justification of our analysis is provided in the Appendix. The four approaches are summarized in Table 2:

Table 2: Government Technology Management Approach

<p>Nuclear</p>	<p>All-out effort, government-led development and utilization</p>	<ul style="list-style-type: none"> • Extraordinary levels of spending and dedication of national resources to nuclear technology continued for many decades after development • From 1940 to 1996, 11% of total federal government spending was related to nuclear weapons, even with arms control and voluntary restrictions • Initially, nuclear technology was treated as classified regardless of origin. Illegal to hold patents on nuclear.
<p>Aerospace</p>	<p>Government-led public private partnership</p>	<ul style="list-style-type: none"> • Heavy government involvement in the aerospace sector with research and development support, acting as an anchor customer, and major regulation • Tech. superiority seen as key to national power; govt. restricted access to aerospace tech. using classification and export restrictions • Despite predominant government role, the U.S. Aircraft industry remained within the American economic model of capitalism and free enterprise
<p>Cyber</p>	<p>Government "seeding and harvesting"</p>	<ul style="list-style-type: none"> • Govt. heavily involved in supportin R&D of tech. underpinnings of computing and internet, but ultimately cedes leadership in most areas to private industry • Govt. retains leadership in the security aspects of computing, using computers in military systems and dev. on cyber attack/defense as early as the 1960s • Govt. initially wants to limit commercial security aspects (e.g. restricting cryptography) but recently sees govt. role in aiding commercial cybersecurity
<p>Biotech</p>	<p>Voluntary restraint</p>	<ul style="list-style-type: none"> • U.S. Govt unilateral ends U.S. bioweapons program in 1969, and ratifies Biological Weapons Convention. However, USSR bioweapons program continues beyond end of Cold War. • U.S./European commercial biotech industries adopt voluntary restrictions on recombinant DNA R&D in 1975 due to ethical and security risk concerns

Government Management Approach “Scorecard”

Next, we evaluate the effectiveness of the government’s technology management approach for each of the four cases. Our evaluation is based upon our assessment of the government’s performance in meeting three key goal:

- 1: Preserve U.S. technological leadership**
Underwrite continued military and intelligence capability superiority
- 2: Support peaceful use of the technology**
Help civil/commercial sectors reap benefits of tech. applications
- 3: Manage catastrophic risks**
Prevent and mitigate dangers from accidental and adversarial use

Our detailed justifications for the scorecard are provided in the Appendix. Our findings are summarized in Table 3.

Table 3: Government Technology Management Approach Scorecard

	1: Preserve U.S. technological leadership	2: Support peaceful use of the technology	3: Manage catastrophic risks
Nuclear	Partial Success U.S. achieved fission and fusion first, and had more nukes and more ways to deliver, but this never gave a usable adv. Espionage hurt U.S. technological edge.	Partial Success Military nuclear tech begets commercial nuclear power and nuclear medicine, but benefits were over-estimated and proliferation risks underestimated	Partial Failure No full accidental detonation, but many nuclear accidents that could have led to detonations; U.S. repeatedly ignores need for safety upgrades/investment
Aerospace	Success Aside from brief periods during WW1 and WW2, U.S. was and is undisputed leader in developing and using military aerospace tech.	Success After WW2, the U.S. emerged as the clear winner in building commercial aircraft for the rapidly growing market in air transportation	Success Main risks are accidental crashes and attacks from superior air forces, both of which the U.S. has responded to effectively
Cyber	Success Though cyber domain is not as amenable to dominance as aerospace, the U.S. clearly has leading tech and capabilities in both cyber and defense	Partial Success U.S. commercial industry leads the world in computing and internet sectors, but U.S. govt. left commercial too vulnerable to criminal and nation-state cyber attacks	Partial Failure While the U.S. developed offensive cyber superiority, the govt. failed for decades to address the asymmetric vulnerability it faced in espionage and attack
Biotech	N/A U.S. voluntarily disbanded bioweapons program, saying deterrent from nukes was sufficient. USSR bioweapons program continued, however.	Success U.S. has largest biotech industry worldwide and the R&D leader in biotech; Favorable government support of R&D and regulations	Partial Success No major bioweapons attacks or accidental releases; most risky research was delayed until risks better understood, BWC helpful but had key failures (USSR)

AI Technology Profile: A Worst-case Scenario?

Comparing the technology profile of AI with the prior technology cases, we find that it has the potential to be a worst-case scenario. Proper precautions might alter this profile in the future, but current trends suggest a uniquely difficult challenge.

Destructive Potential: High

- At a minimum, AI will dramatically augment autonomous weapons and espionage capabilities and will represent a key aspect of future military power.
- Speculative but plausible hypotheses suggest that General AI and especially superintelligence systems pose a potentially existential threat to humanity.⁸⁷ ^o

Cost Profile: Diverse, but potentially low

- Developing cutting-edge capabilities in machine learning and AI can be expensive: many firms are spending billions or hundreds of millions of dollars on R&D.
- However, relatively small teams can leverage open-source code libraries and COTS or rented hardware to develop powerful capabilities for less than \$1 million; leaked copies of AI software might be virtually free.

Complexity Profile: Diverse, but potentially low

- Advancing the state of the art in AI basic research requires world-class talent, of which there is a very limited pool.

^o Nick Bostrom, Elon Musk, Bill Gates, Stephen Hawking, and many others have expressed concern regarding this scenario.

- However, applying existing AI research to specific problems can sometimes be relatively straightforward and accomplished with less elite talent.
- Technical expertise required for converting commercially available AI capabilities into military systems is currently high, but this may decline in the future as AI improves.

Military/Civil Dual-Use Potential: High

- Militaries and commercial businesses are competing for essentially the exact same talent pool and using highly similar hardware infrastructure.
- Some military applications (e.g. autonomous weapons) require additional access to non-AI related expertise to deliver capability.

Difficulty of Espionage and Monitoring: High

- Overlap between commercial and military technology makes it difficult to distinguish which AI activities are potentially hostile.
- Few if any physical markers of AI development exist.
- Total number of actors developing and fielding advanced AI systems will be significantly higher than nuclear or even aerospace.
- Monitors will find it difficult to assess AI aspects of any autonomous weapon system without direct access.

Lessons Learned

Having provided our observations of previous cases, we will now attempt to summarize lessons learned. We recognize that there are vast differences of time, technology, and context between these cases and AI. This is our effort to characterize some lessons which endure nevertheless.

Lesson #1: Radical technology change begets radical government policy ideas

The transformative implications of nuclear weapons technology, combined with the Cold War context, led the U.S. government to consider some extraordinary policy measures, including but not limited to the following:

- **Enacted—Giving one individual sole authority to start nuclear war:** The United States President, as head of government and commander in chief of the military, was invested with supreme authority regarding nuclear weapons⁸⁸
- **Considered—Internationalizing control of nuclear weapons** under the exclusive authority of the United Nations in a collective security arrangement ^{P 89}
- **Enacted—Voluntarily sharing atomic weapons technology** with allies (which occurred) and adversaries including the Soviet Union (which did not)⁹⁰
- **Considered—Atomic annihilation:** Pre-emptive and/or retaliatory atomic annihilation of adversaries, which could have resulted in millions or even billions of deaths^Q

^P This was the so-called Baruch Plan, which the U.S. proposed at the United Nations but abandoned shortly thereafter. To this day there is significant debate over whether the United States offered the Baruch Plan in sincerity.

^Q Senior U.S. military officials, including Lieutenant General Leslie Groves, the director of the Manhattan Project, and General Orvil Anderson, commander of the Air University, publicly argued that the United States should strike the Soviet Union with nuclear weapons to prevent them from acquiring nuclear technology. Respected foreigners including Winston Churchill, John Von Neumann, and Bertrand Russell all advised the United States to do the same. How seriously the United States' senior leadership considered this first strike advice is difficult to say with certainty. Retaliatory nuclear strikes and mutually assured destruction remain the official policy of the United States.

- **Enacted—Voluntarily restricting development in arms control frameworks** to ban certain classes of nuclear weapons and certain classes of nuclear tests

The world has lived with some of these policies for seven decades, so the true extent of their radicalism (at the time they were first considered) is hard to convey. The first example is perhaps the easiest, because it required passage of the Presidential Succession Act of 1947, which laid the foundation for the 25th Amendment to the United States Constitution. Though there were other proximate causes for the 25th Amendment, such as the assassination of President Kennedy, it is only a mild stretch to say that the invention of nuclear weapons was so significant that it led to a change in the United States Constitution.

Though nuclear weapons clearly resulted in the most radical policy proposals, the other cases also led to significant changes. For instance, the Department of Defense ultimately created a full armed service to make use of aerospace technology, the organization now called the U.S. Air Force. Cyber challenges led to the creation of U.S. Cyber Command. These were significant changes, though time has made them familiar.

It remains unclear what the full impact of AI technology on national security will be, and how fast it will arrive. So far, we have argued that it is highly likely to be a transformative military technology. Some, such as Nick Bostrom, believe that the recursive improvement property of AI has the potential to create a superintelligence that might lead to the extinction of the entire human species.⁹¹ If continued rapid progress in AI leads some governments to share Bostrom's view, they may consider policies as truly radical as those considered in the early decades of nuclear weapons. The bigger and more visible the impacts of AI become (and we argue the impacts are likely to be increasingly large and obvious over time) the more policymakers will feel justified in making extreme departures from existing policy.

Lesson #2: Arms races are sometimes unavoidable, but they can be managed

Fears of aerial bombing led to an international treaty banning the use of weaponized aircraft, but voluntary restraint was quickly abandoned and did not stop air war in WWI.

In 1899, diplomats from the world's leading military powers convened in The Hague for a peace conference. One of the more interesting outcomes of the conference was a five-year moratorium on all offensive military uses of aircraft.^R Though the intention was to later make the ban permanent, it was abandoned at the second Hague conference of 1907 once countries saw the irresistible potential of aerial warfare. Accordingly, all the great powers began constructing and planning for the use of aircraft bombers.⁹² In 1910, the combined military air fleets of the European great powers contained 50 airplanes. By 1914, the number reached 700.⁹³ When World War I broke out, the only real limitation on the use of military air power was technology: the primitive airplanes had limited range and bomb-carrying capacity. Still, every European belligerent's capital, save Rome, was bombed from the air.⁹⁴

The applications of AI to warfare and espionage are likely to be as irresistible as aircraft. Preventing expanded military use of AI is likely impossible.

Aerospace technology ultimately became nearly synonymous with military power, and it seems likely that applications of AI will ultimately go the same route. Just as businesses are choosing machine learning because competitively they have no choice, so too will militaries and intelligence agencies feel pressure to expand the use of military AI applications.

Michael Rogers, head of the United States National Security Agency and Cyber Command, agrees: "It is not the 'if.' It's only the 'when' to me. This is coming."⁹⁵ That sense of inevitability derives not only from how useful AI is already proving to be, but also from the belief that current applications have only scratched the surface of what capabilities are likely to come.

Though outright bans of AI applications in the national security sector are unrealistic, the more modest goal of safe and effective technology management must be pursued.

^R At the time, diplomats were primarily concerned with aerial bombardment from motor-driven balloons, but the treaty language was sufficiently broad that it applied to fixed-wing aircraft upon their invention.

The ban of aircraft fell apart, but the United States, its allies, and even its adversaries did develop a framework that sought to limit the risks of aerospace technology. Though many details will remain unclear until the technology is more mature, eventually the United States and other actors will have to develop a regime that limits the risk of military AI technology proliferation.

Lesson #3: Government must both promote and restrain commercial activity

Failure to recognize the inherent dual-use nature of technology can cost lives, as the example of the Rolls-Royce Nene jet engine shows.

After World War II, the United States recognized that facilitating economic growth of the commercial aerospace industry and maintaining military secrecy were often at odds. For instance, the United Kingdom had superior jet engine technology at the end of World War II but faced significant financial challenges. The British engine manufacturers, seeking export revenues, sold 25 of their “commercial” Rolls-Royce Nene Jet Engines to the Soviet Union, which promptly reverse-engineered the Nene engines and designed their MiG-15 fighter around it. The highly effective MiG-15 went on to dominate the skies in the Korean War.⁹⁶ Experiences such as those of the Nene taught the United States that breakthroughs in aerospace technology sometimes had to be kept secret and in the hands of the defense sector. The government expanded its classification and clearance process to include significant numbers of the civilian aerospace workforce, and restrictions were placed on the ability of aerospace companies to sell their technology domestically and especially abroad.

Having the largest and most advanced digital technology industry is an enormous advantage for the United States, but reconciling commercial and national security interests will remain a challenge.

When the United States government set out to regulate the aerospace industry, it did so from a position of extreme strength. The government customer represented a significant majority of total aircraft sales, and the government funded most aerospace R&D. Likewise, as one of us wrote

regarding the nuclear situation: “When nuclear weapons were invented, the best scientists worked for governments, the most advanced technology was possessed exclusively by governments, and governments provided the bulk of scientific research and development funding. That world is so far gone as to be almost unrecognizable.”⁹⁷

The situation for AI will be very different, both because the government is not nearly as large a customer for AI companies and because most of the leading researchers in the field do not work for government. As the White House report on AI points out, the entire U.S. government spent roughly \$1.1 billion on unclassified AI research and development in 2015, while annual U.S. government spending on mathematics and computer science R&D is \$3 billion.⁹⁸ There are multiple Silicon Valley and Chinese companies who each spend more annually on AI R&D than the entire United States government does on R&D for all of mathematics and computer science combined.⁹⁹

To make matters more difficult, the relationship between the U.S. government and the digital technology industry is currently strained, especially in the wake of the Edward Snowden incident and the statements of some political leaders about technology and the tech industry. Google’s DeepMind, seen by many as the world leader in cutting-edge AI research and development, has a strong stance against the military or surveillance use of AI technology. Upon Google’s acquisition of DeepMind, the two organizations agreed that Google would prohibit the use of DeepMind’s technology for military and government surveillance purposes.¹⁰⁰ When Google acquired Boston Dynamics and Schaft—two leading robotics research and development firms that received a significant portion of their funding from DARPA—Google stated that the firms would no longer pursue new military and intelligence contracts.¹⁰¹

Google is in fact more cooperative with the national security community than many leading technology companies. Eric Schmidt, the Executive Chairman of Google’s parent company Alphabet, also serves as Chairman of the Department of Defense Innovation Board. That even Google has significant restrictions on its cooperation with the Department of Defense shows just how tough the current situation is. Though leading digital technology companies are, for the most part, headquartered in the United

States, they are operating in global markets, with customers, suppliers, and partners all over the world.

Whereas the government regulated the nuclear and aerospace industries from the position of most valuable customer and trusted partner, the relationship between the government and some leading AI research institutions is fraught with tension. Fortunately, the same concern applies in the cybersecurity domain, and the United States government has nevertheless been able to build a significant lead in the offensive military and espionage applications of that area. In no small part, this success is due to decades of U.S. government support of the computing and internet industries while they were in their comparative infancy. Nevertheless, the tensions between commercial and government interests in AI will remain a challenge for policymakers, who must effectively support the interests of both constituencies.

Lesson #4: Government must formalize goals for safety and provide resources

In each of the four cases, national security policymakers faced tradeoffs between safety and performance, but the government was more likely to respond appropriately to some risks than to others.

The current Command and Control and safety systems used for each of the four cases took decades to emerge. This in and of itself is not worrisome. What is worrisome is the often very long times between thorough identification of a risk factor and the implementation of a solution. In the case of nuclear weapons, many safety measures that are today considered essential were not implemented for a decade or more after the solution was identified.¹⁰² The institutions responsible for safety repeatedly failed to implement needed safety measures due to cost concerns, biases towards functional reliability (assured destruction of the target) over safety reliability, and bureaucratic infighting.

After surveying the record of nuclear close calls, we agree with former Secretary of Defense Robert McNamara that the absence of a catastrophic nuclear weapons accident can be attributed to luck at least as much as a

reflection of well-designed technological and procedural safeguards.¹⁰³ In an interview with Errol Morris, McNamara stated “I want to say—and this is very important—at the end, we lucked out. It was luck that prevented nuclear war.”¹⁰⁴ The same can be said for the absence of a major cyberattack on United States critical infrastructure. Most cybersecurity experts feel these systems are not actually secure from attack and so the absence of a major attack on one has more to do with the success of U.S. deterrence (and some luck) than it does with appropriate attention and resources being devoted to cyber defense and safety.

Not all communities made this same mistake. The U.S. nuclear submarine community never lost a sub for nuclear technology-related reasons. The aerospace sector likewise managed to achieve continuous and rapid capability improvement while at the same time delivering consistent progress on safety—in both the military and commercial domains.

Across all cases, safety outcomes improved when the government created formal organizations tasked with improving the safety of their respective technology domains and appropriated the needed resources. These resources include not only funding and materials, but talented human capital as well as the authority and access to win bureaucratic fights. The nuclear weapons safety department at Sandia, the Federal Aviation Administration, the Center for Disease Control and Prevention, are all examples of organizations that put safety at the center of their mission, and safety outcomes improved as a result.

As the United States embarks upon the Third Offset and looks to regulate expanded use of AI in the commercial and civilian government sector, it should consider standing up formal research and development organizations tasked with investigating AI safety across the entire government and commercial AI portfolio.

Lesson #5: As a technology changes, so does the United States' national interest

The declining cost and complexity of bioweapons led the United States to change their bioweapons strategy from aggressive development to voluntary restraint.

Based on its own experience, the United States initially believed effective bioweapons were likely to be expensive, complicated and therefore only available to powerful states. During WWII, the United States spent \$400 million in 1945-dollars (\$5.4 billion in 2017-dollars) on bioweapons, roughly one-fifth what was spent on the Manhattan project.¹⁰⁵ Most of this funding went to R&D, since developing mass-production, storage, and effective dispersal methods proved technologically difficult. Biological weapons were seen to have significantly greater destructive capability per cost than chemical or conventional weapons,¹⁰⁶ but bioweapons were perceived as only being available to the United States and other powerful nation-states. The U.S. pursued security through aggressive bioweapons development to underwrite effective deterrence.

By the late 1960s, however, technological progress raised the possibility that bioweapons could become comparable in destructive potential to nuclear weapons and could become available to weaker states that lacked the wealth and technological sophistication of nuclear weapons. Bioweapons had the potential to become “a poor man’s nuke” with an offense-dominant profile. The United States accordingly realized that its primary bioweapons threat was likely to come from unstable small states against which deterrence might not provide sufficient security. In order to shape global norms and arms control frameworks against bioweapons, the United States took the unprecedented step of unilaterally renouncing an entire category of weapons.

As the bioweapons case illustrates, the United States has a strategic interest in shaping the cost, complexity, and offense/defense balance profiles of strategic technologies.

The 1969 National Security Council position paper on biological weapons by Matthew Meselson concluded that “our major interest, is to keep other nations from acquiring them.”¹⁰⁷ Improvements in technology that increased the destructive potential of bioweapons while reducing their cost could not strengthen the United States’ deterrent, which was already well supported by nuclear and conventional armaments, but it might give weak states or terrorists the ability to deter actions by the United States. Equally important, such actors might harm the United States unintentionally through contagious outbreaks. The United States unilaterally disarmed because it determined its primary interest to be in opposing the proliferation of biological weapons.

The broader point is that the United States has a strategic interest in the attributes of dominant military technologies: since the United States has a much larger economy and is much richer than its adversaries, it is better off if the most useful military/intelligence technologies are complex and expensive, so that only it and a minimal number of peers can afford them. The United States is also better off if the performance gap between expensive, state of the art systems and cheaper/older alternatives is very large and would take a long time and considerable resources to close the performance gap.

As the case of stealth aircraft shows, strategic investments can sometimes allow the United States to affect the offense/defense balance in a field and build a long-lasting technological edge.

Consider the case of stealth aircraft. During one 18-day period of the 1973 Yom Kippur War, Soviet-made Surface-to-Air-Missile (SAM) batteries shot down 109 Israeli military aircraft. Since the Israeli Air Force used the most advanced U.S.-made aircraft and electronics, the U.S. military quickly determined that Soviet air defense capabilities were capable of decimating U.S./NATO offensive fighters and bombers.¹⁰⁸ The United States then began a research and development program that ultimately resulted in the creation of stealth aircraft technology. With the introduction of the F-117 nighthawk in 1981, stealth tipped the balance back in favor of the United States’ offensive capabilities. Perhaps most shocking in this story, several of the key underlying scientific breakthroughs that enabled stealth technology originated in 1962 in the Soviet Union with research by Petr Ufimtsev, a physicist at the

Moscow Institute for Radio Engineering. English translations of Ufimtsev's work were not available until 1971.¹⁰⁹ Despite having a nine-year head start, and later an aggressive effort to replicate U.S. advances,¹¹⁰ the Soviet Union never successfully fielded stealth aircraft or developed radars that could reliably detect U.S. stealth aircraft. If the United States had never come across Ufimtsev's breakthrough work, it is possible that the initial invention of stealth aircraft might not have occurred until decades later.

The United States should consider how it can shape the technological profile of military and intelligence applications of AI.

We have argued that the technological profile of AI has the potential to be a worst-case scenario from a technology-management perspective. However, while we view this as the most likely outcome, it is not an inevitable one. There is much the United States could do to make the situation better or worse. As just one example, the Department of Defense Strategic Capabilities Office is currently developing autonomous swarms of aerial micro-drones.¹¹¹ As the United States pursues this sort of military AI research, it should ask whether this is likely to result in a capability that produces a sustainable military advantage for the United States or whether it is likely to accelerate the acquisition of similar capabilities by other countries. Given that aerial micro-drone swarms are also being evaluated by commercial and academic researchers, it may be that whatever advances this program produces can be easily replicated and that the United States is spending money that will ultimately accelerate a technological state of affairs that is worse than the current one. Of course, the program may also result in a breakthrough technological edge that is as decisive and long-lasting as stealth aircraft proved to be.

Part 3: Recommendations for Artificial Intelligence and National Security

Preserving U.S. Technological Leadership

RECOMMENDATION 1:

DoD should conduct AI-focused war games to identify potential disruptive military innovations.

Background: Disruptive innovation theory

Clay Christensen, a professor at Harvard Business School, has characterized two different types of innovation: sustaining and disruptive. Sustaining innovation is where the locus of competition is on “making better products that can be sold for more money to attractive customers.”¹¹² In sustaining innovation competitions, the existing market leaders usually prevail. Disruptive innovation occurs when “the challenge is to commercialize a simpler, more convenient product that sells for less money and appeals to a new or unattractive customer set.”¹¹³ In disruptive innovation, new competitors are likely to beat the incumbents. The disruptive and sustaining innovation pattern has been documented hundreds of times.¹¹⁴

Disruptive innovation theory applies to military domains.

Dr. Gautam Mukunda has observed that these disruptive innovation dynamics also occur in the military sphere,¹¹⁵ and we believe that they are likely to take place in the case of AI. The United States, as the world’s current leading military power, is analogous to the market incumbent: it competes through sustaining innovation, leveraging and improving the extraordinary military capabilities that it already possesses. Other countries and non-state actors—with smaller military budgets and less advanced technology—are analogous to the new competitors. They must consider how to innovate with far fewer existing advantages. The

improvised explosive device (IED) is a classic example of a disruptive military innovation.¹¹⁶ IEDs significantly increased the threat posed by insurgent groups in Iraq, even though they were significantly inferior to U.S. military technology.

Advances in AI will enable new, disruptive innovations for military power.

To the United States, a \$1,000 quadcopter drone might appear completely useless since its performance in nearly every sustaining category is inferior to that of existing military aircraft. To a small-power military or non-state actor, however, the drone might appear as an affordable means for acquiring desirable capabilities that are otherwise too expensive, including reconnaissance or long-range delivery of explosives. As drones and other AI-related capabilities grow in capability and fall in price, the number of disruptive opportunities will increase.

Recommendation: The Department of Defense should fund war-gaming and red-team creative thinking exercises designed to identify how advances in AI might lead to disruptive military innovations that will threaten U.S. military advantages. Specifically, the United States should attempt to identify how AI-enabled capabilities might be useful to different types of actors: powerful nation-states, middle powers, and non-state actors. Once identified, DoD can develop investment strategies to counteract these threats and maintain the United States' military leadership.

RECOMMENDATION 2:

DoD should fund diverse, long-term-focused strategic analyses on AI technology and its implications.

Beyond military war-games, the United States needs prolonged strategic thinking on AI and its implications, like the role the RAND Corporation played in assessing nuclear weapons strategy.

While this study draws heavily upon history for inspiration, there is much about AI technology that is unique and unprecedented. Determining the

correct path forward will require “war-games” not only for military strategy, but also for more complex policy decisions that involve economic, legal, cultural, and technological considerations. Evaluating plausible scenarios, their desirability, and what the optimal response is will require sustained, long-term strategic analyses of AI technology and its implications. We feel that the role played by the RAND Corporation for nuclear strategy during the Cold War is a useful comparison in this regard. RAND’s staff included hundreds of leading scientists, engineers, academics, and former practitioners. These individuals were trusted with sensitive information critical to understanding the nuclear problem, but they remained separate and independent from the government agencies that they advised. They could also serve as an independent voice challenging the conventional wisdom and giving a second opinion before Congress and the executive branch.

Simply put, the U.S. government needs something like a RAND Corporation for AI. The amount of strategic thinking needed on this topic is immense. Of course, the Types of questions that demand substantive evaluation include, but are not limited to the following:

Mandatory IARPA Research Proposal Questions

- What is the first-mover advantage in developing AI technologies? Can fast-followers effectively compete?
- What commercial AI technologies are military “dual-use”?
- What investments in R&D could affect the offense/defense balance for military and intelligence AI applications? And what balance should the United States prefer in various military and intelligence domains
- What AI investments would likely extend the advantages of powerful states, as opposed to weak states or non-state actors?
- How will the growth of artificial intelligence capabilities affect the international balance of economic power?
- When might artificial general intelligence happen? How could the United States know when technology is getting close to general AI? How can the United States effectively plan for or try to affect how it happens?

RECOMMENDATION 3:

DoD should prioritize AI R&D spending on areas that can provide sustainable advantages and mitigate key risks.

AI has the potential to enable many new types of low-cost, high-impact military technologies. Some of these may make DoD's current investments unattractive.

Though the development timeline of many specific AI capabilities is unclear, AI has the potential to be a transformative military technology. Some of these future, AI-enabled capabilities will change the relative attractiveness of procurement and sustainment investments that the Department of Defense plans to make. For instance, the spending justification for some aircraft and naval platforms assumes that they will still have useful military capabilities decades hence. The amount of progress AI technology is poised to make over the next 10-20 years should lead the Department of Defense to revisit those assumptions. If swarms of autonomous, long-range, and low-cost kamikaze drones become available, for example, aircraft carriers as we know them may no longer be relevant to the conflicts of the future. If the United States has a strategic interest in extending the aircraft carrier's military superiority for as long as possible, then it should be investing aggressively in technologies to defend against the threat of drone swarms. Moreover, it should limit spending on any technologies that threaten existing military advantages and that—once demonstrated—will be easily replicated by potential U.S. adversaries. Some military AI technologies that the United States develops may ultimately be more beneficial to its adversaries than to itself and its allies.

However, it may also be the case that by investing to extend the relevance of the United States' existing advantages, it is merely wasting time and resources to fight inevitable technological progress in AI. Doing so may allow the United States' pacing competitors to move first in developing and fielding disruptive technologies and to reduce the amount of time that the United States has in which to develop an effective change in approach.

Determining which of these situations is the case and what is the optimal investment portfolio will be difficult and require constant reassessing as technology evolves. One of the key ways that a country expresses its strategy in peacetime is through choices of buying and researching weapons systems. As the Department of Defense develops its military and intelligence AI research agenda, it should consider what types of strategic outcomes it is seeking and how to avoid counterproductive “races to the bottom.” When evaluating research proposals, IARPA requires applicants to answer a series of questions, which are highly relevant to the sorts of questions that the United States should consider across its AI R&D portfolio:

- What is your estimate for how long it would take a major nation competitor to weaponize this technology after they learn about it?
- What is your estimate for how long it would take a non-state terrorist group with resources like those of Al-Qaeda in the first decade of this century?
- If the technology is leaked, stolen, or copied, would we regret having developed it?
- How could the program be misinterpreted by foreign intelligence? Do you have any suggestions for reducing that risk?
- Can we develop defensive capabilities before offensive ones?
- Can the technology be made less prone to theft, replication and mass production? What intrinsic design features could create barriers to entry?
- What red-team activities could help answer these questions? Whose red team opinion would you particularly respect?

RECOMMENDATION 4

The U.S. defense and intel communities should invest in “counter-AI” capabilities for both offense and defense.

Machine learning-based systems^S have different strengths and weaknesses from traditional software development.

In traditional software development, programs are hand-coded as a long series of sequentially executed instructions. Machine learning is different. In a sense, the computer programs itself by applying an algorithm to a set of training data examples. With this different paradigm come different strengths, including a superior ability to analyze unstructured sensor data, and different weaknesses, including unpredictable behavior in response to data not found in the training data set.

Researchers have only just begun to explore the vulnerabilities and potentially exploitable aspects of machine learning-based systems, so-called “counter-AI.”

Recent research has made progress in identifying what sort of predictable and exploitable vulnerabilities exist within a machine learning system. For example, researchers at the University of Wyoming and Cornell University have demonstrated that adversaries with access to the training data of an image classification machine learning algorithm can apply transformations to any image that will cause the algorithm to predictably misclassify the result.¹⁷ This field of “counter-AI” is in its infancy but will take on increasing importance going forward.

The United States defense and intel communities should seek a leading position in “counter-AI” capabilities.

Machine learning is likely to be incorporated into a large and diverse set of systems over the coming decade. Much as the United States developed a leading capability in offensive cyber operations in the early days of the

^S Note: There are many different paradigms of machine learning. Most recent technological progress has been within the neurology-inspired connectionist paradigm, which includes Deep Learning.

internet, it must now invest to develop capabilities that exploit vulnerabilities in an adversary's machine learning systems. At the same time, it must invest to secure its own systems against these same types of threats. Given the early stage of this research, it is probably best supported through grant-based funding for academic institutions, but eventually research will need to be moved into the classified community.

Supporting Peaceful Use of AI Technology

RECOMMENDATION 5:

DARPA, IARPA, the Office of Naval Research, and the National Science Foundation should be given increased funding for AI-related basic research.

Skilled researchers with expertise in AI are in high demand. Many are leaving academia for significantly higher salaries within the private sector. For instance, in 2015 a single company, Uber, hired 4 faculty and 35 technical staff away from Carnegie Mellon University's Robotics Institute, part of the School of Computer Science, in one swoop. "How to retain people who are worth tens of millions of dollars to other organizations is causing my few remaining hairs to fall out" said department head Andrew Moore.¹¹⁸

This trend runs the risk that talent and information on cutting-edge AI research will be locked up by proprietary enterprises who do not view the national security community as a significant potential customer. Perhaps worse, poaching academic talent runs the risk of eating the AI "seed corn" of instructors who are desperately needed to train a much larger AI workforce and causing the publicly funded research community to fall behind the corporate sector.

To combat these trends, the U.S. government should increase funding for basic AI research at universities to ensure there are many more exciting and well-funded projects for instructors and students alike to collaborate on.

RECOMMENDATION 6:

DoD should release a Request for Information (RFI) on Dual-Use AI Capabilities.

As a General-Purpose Technology, AI will affect many areas of the commercial and military sectors. DoD should seek to determine what AI-capabilities (if any) are inherently military or inherently commercial.

AI is a broad field covering many areas. Some of these areas, such as the incorporation of AI into autonomous weapons, are likely to be inherently military in nature, while others are likely to be either dual-use or inherently commercial. Since the commercial sector also has security needs, these distinctions are not easily resolvable. By releasing an RFI and holding hearings through the Defense Innovation Board, DoD should seek clarity on these distinctions. A greater understanding of which aspects are inherently military or have relatively few civilian uses can then be used to inform future regulations on sensitive AI technology. This would assist the U.S. national security community in threading the needle between preserving military superiority and supporting the peaceful and commercial use of AI technology.

RECOMMENDATION 7:

In-Q-Tel should be given additional resources to promote collaboration between the national security community and the commercial AI industry.

In-Q-Tel is a not-for-profit venture capital firm that invests in technology companies to promote links between these companies and the national security community

In-Q-Tel has a proud history of making venture capital investments in companies that later go on both to make significant contributions in national security and to find success in the private sector. Though its full budget is not public, public estimates of In-Q-Tel's annual budget are in the range of \$120 million.¹¹⁹ This is a drop in the bucket compared to the more

than \$75 billion of annual venture capital funding that occurs in the United States.¹²⁰ Venture capital is an increasingly important source of U.S. R&D funding for groundbreaking technological areas such as AI. Given that U.S. Government Defense and Intelligence spending is more than 3.5% of GDP, In-Q-Tel should comprise more than 0.0016% of annual U.S. venture capital investment.[†]

These investments should go toward firms interested in pursuing both commercial and national security customers.

Most experts in the field believe that leading AI companies are primarily and in many cases exclusively serving commercial, non-defense customers. It is unrealistic to believe that the national security community will be a primary source of revenue for most of these firms. Where possible, the government should seek to ensure that promising startups are also pursuing relevant opportunities in the government space. These venture investments should therefore include companies whose primary market orientation is commercial, so long as they also have the strong potential to contribute to the government mission.

[†] In-Q-Tel is not the only source of venture capital for defense and intelligence-focused firms, since firms can also seek funds from traditional VC sources. Nevertheless, In-Q-Tel's relationship with the classified community means that it plays a critical, unique, and highly beneficial role.

Mitigating Catastrophic Risk

RECOMMENDATION 8:

The National Security Council, the Defense Department, and the State Department should study what AI applications, if any, the United States should seek to restrict with treaties.

While it is highly unlikely that all military and intelligence applications of AI could be restricted via treaty, there may be certain AI applications that powerful states can agree to not develop and deploy.

Arms control treaties are a difficult and imperfect instrument, but they have been helpful in reducing the risks posed by military technologies. Treaties limiting nuclear testing, banning development of certain classes of nuclear weapons, and banning of biological weapons use and development all played a significant role in reducing risk. The future applications of AI are uncertain, but even now there may be areas where treaties can be helpful in mitigating future risk. For instance, states can hopefully all agree that entrusting strategic nuclear weapons to the control of AI “dead man’s switches” would run a tremendous and highly unjustified risk. The current moment, in which the competitive pressures to develop military AI systems are more distant, is the proper time to consider what capabilities the U.S. should seek to restrict or ban via treaty. The United States should also establish a government-wide policy on autonomous weapons systems that can harmonize policy across military and intelligence agencies and also be incorporated into the United States’ stance in diplomatic discussions about AI.

RECOMMENDATION 9:

DoD and the Intelligence Community should establish dedicated AI-safety organizations.

The *National Artificial Intelligence Research and Development Strategic Plan* established a strong agenda for research into AI-safety, covering improving explainability and transparency, building trust, and enhancing verification and validation.¹²¹ These are the right priorities, but it is a separate task to

ensure that research findings on AI safety are effectively incorporated into the plans, systems, and activities of the national security community. As the experience with nuclear weapons shows, establishing dedicated safety organizations is critical to ensuring that safety is given its due against the sometimes (though less often than is argued) competing interest of performance.

Establishing formal AI-safety organizations at DoD and the relevant Intelligence agencies would serve three purposes. First, these organizations can serve as a shared resource for learning about best practices and the latest research on AI-safety. Second, they can serve as a champion of safety as a priority in bureaucratic politics. Third, they could serve as an effective point of interface with private, outside groups.

RECOMMENDATION 10:

DARPA should fund research on fail-safe and safety-for-performance technology for AI-systems.

One difference between the U.S. nuclear submarine community, which had a spotless nuclear safety record, and the U.S. nuclear weapons program, which did not, is that safety is an inherent requirement for high performance on a nuclear submarine. If a nuclear submarine is a danger to its crew or itself, it is significantly less likely to achieve its mission. With nuclear weapons, some safety measures might decrease the chance of mission success if they make it more likely that the bomb will fail to detonate during an attack. This justification was used successfully for decades by Strategic Air Command leadership to refuse the introduction of even commonsense safety measures such as placing a combination lock on each weapon.

Applying the lesson of the remarkable safety record of the nuclear submarine community to AI, DoD should fund DARPA to investigate approaches and technologies that can simultaneously increase safety and performance in the development and fielding of AI-enabled systems. The goal should be

to give future developers a strong, performance-based incentive to pursue safety, rather than merely directives and requirements to do so.

RECOMMENDATION 11:

NIST and the NSA should explore options for countering AI-enabled forgery.

AI-enabled forgery will challenge Command and Control organizations and increase the threat of social engineering hacks for all organizations.

See Part 1 for a full explanation.

The National Institute of Standards and Technology (NIST) and the National Security Agency (NSA) should explore technology options for limiting the effectiveness of AI-enabled video and audio forgery.

Just as there are some (admittedly imperfect) technological solutions that attempt to prevent image software like Photoshop from being used to counterfeit money, there may be technological solutions that can mitigate the worst impacts of AI-enabled forgery. For instance, cameras could be designed that would hash encrypted video files in a block chain. This would not prevent later editing and forgery, but it would allow definitive, cryptographically secured evidence that a given version of a video or audio file existed at a given date. Though lay people would still struggle to know the truth, this might allow sophisticated investigators to definitively confirm that at least some versions were edited, since their hash date would be later than the original. This is but one potential research avenue to limit the impact of AI-enabled forgery. There may be significantly better alternatives discovered later. Regardless, the worst-case scenarios for widely available audio and video forging technology indicates that both technical and regulatory options should be explored. While NIST and the NSA are the best leads for this type of activity, it may make sense to support research through other organizations such as the National Science Foundation and DARPA.

Conclusion

We stand at an inflection point in technology. The pace of change for Artificial Intelligence is advancing much faster than experts had predicted. These advances will bring profound benefits to humanity as AI systems help tackle tough problems in medicine, the environment and many other areas. However, this progress also entails risks. The implications of AI for national security become more profound with each passing year. In this project, we have sought to characterize just how extensive these implications are likely to be in coming years.

We find that AI is likely to display some, if not all, of the most challenging aspects of prior transformative military technologies. In examining how national security policymakers responded to these prior technologies we agree with Scott Sagan, who pointed out that our forebears performed worse than we had known but better perhaps than we should have expected. The challenges they faced were tremendous.¹²²

Unfortunately, AI has the potential to be every bit as fraught with risk as these prior cases, perhaps more so given the speed of technological progress and the more complicated relationship between government and industry in the current era. Though we are encouraged by the bevy of high-quality AI reports that have been released in the past few years, we find that they are somewhat hampered by conservatism. In this work, we sought to honestly characterize the AI revolution as revolutionary, not merely different. The government will need to be ambitious to respond effectively.

Appendix: Transformative National Security Technology Case Studies

Case Study #1: Nuclear Technology

History

The concept of nuclear-powered superweapons that would transform warfare was discussed by scientific and political elites for decades prior to the weapons' creation.

The possibility of using radioactive material to produce super-powerful bombs was raised in popular science fiction as early as 1914. That year, H.G. Wells' novel *The World Set Free* described "atomic bombs" made from uranium dropped from planes that "would continue to explode indefinitely" thereby destroying whole cities in a world war to come.¹²³ Wells was friends with many of the preeminent scientists and politicians of the day, including Winston Churchill, and his idea was well known among elite scientific and political circles.¹²⁴

Starting in 1939, the United States government committed extraordinary financial, organizational, and human resources to nuclear weapons research and production.

The possibility of a technology capable of winning the war for whichever country developed it first was enough to justify unprecedented expense. After getting fully underway in 1942, the Manhattan Project's three-year cost of \$2 billion (in 1940's dollars) comprised nearly 1% of 1945 U.S. GDP.¹²⁵ The government enlisted many of the world's leading scientists, engineers, and mathematicians, both American and foreign, for the Manhattan project.

Extraordinary levels of spending and commitment of national resources to nuclear technology continued for many decades afterward. From 1947-1952, spending on nuclear weapons averaged 30% of total defense spending,¹²⁶ which in 1952 was 15% of U.S. GDP.¹²⁷ From 1940 to 1996, 11% of total government spending was related to nuclear weapons.

Table 4. U.S. Government Spending by Function, 1940-96¹²⁸	
	Billions (\$1996)
Nuclear Weapons and Infrastructure	\$5,821.0
Building the bomb	\$409.4
Deploying the bomb	\$3,241.0
Targeting and controlling the bomb	\$831.1
Defending against the bomb	\$937.2
Dismantling the bomb	\$31.1
Nuclear waste management and environmental remediation	\$365.1
Victims of U.S. nuclear weapons	\$2.1
Nuclear secrecy	\$3.1
Congressional oversight of nuclear weapons programs	\$0.9
Non-Nuclear National Defense	\$13,213.0
All Other Government	\$32,523.0
<i>Total National defense</i>	<i>\$19,034.0</i>
<i>% Nuclear</i>	<i>31%</i>
<i>Total Government Spending</i>	<i>\$51,557.0</i>
<i>% Nuclear</i>	<i>11%</i>

Nuclear weapons were immediately and widely seen as a game-changing technology, and the U.S. national security community transformed to adjust to the implications.

After WWII, the United States continued to devote ever-increasing resources to nuclear weapons. By 1948, the U.S. had enough parts for 56 atom bombs.¹²⁹ By 1950, that figure had increased to 300.¹³⁰ In 1967 the size of the U.S. nuclear arsenal peaked at 31,255 nuclear warheads.¹³¹

In 1952, the United States tested its first nuclear fusion device which, like the fission bomb, was the result of a crash research and development effort personally approved by the U.S. President. Nuclear weapons were the central basis of military power for the Truman and Eisenhower presidencies, which dramatically reduced the size of conventional military forces in favor of nuclear-capable bombers, artillery, and other weapons. The Department of Defense under both Presidents developed war plans that called for extensive use of nuclear weapons.

Key Technology Aspects

Destructive potential: Very high

The destructive power of nuclear weapons is immense, assured, and easily demonstrated. Skeptical adversaries rely on intelligence and analysis of performance in exercises and hostile engagements to accurately assess of an adversary's conventional military capability. With nuclear weapons, however, the destructive capability from even a single weapon test is both immense and obvious, as the fission bomb attacks on Hiroshima and Nagasaki—bombs a thousand times less powerful than later fusion bombs—proved.

Cost profile: Very high

Developing nuclear weapons required a significant portion of total government financial capacity of the first five nuclear weapons states and remains expensive today.

As mentioned above, the United States spent a significant portion of its total government budget on nuclear weapons. Though limited data is available, estimates from academics and intelligence agencies suggest that the financial burden of developing nuclear weapons was even greater for the Soviet Union (despite having gathered helpful espionage from the United States) and for China.¹³² In more recent decades, both Iraq¹³³ and North Korea¹³⁴ are each estimated to have spent billions to develop atomic weapons.

Technical complexity profile: Very high

Development of nuclear technology requires advanced scientific and engineering knowledge.

Lack of availability of weapons-grade material and the expertise on how to refine uranium ore into weapons-grade nuclear fuel are the most important barriers to nuclear proliferation. In the early stages, development of nuclear fuel manufacturing required the involvement of many of the world's foremost scientists and engineers. In 1964, China, whose scientists lacked deep expertise in underlying technologies, nevertheless succeeded in testing a nuclear weapon, but they might not have been able to do so without having received critical technical assistance from the Soviet Union from 1955-1959.¹³⁵

Today the development of the lowest-tech, lowest-yield nuclear bombs is within the technical capability of many states.¹³⁶

Military/Civil Dual-Use Potential: High

Nuclear technology has important civilian and commercial applications in energy and medicine, but both carry significant risks of nuclear proliferation.

In the United States, 20 percent of electricity is generated from nuclear power plants. Some countries rely on nuclear for significantly more. Nuclear power facilities are either government controlled or heavily regulated due to the risks of nuclear accident, terrorism, and due to the use of nuclear energy technology in weapons development.

Radioactive nuclear materials have important medical applications in the diagnosis and treatment of diseases, especially cancer. Many of these nuclear medicine applications and technologies were invented by government scientists at laboratories that also conducted nuclear weapons R&D. Radiopharmaceuticals are frequently produced using weapons-grade uranium, and the production of radiopharmaceuticals could be a significant source of nuclear terrorism and nuclear proliferation risk.¹³⁷

Difficulty of Espionage and Monitoring: Moderate

Shortly after the original development of nuclear technology, advances in aerospace reconnaissance and radioactive tracing made nuclear monitoring generally effective.

Thanks to aircraft and satellite overflights, combined with human intelligence and Signals Intelligence (SIGINT), the U.S. and its allies detected every nuclear weapons program before completion of development. However, in some cases, facilities were under construction or even operational for years before they were detected.¹³⁸

Government Management Approach

The U.S. government responded aggressively to the challenges presented by nuclear technology, creating new civilian and military agencies, forming extensive partnerships with the private and non-profit sectors, and devoting tremendous resources.

Immediately following the Manhattan Project, those organizations created to enable it were made permanent and were augmented by many new ones comprising what would later become today's Department of Energy. National laboratories were transferred from military to civilian control under the newly formed Atomic Energy Commission, which was given significant authority to regulate the entire nuclear domain. These labs were government-owned but were run in partnerships with academia and industry. For example, Sandia National Lab was managed and operated by AT&T corporation.

Nuclear weapons and national security was the single most important political issue following WWII and was subject to uniquely high levels of political and media scrutiny.

Congress established the Joint Congressional Committee on Atomic Energy less than a year after the bombing of Hiroshima. The Committee was given unprecedented legislative powers, including the ability to veto executive actions in advance, to demand information from and assistance from executive agency personnel, to authorize legislation without a vote by the full House and Senate, and (to some extent) to disregard spending limits from other laws.¹³⁹ Moreover, the Committee was staffed by some of the most senior and most powerful Representatives and Senators from both political parties.

The media paid close attention to developments in nuclear weapons, including technological developments and government actions. The media's efforts were hampered, however, by government secrecy and the repeated willingness of government officials to lie about incidents involving nuclear weapons for the goal of national security.¹⁴⁰

Results of The Government's Management Approach

Preserving U.S. Military Technological Leadership: Partial Success

The United States was the first country to acquire an atomic bomb and maintained a significant edge in nuclear armaments throughout the first three decades after their invention. The United States had more mega-tonnage of nuclear warheads and more ways to deliver them through at least 1972, at which point both superpowers were capable of unilaterally destroying the world's cities many times over.

The United States also developed nuclear powered submarines four years earlier than the Soviet Union. Moreover, U.S. nuclear submarines and later nuclear powered surface ships had significantly better safety and performance records than not only their Soviet counterparts,¹⁴¹ but also the commercial nuclear power plants in the United States and elsewhere.¹⁴²

The primary blemishes on this record are the failure of the U.S. nuclear community to prevent the unintentional transfers of its nuclear secrets to both its adversaries and allies. Soviet spies infiltrated the Manhattan Project early and stole material that accelerated their development of nuclear weapons by years.¹⁴³ One of the first Soviet atomic bomber designs was an inch-for-inch reverse engineered design of a captured U.S. B-29 bomber, the same type as the Enola Gay.¹⁴⁴ Later, the John Walker spy ring, active from 1967 through 1985, successfully gave the Soviet Union access to huge numbers of highly sensitive U.S. documents, including many secrets related to the nuclear submarine fleet and the operations of U.S. nuclear forces.¹⁴⁵

However, it is unclear that the United States' achievement of nuclear superiority brought it safety. During the Cuban Missile Crisis, the United States had a nuclear arsenal seventeen times as large as that of the Soviet Union. However, the deterrent effect of this was reduced since tactical nukes in Cuba were under local control of Soviet forces in Cuba. Leaders of these forces stated after the end of the Cold War that their plan was to retaliate with nuclear weapons in the event of a conventional U.S. invasion of Cuba. Robert McNamara has stated that the U.S. leadership made all its decisions

during the crisis under the false assumption that Cuban nuclear forces were under direct Kremlin control.¹⁴⁶

Supporting Peaceful Use of Nuclear Technology: Partial Success

After the Manhattan Project, scientists and resources were directed toward the possibility of generating electricity using nuclear energy, primarily for naval vessels.¹⁴⁷ For the first eight years after Hiroshima, nuclear power technology was considered too dangerous to be outside government control. The Atomic Energy Act of 1946 explicitly banned patents on nuclear technology not exclusively owned by the government.¹⁴⁸

The Eisenhower administration reversed this policy in 1953 and began promoting civilian and commercial use of nuclear technology. The government declassified important aspects of nuclear technology to allow nongovernmental use and reinstated private patent authority. Those companies in the private sector defense industry that were involved in the design of nuclear propulsion systems for submarines, e.g. General Electric and Westinghouse, were encouraged to invest and develop commercial nuclear power.¹⁴⁹ Additionally, Eisenhower's "Atoms for Peace" program encouraged American companies to develop commercial nuclear power in other countries, which became an important U.S. export industry and helped secure American leadership in commercial nuclear technology for several decades.

While the policy did result in benefits, the changes were less significant than expected. Changes in the power industry are a lengthy process, however, and nuclear power did not comprise a significant portion of overall U.S. electricity generation until the mid-1970s.¹⁵⁰ Nor did the nuclear power industry ever achieve its most boastful promises, such as the 1954 public claim by Atomic Energy Commission Chairman Lewis Strauss that nuclear power would bring "electrical energy too cheap to meter" within a single generation.¹⁵¹

Additionally, advocates of exporting peaceful nuclear energy underestimated the risks of proliferation. India, for example, acquired its first nuclear weapons using plutonium from a reactor built with Canadian and

United States assistance, which they had previously promised would be used for exclusively peaceful purposes.¹⁵²

Mitigating Catastrophic Risks of Nuclear Technology: Partial Failure

After the bombing of Nagasaki, the world did not experience a single nuclear weapons attack or a single unintentional nuclear detonation of an atomic bomb. We feel, however, that characterizing this record as a success would be inaccurate due to the high number of near misses of both accidental nuclear war and accidental nuclear detonation. This is not intended to be unduly critical of the organizations charged with managing the U.S. nuclear arsenal. Their mandate to deliver perfect readiness and perfect safety was uniquely difficult.

- 1. The United States experienced numerous nuclear weapons accidents, many of which were near misses that did result in significant release of radioactive material and could have resulted in full nuclear or thermonuclear explosions.**

A report compiled by the Nuclear Safety Department of Sandia National Laboratory found that between 1950 and 1968, no fewer than 1,200 nuclear weapons were involved in “significant” incidents and accidents. This number undercounts the true number of accidents by potentially as much as half, since the military did not keep accurate records on the subject until 1959.¹⁵³ This number includes relatively minor accidents but also includes nearly catastrophic ones such as the dropping of two armed thermonuclear warheads on Faro North Carolina in which every safety mechanism failed except for one, a safety switch which itself was later found to have failed in dozens of other, separate instances.

The accident in North Carolina is but one of a terrifying record. After surveying the record of nuclear close calls, we agree with former Secretary of Defense Robert McNamara that the absence of a catastrophic nuclear weapons accident can be attributed to luck at least as much as to well-designed technological and procedural safeguards.¹⁵⁴ This is especially compelling when the nuclear weapons record is compared with the failure-free performance of the nuclear submarine community.

2. The United States conducted nuclear weapons tests and deployed nuclear weapons without adequate evaluations of the risks.

As is famously known, some of the senior scientists involved in the invention of the first atomic bomb were at least modestly concerned that it would cause a chain reaction igniting all the nitrogen in the atmosphere and thereby end all life on Earth. The program's leaders conducted the test anyway.

Later, the scientists who conducted the first thermonuclear weapons test were astonished at the quantity and spread of deadly radioactive fallout over hundreds of miles around the testing zone, which vastly exceeded their experience with fission weapons and even their worst-case expectations for fusion weapons.¹⁵⁵

The nuclear program leadership's willingness to conduct these tests—in the absence of confidence about nuclear testing's effect on the atmosphere and without having imagined the risks from thermonuclear fallout—is strong evidence of their prioritizing technological progress over mitigating risk from their ignorance of nuclear outcomes. They were more concerned with mitigating the risk of deterrence failure.

3. Even where the risks of using nuclear weapons were clear, the responsible institutions repeatedly failed to implement needed safety measures due to cost concerns, biases towards destructive reliability over safety, and political infighting.

The first report on increasing nuclear weapons safety, authored by the Pentagon's Armed Forces Special Weapons Project, was not initiated until the middle of July 1957, more than twelve years after American warplanes began carrying them over U.S. soil.¹⁵⁶ The report found that nuclear weapons were highly vulnerable to accidental detonations from mechanical failure, human error, or malicious intent. The report was circulated at the highest levels of Pentagon leadership and suggested badly needed changes to the designs of existing and future nuclear weapons as well as the procedures surrounding their use. However, the responsible organizations resisted the needed steps. Even though most

of the recommendations in this and other safety reports would later be implemented, in general the fixes took decades or more to make the transition between the identification of serious risk and implementation of a resolution plan.¹⁵⁷

4. The United States' senior leadership did not always understand the extent to which they were not in control of every aspect of the nuclear arsenal.

The United States' nuclear forces were massive networks of disparate organizations responsible for training and managing hundreds of thousands of individuals and tens of thousands of weapons systems over multiple decades. To its credit, U.S. Strategic Air Command (SAC), which had principal responsibility for the operations of the airborne nuclear arsenal in the atomic age's first decades, created a strong culture of reliability and structured discipline.¹⁵⁸ Nevertheless, these procedures often failed to anticipate key challenges in nuclear technology management during real-world crises. For example, in the middle of the Cuban Missile Crisis, a previously planned ICBM test launch was conducted completely unbeknownst to the President and other leaders and despite the possibility that any launch might be interpreted by the Soviet Union as the beginning of a full scale nuclear first strike.¹⁵⁹ This is but one failure. The Union of Concerned Scientists maintains a list of more than a dozen declassified high-risk incidents.¹⁶⁰

5. The United States transferred custody of nuclear weapons

The United States transferred custody of nuclear weapons to NATO allies with inadequate security precautions and failed to sufficiently supervise their activity.

A 1960 Congressional investigation into U.S. owned nuclear weapons stored in NATO countries found frightening evidence of nuclear mismanagement. In the case of Italy numerous nuclear missiles were guarded by a single U.S. soldier with a handgun and the launch key tied around his neck. During that period, the Italian Communist Party was actively supported by the Soviet Union and was popular in the region where the nuclear weapons were stored.¹⁶¹ Security for U.S. nuclear weapons in Turkey were even worse.¹⁶²

Case Study #2: Aerospace Technology

History

Military aviation began with the use of balloon airships in Europe in the late 1700s, but lack of steering and logistical challenges limited their effective use to reconnaissance and communications for a century.

Balloon technology, invented in France in 1783, was quickly recognized as a useful technology for military reconnaissance and saw extended use in conflicts such as the Napoleonic wars, the U.S. Civil War, and the Franco-Prussian War. By the 1880s most European armies had dedicated corps of balloon engineers. During this time, optimism grew about potential offensive capabilities of future balloons.¹⁶³ Both the American and the European balloon and later airplane industries viewed governments as their primary prospective customer from the earliest days of flight onward.

Science fiction of the late 1800s routinely described futures with cannon and bomb-armed airships.

Popular science fiction from Jules Verne's *Clipper of the Clouds* (1873) and Albert Robida's *War in the Twentieth Century* made aircraft engaging in dogfights and dropping bombs on populated cities a well-known concept long before technology made it possible. Military theorists such as Giulio Douhet and inventors such as Count Ferdinand von Zeppelin were counseling generals that technology would make such envisioned futures inevitable.¹⁶⁴

Fears of aerial bombing led to an international treaty banning the use of weaponized airships, but voluntary restraint was quickly abandoned and did not stop air war in WWI.

At an international arms control conference of 1899, Czar Nicholas II successfully lobbied for a "prohibition of the discharge of any kind of projectile or explosive from balloons or by similar means." The ban lasted five years and was observed by all the European great powers. The second

Hague conference of 1907 addressed renewing the ban but failed. Every European belligerent's capital (save Rome) was bombed from the air.¹⁶⁵

Key Technology Aspects

Destructive potential: Moderate

Individual aircraft carrying conventional explosives can cause damage, but only in vast quantities do aircraft pose a threat remotely comparable to nuclear weapons. The real destructive risk in aerospace technology comes not from individual aircraft, but from air forces. A nation faces existential risk from conventional aerospace technology only in the possibility that a military opponent with superior capability will repeatedly bomb it with large fleets, as happened to Germany and Japan in World War II and Iraq in the Gulf War.

Cost profile: Initially low, then high

Today's military aircraft cost millions or billions of dollars per unit, but in the first few decades after invention, cutting-edge aircraft were affordable for affluent civilians.

During World War I, the main U.S. fighter aircraft cost a little more than ten times the price of a civilian car. By 1945, fighter aircraft were roughly 50 times as much as a new civilian car¹⁶⁶ while advanced bombers were more than 650 times as costly. To research, design, and build the B-29 bomber, the U.S. government spent \$3.7 billion (in 1945 terms), nearly twice the amount spent on the Manhattan Project.¹⁶⁷

Technical complexity profile: Initially moderate, then high

Aerospace technology attracted some of the best scientific and engineering minds from its beginning. Early military aircraft were straightforward enough that car and even bicycle mechanics could build and modify them. By World War II, however, aircraft cost and complexity had ballooned to

the point where only the most sophisticated organizations could push the state of the art.

Military/Civil Dual-Use Potential: High

Through World War II, there was minimal difference between commercial and military aircraft technology, and significant overlap with other scientific and industrial sectors.

After World War I, Germany was banned from producing military aircraft. The enforcers of the peace treaty faced major challenges in that performance requirements for military and commercial aircraft were essentially identical. In Europe, one of the first commercial airlines built its passenger service business using reconfigured WWI bombers.¹⁶⁸

In terms of manufacturing and industrial requirements, the aircraft industry also shared many similar needs with other industries, especially the automobile industry. In both the first and second World Wars, automobile manufacturers reconfigured their plants to build engines, other systems and even whole military aircraft.¹⁶⁹

Difficulty of Espionage and Monitoring: High

Commercial aerospace facilities require similar talent and equipment to the military aerospace facilities and are not especially amenable to Intelligence, Surveillance, and Reconnaissance (ISR) monitoring.

For a large portion of the 20th century, there was significant overlap between military aerospace R&D and manufacturing and general commercial industry, such as the automobile industry. This dual-use issue made it difficult to monitor military aerospace development programs, except for rocketry.

Government Management Approach

The U.S. government has always played a very active role in the aerospace market: providing R&D support, acting as an anchor customer, and developing regulations and standards to enforce use of safety-enhancing technologies and procedures.

Research and development support: Congress established the National Advisory Committee for Aeronautics (NACA) as part of the Naval Appropriation bill in 1915. NACA, which ultimately evolved into the National Aeronautics and Space Administration (NASA), began small but grew rapidly into the primary government aerospace research institution, complete with its own national laboratory in 1917—one of the first U.S. government laboratories in any scientific discipline. Research produced and shared by NACA, especially related to its wind tunnel technology, had a critical impact on the success of the U.S. aircraft industry improving performance and safety.¹⁷⁰ NACA also played an important role in visiting aerospace companies and researchers in Europe and disseminating their latest advances to U.S. companies. The Army and Navy established Aircraft Technical Boards to draw up requirements and assist the industry in meeting military needs. Later, the military established their own laboratories and funded significant research and development at both academic institutions and private contractors. Such approaches are standard now, but they were revolutionary at the time.

Acting as an anchor customer: Military orders in World War I led to an explosion in aircraft demand. Annual U.S. aircraft production exploded from 411 in 1916 to 14,000 in 1918, employing a reported 175,000 personnel in the process.¹⁷¹ Demand crashed after the war's end and by 1923 was again below 1916 levels, causing many firms to go under. The government responded by passing the Air Mail act of 1925. This made commercial companies responsible for government air mail delivery operations, thereby providing stable revenues for aircraft manufacturers and operators and allowing them to reach sustainable scale economies and to compete successfully in commercial markets.¹⁷² Though the aircraft industry remained tiny compared to the automobile industry, it did not collapse despite weak demand and strong European competition following the WWI.

Year	Total	Military	Civilian
1913	43	14	29
1918	14,020	13,991	29
1923	743	687	56
1928	4,346	1,219	3,127
1933	1,324	466	858
1938	3,623	1,800	1,823
1943	85,433	85,433	0
1948	9,838	2,536	7,302
1953	13,112	8,978	4,134
1958	10,938	4,078	6,860
1963	10,143	1,970	8,173
1968	19,362	4,440	14,922
1973	15,952	1,243	14,709

Regulation and standardization: The early air industry suffered from high rates of costly crashes and fatalities that frightened customers and ruined company finances. The government played an important role in addressing this problem with the Air Commerce Act of 1926. The Act required that pilots be trained and licensed, developed uniform standards for safety among both manufacturers and operators, and funded the development of a safety infrastructure. This was all supervised by a new aviation branch of the Department of Commerce, which would later evolve into today's Federal Aviation Administration (FAA).¹⁷⁴ Once this agency got underway in 1928, crash rates, though still unacceptable by today's standards, made continuous progress each year.

Year	Fatalities per Airline Passengers Carried
1930	1 per 50,000
1950	1 per 100,000
2012	1 per 9,900,000

The modern air transport regulatory complex represents the effective implementation of more than a century of technological and process wisdom for maximizing safety. Each commercial airline crash or major problem is investigated thoroughly by both industry and government officials, after which, procedures and technologies are implemented to minimize the risk of that specific crash cause occurring a second time.

During WWII aerospace technology and operations became one of the primary activities of the U.S. military and government. Aerospace technology became nearly synonymous with modern national power.

In 1941, the Army Air Corps was renamed the Army Air Force, a unit that grew so large and vital that it ultimately became an independent service branch, co-equal with the Army and the Navy. The Navy, for its part, also acquired significant aerospace capabilities to use aircraft carriers and to execute combined air/sea operations. During and after World War II, millions of American military service members and civilian support personnel were involved in conducting military and intelligence operations that were enabled by aerospace technology. The scale of these activities was colossal, comprising a significant portion of overall U.S. GDP. Air superiority and air power would be foundational goals for U.S. military strategy and operations from WWII onward.

Despite heavy government involvement in the aerospace industry, the U.S. aircraft industry remained fundamentally undergirded by the American economic model of capitalism and free enterprise.

The U.S. government played a more interventionist role in aircraft than in most other industrial sectors. Yet, even during the height of World War II and the Cold War, the U.S. government generally did not engage in aerospace production directly through government organizations or state-owned companies. These activities were left to private firms who competitively bid for government production contracts.

Additionally, the United States encouraged the commercialization of military aerospace technologies. For instance, the digital computer industry received significant early support from defense organizations who needed computer chips for their guided missile avionics. This commercialization allowed producers to expand to new markets outside of aerospace, which in turn allowed them to reach even greater economies of scale and reduce costs for aerospace customers.

During WWII and the Cold War, the United States engaged in industrial espionage on behalf of its military aerospace companies:

Where U.S. intelligence agencies uncovered superior foreign aerospace technologies, these were shared with government defense contractors who could incorporate these advances into their own designs. The longstanding, official U.S. policy on industrial espionage is not to engage in it in outside of national security industries, but U.S. defense aerospace companies have long benefitted from industrial espionage.¹⁷⁶

At times, the United States government implemented major changes in the overall American economy and education system, based on its goals for the aerospace industry.

During World War II, the industrial production agencies of the U.S. government developed prioritization quota systems for different raw materials, such that different industries received the quantities they needed to meet their production targets. In this regard, aerospace was no different from other wartime industries such as tank or ship manufacturing, although aerospace often has unique requirements, such as exotic materials.

During the Cold War, the U.S. government showed a continued willingness to reorient the nation's economy around improving aerospace industry competitiveness. Soviet advances in rocketry, as demonstrated by the launch of the Sputnik satellite, directly led to major reforms in the American education system with the National Defense Education Act of 1958. The Act provided annually one billion dollars of federal funding to American schools to expand and improve science and engineering education.¹⁷⁷

Results of the Government's Management Approach

Preserving U.S. Military Technological Leadership: Success

Much of the 20th century can be accurately summed up as an aerospace arms race. The United States was not always the clear, unambiguous leader. For example, the U.S. never fielded a U.S.-designed airplane in WWI, and the U.S. was notably behind Germany and the United Kingdom in early jet engine technology. However, The United States set the pace in many technological and operational domains and generally caught up rapidly in those instances when an adversary jumped ahead with a technological breakthrough. Though European aerospace industries had an edge during World War One and immediately prior to World War Two, the United States' overall record is best in class. The U.S. was first to invent the airplane in the early 1900s, first to cross the Atlantic in the late 1920s, and first to the moon in the 1960s. Even in the most famous instances of the United States being behind in aerospace—the early days of the Space Race—the deficit was less severe than is popularly imagined. When the Soviets were first to launch an uncrewed satellite in 1957, the United States matched the accomplishment just 14 weeks later. By 1961, the United States successfully launched their human spaceflight capsule a week prior to Yuri Gagarin's first human spaceflight. Had that uncrewed capsule carried a human, which it successfully could have, the U.S. would have had the first human in space. Later, the United States decisively proved aerospace leadership with the Apollo moon program. The U.S. government's technological management approach helped it in building the world's leading military and commercial aerospace industry.

Supporting Peaceful Use of Aerospace Technology: Success

Since there was such a significant overlap between civilian and military aerospace technology in the first four decades after the invention of the airplane, the strong performance mentioned above was replicated in the commercial sphere. After WWII, the United States emerged as the clear winner in building commercial aircraft for the rapidly growing market in air transportation. The Soviet Union, with less effect, used politics to

pressure its allies and clients to buy Soviet airplanes wherever the USSR held sufficient sway.¹⁷⁸

Mitigating Catastrophic Risks of Aerospace Technology: Success

Catastrophic risk in aerospace is very different from that of nuclear weapons. Rather than a single nuclear device being responsible for the death of millions, aerospace's primary risks are the small (compared to nuclear) loss of life incidents from airplane crashes. As mentioned above, the government's approach to reducing the risks of civilian and military air transportation has been spectacularly successful, and flying has gotten progressively safer over time.

The other primary risk for aerospace is that of falling behind in technology and air power, which the U.S. adequately addressed by building powerful military aerospace capabilities.

Case Study #3 Internet and Cyber Technology

History

The government played an integral role in the technological evolution of digital computing, internet networking, and cryptography, the three fundamental technologies enabling all cyberspace operations.

“Cyberspace,” according to P.W. Singer and Allan Friedman, refers to “the realm of computer networks (and the users behind them) in which information is stored, shared, and communicated online.”^{U 179} Modern cyberspace was enabled by three technologies:

- 1. Digital computing** (especially using silicon integrated circuits), which allows storage and processing of information by machines
- 2. Internet networking**, which allows for the connection and unification of different types of networks according to a single standard, namely internet protocol
- 3. Cryptography**, which allows for unrelated users to share data and infrastructure while maintaining data confidentiality and integrity

All three technologies were actively supported by the U.S. government. This support was crucial to the development of the internet from its early inception in the 1970s through the mid-1990s, when the internet took on a more commercial nature.

^U Cyberspace thus predates the invention of the internet and its predecessors, though in modern language cybersecurity and internet security are used interchangeably.

Key Technology Aspects

Destructive Potential: Moderate

As more and more of the world's systems become linked to computers and in turn to the internet, the destructive potential of a cyberattack has grown accordingly. The most typical cyberattack's destructive power is quite low, but there are indications for much greater potential. Three examples illustrate the destructive power available for skilled cyberattackers:

- **Cyber capabilities can augment physical military attacks:** In 2006, the Israeli intelligence agency Mossad reportedly used a cyberattack to spoof the entire Syrian air defense radar network, allowing the Israeli Air Force to enter Syrian airspace unnoticed until the missiles began exploding.¹⁸⁰ Hacking may be able to allow an adversary access to systems related to nuclear weapons, though how feasible this is unclear.¹⁸¹
- **Cyber capabilities can directly damage physical infrastructure:** In 2010, the Iranian nuclear program was set back many years when a cyberattack caused centrifuges to violently self-destruct (Singer and Friedman 117). This type of attack could in principle be used to damage many types of commercial and military infrastructure.
- **Cyber-espionage can acquire sensitive information:** The Chinese government has reportedly hacked many of the U.S. defense contractors and military organizations associated with the F-35 program. The R&D cost of the F-35 exceeded \$50 billion, and the Chinese are believed to have acquired nearly all the intellectual property associated with the plane. The Chinese are also believed to have hacked extremely sensitive information related to the U.S. nuclear arsenal.¹⁸²

However, cyber is distinct from nuclear or aerospace capabilities in that testing and demonstrating the destructive potential of a cyber capability can be difficult. Openly announcing that one was exploiting a vulnerability in an enemy's network will generally lead them to resolve that specific vulnerability. Accordingly, the game theory aspects of cyberweapons are still unclear and debated. For an in-depth discussion of these issues, see the author's article in *Vox*.¹⁸³

Cost Profile: Inexpensive

Cyber capabilities are cheap enough that even terrorists and criminals can afford quite useful capabilities. As with all existing productized software, the marginal cost of additional production is near zero.^v For those groups or individuals who are merely using cyber exploits developed by others, the price is often very low.

Individual attacks are cheap: The Department of Defense reports experiencing more than 10 million incursion attempts daily.¹⁸⁴ However, some actors, notably the United States, see value in spending heavily on cyber. For Fiscal Year 2017 budget, then President Obama requested \$17 billion for cybersecurity, an increase of 35% over the previous request.¹⁸⁵ This figure reflects the scale of both the challenge the U.S. faces in securing its expansive data networks and its ambitions in exploiting weaknesses in the networks of others.

Nevertheless, cyber delivers capabilities at costs that are multiple orders of magnitude below what they would otherwise cost. As Bruce Schneier points out, “the exceptionally paranoid East German government had 102,000 Stasi surveilling a population of 17 million: that’s one spy for every 166 citizens, or one for every 66 if you include civilian informants.”¹⁸⁶ With digital surveillance, intelligence agencies and even corporations can collect data on hundreds of millions or even billions of individuals with far fewer resources than the Stasi.

Technical Complexity Profile: Moderate

As stated previously, there is broad diversity in the type, sophistication, and impact of cyber operations. The technical sophistication required varies accordingly. Some attacks, such as the Stuxnet virus that knocked out one-fifth of Iran’s nuclear centrifuges, likely require resources and capabilities likely to reside only within military and intelligence agencies.¹⁸⁷ Others, such as spear “phishing” attacks to acquire user credentials, can be executed by so called “script kiddies,” hackers who lack detailed technical

^v There is an important distinction, though, between those cyber exploits which do not require high levels of customization, such as those that apply to widely used desktop computer operating systems, and those which do require high levels of customization, such as industrial control software. Only in the former is their minimal marginal cost of utilization.

understanding of the exploits they are using.¹⁸⁸ Depending on the system authorizations of the stolen credentials, however, spear phishing attacks can be highly impactful.

Military/Civil Dual-Use Potential: High

The basic requirements for using and accessing digital networks are similar for both commercial and military users. In 2011, more than 90 percent of military digital communications took place over civilian networks.¹⁸⁹ Militaries likewise make extensive use of commercial computing hardware, though it is sometimes modified to meet their security or operational requirements.

In terms of cyber defense operations, the military and civilian communities share many of the same needs—preserving data confidentiality, integrity, and service availability. Both groups need to secure their data with strong cryptography, and rapidly patch vulnerabilities in the systems they use. As U.S. corporations become increasingly under threat from cyber criminals and adversarial states, their cybersecurity needs have correspondingly increased. The commercial cybersecurity market was estimated at \$75 billion in 2015, and may double that figure as soon as 2020.

Only governments have a strong case for needing cyber offense capabilities, whether attack or exploitation. But, both defense and offense involve looking for vulnerabilities. Only the hunting ground changes.

The best evidence for a high degree of technological overlap is that individuals with job experience in government cyber organizations are in high demand among commercial firms looking to secure their networks. In 2015, the U.S. National Security Agency began licensing its cyber defense software to commercial companies and saw strong demand.¹⁹⁰

Difficulty of Espionage and Monitoring: Very Difficult

Cyber capabilities are difficult to monitor. Military cyber equipment is generally very similar to commercial information technology equipment. As internet security becomes increasingly important to commercial entities, the staff's training and experience are likely to increasingly resemble that of an offensive cyber entity. To some extent, the fact that cyber offense and exploitation are so much easier than defense¹⁹¹ has allowed for mutual infiltration and monitoring of many of the more sophisticated government military organizations, but in practice much remains secret and unknown. Moreover, criminal and terrorist groups have had considerable success in hiding their online activities.

Government Management Approach

The government was a highly active supporter of the U.S. semiconductor industry, which was a key technological enabler of modern digital computing.

U.S. government intervention was crucial to the semiconductor industry's progress in both early and mature stages. In the early stage, the U.S. military's role as an R&D subsidizer and an anchor customer was crucial to driving investment, innovation, and growth.¹⁹² In 1987, the Department of Defense matched R&D investments up to \$100 million annually in the U.S. semiconductor industry research consortium, which was crucial in restoring U.S. competitiveness against Japan.¹⁹³

The U.S. government was a highly active supporter of the development of internet networking technologies and computer science research generally.

The U.S. DOD's Defense Advanced Research Projects Agency (DARPA, previously ARPA) is the primary government defense organization funding long-term advanced research and development projects. University scientists, working as DARPA program managers and with DARPA

funding, developed ARPANET, a network for sharing computing resource access. In 1973, Stanford professor Vint Cerf, working with Robert Kahn of ARPA, developed the internet protocol that would ultimately evolve into a common standard that can be used to connect any two information networks.¹⁹⁴ After the invention of internet protocol, the government continued to support the development of the internet by funding procurement of internet backbone infrastructure, promoting use of the internet at government science agencies, and funding continued technological R&D and standardization.¹⁹⁵

From 1975 through 1996, unclassified federal government funding for computer science research increased nearly five-fold, from roughly \$200 million to nearly \$1 billion (\$1995).

The U.S. government invested heavily in developing advanced cryptography mathematics and technology, but restricted its use to government organizations for several decades.

Since its founding, the U.S. National Security Agency (NSA) has had an intimate relationship with the study of cryptography. The NSA's dual mandate is to secure the confidentiality of communications of the U.S. and simultaneously to intercept the communications of other governments. Accordingly, it has, since its inception, employed large numbers of mathematicians and engineers to develop advanced cryptography and other information security technologies.

The U.S. military and intelligence communities were also the largest customers for cryptography technology. As a result, the best cryptographic capabilities resided in government. Unlike digital computing and networking technologies, however, the U.S. government's official policy, for many decades after the war, was that cryptography was a sensitive enough technology to be legally treated as a military munition. Accordingly, the U.S. government banned overseas sale of advanced cryptography software by U.S. firms.

Rather than develop different software versions for domestic and international markets, nearly all U.S. information technology firms used

cryptography software weak enough to meet U.S. government export restrictions in both markets. The weaker cryptography standards were easily cracked by interested parties, but in the nascent days of the internet, the U.S. government considered this a minor risk. The law was only relaxed in the late 1990s, by which time non-NSA affiliated academics had made considerable advances in developing strong cryptography, and competing high-quality foreign cryptography software became widely available.¹⁹⁶

Results of the Government's Management Approach

Preserving U.S. Military Technological Leadership: Success

As Bruce Schneier points out, the United States is the undisputed leader in cybersecurity technology because of three key advantages:

It has a larger intelligence budget than the rest of the world combined. The Internet's physical wiring causes much of the world's traffic to cross U.S. borders, even between two other countries. And almost all the world's largest and most popular hardware, software, and Internet companies are based in the U.S. and subject to its laws. It's the hegemon.¹⁹⁷

The United States has by far the most advanced capabilities in both cyber offense and cyber defense, but it is not clear that dominance in cyber will ever be comparable to dominance of the air, where the United States can establish undisputed air superiority and can prevent other militaries from even operating in a given airspace. It is unlikely that any adversary could, for example, deploy a bomber to destroy a U.S. power plant or radar installation. With cyber, however, many U.S. potential adversaries now possess the capability to destroy U.S. mainland power plants or take radars offline. In 2014, Admiral Michael Rogers, director of the NSA, testified before congress that China, as well as other countries, currently possesses the ability to use a cyberattack to take down the U.S. power grid. This could be evidence that the United States failed to invest and plan sufficiently for

its cyber defense, or it may simply reflect the uniquely difficult technical realities of cyberspace as a warfare domain.¹⁹⁸

Supporting Peaceful Use of Cyber Technology: Partial Success

The United States internet and information technology industries are unambiguously the leaders worldwide. Across the internet technology industry, U.S. companies lead in search, social networking, mobile hardware, internet infrastructure, and delivery of cloud-based services. In general, U.S. internet policy has supported economic growth and U.S. competitiveness across this domain. As President Barack Obama stated in 2015,

*[The United States has] owned the internet. Our companies have created it, expanded it, and perfected it in ways that they can't compete. And oftentimes what is portrayed [by foreign countries] as high-minded positions on issues sometimes is just designed to carve out some of their commercial interests.*¹⁹⁹

Not all U.S. policies have been supportive of U.S. commercial competitiveness in the internet industry, however. The NSA's restriction on use of advanced cryptography through the mid-1990s at one point made European software more competitive than it otherwise might have been.²⁰⁰ Additionally, many U.S. companies have claimed that government surveillance of U.S. digital equipment and networks hurts the competitiveness of American firms in export markets. Referring to reports of U.S. government surveillance in 2013, Facebook CEO Mark Zuckerberg said, "The government response was, 'Oh, don't worry, we're not spying on any Americans.' Oh, wonderful: that's really helpful to companies trying to serve people around the world, and that's really going to inspire confidence in American internet companies."²⁰¹ In recent years, American technology firms such as Apple have shown increased willingness to resist government requests for cooperation in enabling government digital surveillance.

Perhaps more problematic, however, is how the U.S. government supported the commercial development of the internet while not taking adequate steps to ensure security for individuals and organizations that use the internet.

Mitigating Catastrophic Risks of Cyber Technology: Partial failure

While the United States has had astounding success in cyber offense, the government failed for decades to develop a strategy that adequately addressed the asymmetric vulnerability it faced in terms of cyber defense. As former Director of National Intelligence Mike McConnell stated in 2010 testimony before Congress, “If the nation went to war today, in a cyberwar, we would lose. We’re the most vulnerable. We’re the most connected. We have the most to lose.”²⁰² The previously mentioned loss of the F-35 intellectual property illustrates this asymmetric vulnerability in another way: when China or another adversary hacks the United States, they can spend nearly nothing to steal cutting edge technology and designs that cost \$50 billion to develop.^w When the U.S. hacks China, they can only learn about older, essentially obsolete military technology, though this will likely not always be the case. No one has yet died from the theft of the F-35 plans, but in the event of a future conflict, China would have military capabilities, perhaps new missiles or planes or electronic countermeasures, that they would not otherwise have. In a war, this type of failure would cost the lives of American military personnel.

The situation has improved in the years since McConnell’s testimony. The United States federal government, especially national security and homeland security agencies, have provided increased support to commercial firms to secure their networks and systems.²⁰³ Still, a 2017 report by the Government Accountability Office found that the federal government still needed to do significantly more to protect its own networks.²⁰⁴

^w Note: acquiring research data and designs are not equivalent to acquiring a capability in many domains. Inability to use the results of cyber-espionage to foster other required skills—for example, the fabrication process for jet engines—remains an important limitation.

Case Study #4 Biotechnology

History

Militaries have intentionally used biological disease as a weapon for thousands of years.

Humanity has suffered from disease outbreaks for as long as there have been humans. The worst outbreak was likely the Black Death, which killed an estimated 200 million people (including roughly 1/3 of the European population) during the 14th century.²⁰⁵ Even at the peak of the Black Death's devastation, militaries made use of it for warfare: At the 1346 Siege of Caffa, the Mongol army used catapults to hurl plague-infected corpses over the walls of the besieged city.²⁰⁶ Evidence exists for much earlier wartime uses of infectious disease, as early as 600 BCE.

Disease has also seen more recent wartime use on the North American continent. During the Seven Years' War, "the British army used a few infected blankets to start a smallpox epidemic in an enemy American Indian tribe."²⁰⁷

The modern history of biological weapons is interwoven with that of chemical weapons, which saw extensive use during WWI. The Geneva Protocol of 1925 banned use of both biological and chemical weapons.

Though The Hague Declaration of 1899—ratified by all major powers except the United States—prohibited the military use of "Asphyxiating Poisonous Gases," all World War I belligerents ultimately made use of chemical weapons.²⁰⁸ Despite that failure of pre-war diplomacy and voluntary restraint, the great powers again banned military use of chemical weapons in the Geneva Protocol of 1925.

Biological weapon attacks did not play a significant role in WWI, though disease certainly did. The so-called "Spanish" Influenza of 1918-1919 infected an estimated one-third of the world's population (500 million people) and killed an estimated 50 million.²⁰⁹ The Geneva protocol likewise banned "the use of bacteriological methods of warfare."²¹⁰ The United States

signed the Geneva protocol in 1925 but did not ratify it until 1975. After the required number of countries ratified the treaty, it went into effect in 1928.

The Geneva Protocol only banned the military first use of bioweapons, not their development or stockpiling. After WWI and especially during WWII, many militaries, including the United States, worked to develop industrialized biological warfare.

The immense destructive potential of disease did not dissuade countries from developing biological weapons. Rather, some nations sought to utilize their improved understanding of medicine, public health, and chemical weapons to develop powerful biological armaments that were orders of magnitude more destructive than chemical weapons. The French, who had a rich medicinal science legacy dating back to Louis Pasteur, were the most aggressive and sophisticated in developing bioweapons during the interwar period.²¹¹ After Germany occupied France, the United Kingdom, fearful that the Germans would inherit the advanced French program, began a bioweapons effort of their own. The UK, in collaboration with the United States and Canada, successfully mass-produced bioweapon munitions during WWII. However, the offensive elements of the United States' biological program were officially conceived as a deterrent against adversarial use of bioweapons on the United States. The official U.S. policy was no-first-use.²¹²

Despite bioweapons R&D and manufacturing by many WWII belligerents, only the Japanese made offensive use of biological weapons.

The architect of Imperial Japan's biological warfare program, Shiro Ishii, successfully persuaded the leaders of Japan's military that widespread acceptance of the Geneva Protocol (which Japan signed in 1925 but did not ratify until 1970) meant that Japan should aggressively develop a biological weapons program. Ishii and Japan believed that the Geneva Protocol meant other countries would foolishly neglect to develop biological weapons and that Japan could provide itself with a significant military advantage.²¹³ However, Japan's separation of its bioweapons program from its chemical weapons activity left it at a disadvantage in solving complicated problems

related to agent disbursement and munitions production. Japan used biological weapons against Chinese civilians and attempted to use them against Soviet forces, but Japan's primary attack vectors were disbursement of disease-infected fleas, poisoning of water wells, and use of infected kamikaze soldiers. Though they produced significant suffering, especially among Chinese civilians, they did not provide Japan with any significant wartime advantage.

During the first decades of the Cold War, both sides saw biological weapons as having destructive potential comparable to nuclear weapons, and both massively expanded their bioweapons programs.

After a brief, post-WWII reduction in activity, both the United States bioweapons program and its Soviet counterpart were restarted. In 1945, the U.S. and its allies foresaw future bioweapons having destructive potential rivaling nuclear weapons.²¹⁴ By the mid-1960s, the United States was spending \$300 million annually (not inflation-adjusted) on chemical and biological weapons and even seriously considered first use of biological weapons during wartime: In 1956, the U.S. Army manual, *The Law of Land Warfare*, removed all statements about biological weapons being “retaliation only” and stated explicitly that the United States was not party to any treaty that would restrict the use of biological weapons.²¹⁵ The United States did make significant use of chemicals during its conflicts, notably the Agent Orange herbicide during Vietnam, but there is no credible evidence that the United States ever used biological weapons during wartime.²¹⁶

The United States terminated its offensive biological weapons program in 1969 and began working to create an international treaty to ban biological weapons. The US' efforts culminated in the Biological Weapons Convention (BWC) of 1972.

After a formal policy review in 1969, then-president Nixon stated that the United States would dismantle its offensive biological weapons program and thereafter only devote U.S. efforts to “research and development for defensive purposes.”²¹⁷ The U.S. then began negotiating with the Soviet Union and other nations, which resulted in the BWC of 1972. The BWC banned all non-defensive biological weapons activity but lacked effective

enforcement or monitoring mechanisms. The United States and the Soviet Union both signed the treaty in 1972, and it went into effect in 1975.

Over time, both the technological potential of peaceful biotechnology and the technological barriers to bioweapons development have changed significantly. This poses a challenge for managing proliferation.

Though only one terrorist group (Japan's Aum Shinrikyo) is known to have had an advanced bioweapons program,²¹⁸ the U.S. government has spent billions on both biodefense and technology management to address the threat of terrorists armed with biological weapons.

The rise of a commercial biotech industry has complicated these efforts by making the materials, the systems, and the technical knowhow needed for a biological weapons program more widespread, affordable, and more easily concealed under the auspices of a commercial effort.²¹⁹

Key Technology Aspects

Destructive potential: High

Military planners had credible evidence that non-contagious bioweapons (anthrax) could feasibly kill millions of people within days. Prior contagious disease outbreaks had a demonstrated ability to kill tens or hundreds of millions of people.

In 1944, during intense aerial bombardment of Germany, the UK Joint Planning staff drew up plans for bioweapon attacks on German cities that would have used four million air-dropped anthrax bombs (mostly manufactured in the United States) to kill an estimated three million German civilians.²²⁰ The accuracy of these estimates are difficult to prove but are plausible given the technology of the time.

Natural disease likewise proves how destructive biological weapons could become. The naturally occurring influenza outbreak of 1918 killed an estimated 50 million people worldwide.

Cost Profile: Initially High, Currently Low

Though the United States and other countries spent heavily to develop bioweapons, they were viewed as being comparatively cheap, having greater destructive potential per dollar cost than the alternatives.

During WWII, the United States spent \$400 million in 1945-dollars (\$5.4 billion in 2017-dollars) on bioweapons, roughly one-fifth what was spent on the Manhattan project.²²¹ Most of this funding went to research and development. Biological weapons were seen as having significantly greater destructive capability per cost than chemical or conventional weapons.²²²

By the mid-1960s, the United States was spending \$300 million annually (not inflation-adjusted) on chemical and biological weapons. Most of this was going towards chemical weapons that were being used in the Vietnam War.

Today, biological weapons are within the grasp of well-funded terrorist groups, as demonstrated by the Japanese Terrorist organization Aum Shinrikyo.

Aum Shinrikyo, whose budget was in the tens of millions of dollars, had an advanced chemical and biological weapons program. They successfully managed to cultivate anthrax and were struggling but making progress on agent-dispersal technologies. Fortunately, the terrorists were working with less-virulent strains of the anthrax bacteria²²³ and were unsuccessful in their attempts to convert benign anthrax into a weaponizable form.²²⁴

Technical Complexity Profile: Initially High, Currently Low

At first, the USA believed that only industrialized countries could develop bioweapons, and that development and use of bioweapons could be effectively controlled.

In 1951, the U.S. Joint Chiefs of Staff released a report that commented favorably on the cost-profile of bioweapons compared with conventional and nuclear weapons. The report, however, assumed that only industrialized nations would be able to successfully develop biological weapons.²²⁵ Weaponization, especially mass production of disease agents, development of reliable storage technologies, and development of delivery mechanisms had proven highly challenging using the technologies and disease agents available during WWII.

Notably, the disease agent that was weaponized most extensively by both the United States and the Soviet Union was Anthrax, which is not human-to-human contagious.²²⁶ Therefore the impact of an anthrax attack, while devastating to the affected region, would not trigger an infectious plague outbreak that could “boomerang” and spread to the attacker’s home territory and population.^{x 227} Likewise, accidental infection at a manufacturing or research site could have deadly consequences, but would not trigger a contagious outbreak.

By the late 1960s, the United States’ assessment of the technological situation and its interests had changed. The U.S. saw bioweapons as unnecessary, given the existing atomic deterrence, and less controllable, given rapidly decreasing technological barriers.

As the technology for developing bioweapons increasingly became within the reach of lesser powers, the strategic calculus for the United States changed. In 1969, the U.S. National Security Council led a review of U.S. biological weapons policy and concluded in a position paper that the U.S.’ “major interest [...] is to keep other nations from acquiring them.”²²⁸ Possessing bioweapons did not improve the U.S.’ deterrent, which was primarily underwritten by nuclear weapons. But, the proliferation of biological weapons—viewed as a “poor man’s atomic bomb”—increased the threat to the U.S. posed by lesser powers and terrorists.

X The Soviet Union was less concerned about boomerang risk and did develop weaponized contagious diseases including smallpox (*Variola virus*) and pneumonic plague (*Yersinia Pestis*).

Military/Civil Dual-Use Potential: High

Most of the R&D workforce for WWII biological weapons programs was drawn from the medical and biological research communities.

The authors of the Rosebury-Kabat report, which led to the creation of the U.S. bioweapons program, were academic medical professionals, as was much of the research staff of the U.S. bioweapons program. Most of their equipment was purchased from the civilian medical or chemical industries. However, there was significant expertise drawn from the chemical weapons industry, which had important technical insights on weaponization and storage.

Difficulty of Espionage and Monitoring: High

Despite signing Biological Weapons Convention, the Soviet Union continued its biological weapons program unbeknownst to the United States. The Soviet program was reportedly dismantled after the dissolution of the Soviet Union but may continue.

Unlike the Geneva Protocol, the Biological Weapons Convention prohibited the development, manufacturing, and stockpiling of biological weapons, not merely their use in wartime. However, the treaty lacked effective provisions for inspection and monitoring. This, combined with the fact that offensive biological weapons programs are difficult to distinguish from defensive and public health programs, meant that the United States did not know that the Soviet Union never ended its offensive program. Some in the United States suspected, however, especially after the 1980 Anthrax outbreak in the Soviet city of Sverdlovsk. The Soviets plausibly blamed the outbreak on naturally occurring anthrax from a local textile mill, but the Russian government in 1992 revealed that the outbreak was caused by a leak from a nearby offensive bioweapons facility.²²⁹ The Russian government officially ended its bioweapons program after the end of the Cold War, but a high-level defector from the Soviet bioweapons program reported in 1998 that the Russian bioweapons program continues in a reduced form.²³⁰

After the end of the Cold War, the United States' primary concern was proliferation of biological weapons to smaller states such as Iraq and to terrorist groups. The United States has struggled to accurately monitor and stop these efforts.

Iraq began its bioweapons research and development program in 1984, the same year that the United States restored diplomatic relations.²³¹ By 1988, Iraq had begun mass production, unbeknownst to the international community.^Y The United Nations worked to compel Iraq to dismantle its program in the mid-1990s, which it did. However, due to Iraq's continued unwillingness to cooperate with inspections, many senior officials in the United States did not accept Iraq's claim that the program had ended. In 2003, then Secretary of State Colin Powell cited Iraq's purported continued possession of an advanced bioweapons program as a major justification for the U.S. invasion of Iraq.²³²

Government Management Approach

After a nearly thirty-year period of active bioweapons development without military use, the United States adopted a policy of total voluntary restraint, whereby it renounced bioweapons and worked to end them as a tool of war.

During WWII and the first decades of the Cold War, the United States amassed a major bioweapons arsenal and munitions production capability. For nearly all this time, the United States had an official no-first-use policy of bioweapons, meaning that the United States would only use bioweapons to retaliate against a military that attacked the United States with bioweapons. This is notably in contrast with the U.S. policy on nuclear weapons, where it has always refused to adopt a no-first-use policy.²³³

Beginning with the Nixon administration, the United States went further by unilaterally disarming its offensive bioweapons program and working on domestic and international non-proliferation regimes.

^Y It is unclear what level of knowledge United States intelligence agencies possessed regarding the Iraqi biological weapons program prior to the Gulf War

As commercial biotechnology has grown more capable and sophisticated, the biotech regulatory regime has grown more extensive both domestically and internationally.

The 1994 U.S. Senate's Riegal Report showed that during the 1980s the United States exported significant quantities of biotech machinery and materials—including four strains of anthrax in a sale approved by the U.S. Commerce Department—that ultimately were used for Iraq's development and manufacturing of bioweapons.²³⁴ This experience showed that effectively countering bioweapons proliferation would require extensive regulation of the commercial biotech industry. As Jonathan B. Tucker writes, the U.S. approach to managing biotech dual-use has “traditionally revolved around the materials, methods, and products involved in misuse, and governance strategies have also taken an ‘artifact-centric’ approach by seeking to control the availability of dual-use products and services.”²³⁵

The commercial biotech and civilian research communities have also adopted a voluntary restraint approach, notably with the Asilomar conference on recombinant DNA of 1975.

Recombinant DNA, a technology that involves inserting DNA from one organism into another organism's DNA, was a breakthrough in genetic engineering technology when discovered in 1972. The genetics research community quickly realized the significant implications of this technology and called for a temporary moratorium on recombinant DNA research. The field's leading researchers held a conference in 1975 to develop guidelines for research risk mitigation. As Katja Grace writes,

The conference ultimately recommended that the science continue and offered guidelines under which they thought it could do so safely. The resulting guidelines were adopted by the National Institutes of Health as a condition for funding, and were adhered to by others voluntarily. Over the years, the guidelines have become less restrictive as new information has emerged.

While the guidelines were generally adhered to in the West, the Soviet Union violated its obligations under the Biological Weapons Convention

and proceeded, with ultimately successful research, to weaponize recombinant DNA technology.²³⁶

Results of the Government's Management Approach

Preserving U.S. Military Technological Leadership: Not Applicable / Voluntary Restraint

The United States, along with the United Kingdom and Canada, had the most advanced biological weapons program during WWII. During the second and third decade of the Cold War, it is possible, though unclear, that the Soviets may have had a more advanced bioweapons program. After 1969, the United States unilaterally disarmed because it was comfortable ceding leadership in biological weapons given the strength of its nuclear deterrent and its primary interest in opposing the proliferation of biological weapons.

Supporting Peaceful Use of Biotechnology: Success

The United States is generally regarded as the world leader in the biotechnology industry, a position that it has maintained since the end of WWII.²³⁷ Biotechnology has seen many important technology advances in that time, including recombinant DNA, cloning, gene sequencing, synthetic biology, and gene editing. Throughout each, the United States has managed to remain at the cutting-edge in both research and commercial exploitation.

Mitigating Catastrophic Risks of Biotechnology: Partial Success

The post-BWC U.S. response to the risks of bioweapons and bioterrorism has been extraordinary. As L.P. Knowles writes, “the United States leads the rest of the world with respect to the extent and detail of its biosecurity legislation.” The United States has spent billions of dollars to establish the capacity to prevent the spread of biological weapons, manage the risks of dual-use biotechnology, and establish defenses against deliberate and accidental biotechnology risks. According to one estimate, the U.S. federal

government spent \$79 billion on civil biodefense between 2001 and 2014, with recent annual budgets nearly \$7 billion.²³⁸

Given that part of the stated justification for the Iraq War was belief in Iraq's possession of biological weapons, the United States has also demonstrated willingness to use military force to prevent proliferation of biological weapons.

Nevertheless, there are two important criticisms of U.S. policy that lead us to characterize its risk-management regime as only moderately successful. First, the United States did not develop effective tools for monitoring and countering the Soviet post-BWC bioweapons program. Three factors were especially worrisome:

1. The Soviets were experimenting with highly-contagious and highly-lethal pathogens;
2. The Soviets were using recombinant DNA and other techniques to increase the lethality and resistance to treatment of their weaponized pathogens; and
3. The Soviets had unsafe containment procedures and experienced several major containment failures and infectious outbreaks,

Combined, these aspects suggest that the United States' lack of knowledge about the post-BWC Soviet bioweapons program put the U.S. at significant risk despite its best efforts.

Second, the United States was late in developing its counter-proliferation approach to dual-use technologies. As a result, many of the most important assets that Iraq needed to develop its biological weapons program were acquired in legitimate trade with the United States.

Citations

- 1 Levinovitz, Aian. "The Mystery of Go, the Ancient Game That Computers Still Can't Win." *Wired*. May 12, 2014. Accessed January 1, 2017. <https://www.wired.com/2014/05/the-world-of-computer-go/#slide-1>.
- 2 Silver, David. "Mastering the game of Go with deep neural networks and tree search." *Nature* 529, no. 7587 (January 26, 2016): 484-89. January 26, 2016. doi:10.1038/nature16961.
- 3 Moravik, Matej, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. "DeepStack: Expert-level Artificial Intelligence in heads-up no-limit poker." *Science*, 2017. doi:10.1126/science.1256960.
- 4 Xiong, Wayne, Jasha Droppo, Xuedong Huang, Frank Seide, Mike Seltzer, Andreas Stolcke, Dong Yu, and Geoffrey Zweig. "Achieving human parity in conversational speech recognition." arXiv preprint arXiv:1610.05256 (2016).
- 5 Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097-1105. 2012.
- 6 Ernest, Nicholas, and David Carroll. "Genetic Fuzzy based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions." *Journal of Defense Management* 06, no. 01 (2016). doi:10.4172/2167-0374.1000144.
- 7 Müller, Vincent C., and Nick Bostrom. "Future Progress in Artificial Intelligence: A Survey of Expert Opinion." *Fundamental Issues of Artificial Intelligence*, 2014, 555-72. doi:10.1007/978-3-319-26485-1_33.
- 8 "Reagan Defense Forum: The Third Offset Strategy." U.S. Department of Defense. November 7, 2015. Accessed January 1, 2017. <https://www.defense.gov/News/Speeches/Speech-View/Article/628246/reagan-defense-forum-the-third-offset-strategy>.
- 9 United States. Department of Defense. Office of Net Assessment. *Summer Study - (Artificial) Intelligence: What questions should DoD be asking?* Chair and Editor Matthew Daniels. 2016.
- 10 Moy, Timothy. *War Machines: Transforming Technologies in the U.S. Military, 1920-1940* (Texas A & M University military history series; 71). 1st ed. Texas A&M University Press, 2001. Page 4.
- 11 United States. Department of Defense. Office of Net Assessment. *Summer Study - (Artificial) Intelligence: What questions should DoD be asking?* Chair and Editor, Matthew Daniels. 2016.
- 12 "Preparing for the Future of Artificial Intelligence." The Administration's Report on the Future of Artificial Intelligence. Executive Office of the President—National Science and Technology Council. October 2016. Accessed March 1, 2017. <https://obamawhitehouse.archives.gov/biogr/2016/10/12/administrations-report-future-artificial-intelligence>.
- 13 Domingos, Pedro. *The master algorithm: how the quest for the ultimate learning machine will remake our world*. London: Penguin Books, 2015. Page 279
- 14 Roff, Heather M. *Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons*. Report. Global Security Initiative, Arizona State University, Geneva, 2016.
- 15 Moy, Timothy. *War Machines: Transforming Technologies in the U.S. Military, 1920-1940* (Texas A & M University military history series; 71). 1st ed. Texas A&M University Press, 2001.
- 16 "What is 'fire and forget' principle used in the BrahMos missile? - Times of India." *The Times of India*. April 05, 2003. Accessed January 1, 2017. <http://timesofindia.indiatimes.com/What-is-fire-and-forget-principle-used-in-the-BrahMos-missile/articleshow/42505499.cms>.
- 17 United States. Department of Defense. Office of the Secretary of Defense. Department of Defense Directive 3000.09 by Secretary of Defense Ashton B. Carter. November 21, 2012. Accessed January 1, 2017. <http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf>.
- 18 Sander, Alison, and Meldon Wolfgang. "BCG Perspectives. The Rise of Robotics." August 27, 2014. Accessed January 1, 2017. https://www.bcgperspectives.com/content/articles/business_unit_strategy_innovation_rise_of_robotics/
- 19 Belton, Padraig. "Game of drones: As prices plummet drones are taking off." *BBC News*. January 16, 2015. Accessed March 24, 2017. <http://www.bbc.com/news/business-30820399>.
- 20 Gates, Bill. "A Robot in Every Home." *Scientific American*. February 1, 2008. Accessed March 24, 2017. <https://www.scientificamerican.com/article/a-robot-in-every-home-2008-02/>.

- 21 "Long-term price trends for computers, TVs, and related items: The Economics Daily." U.S. Bureau of Labor Statistics. October 13, 2017. Accessed March 1, 2017. <https://www.bls.gov/opub/ted/2015/long-term-price-trends-for-computers-tvs-and-related-items.htm>.
- 22 Scharre, Paul. "Robotics on the Battlefield Part II: The Coming Swarm." Center for a New American Security, October 2014. Accessed March 1, 2017. <https://www.cnas.org/publications/reports/robotics-on-the-battlefield-part-ii-the-coming-swarm>.
- 23 Pratt, Gill A. "Is a Cambrian Explosion Coming for Robotics?" *Journal of Economic Perspectives* 29, no. 3 (2015): 51-60. doi:10.1257/jep.29.3.51.
- 24 Schneier, Bruce. *Data and Goliath: the hidden battles to collect your data and control your world*. New York, NY: W.W. Norton & Company, 2016. Page 27
- 25 Marquis-Boire, Morgan, Bill Marczak, Claudio Guarnieri, and John Scott-Railton. *You Only Click Twice: FinFisher's Global Proliferation*. Publication. March 2013. Accessed January 1, 2017. <https://citizenlab.org/wp-content/uploads/2009/10/You-Only-Click-Twice-FinFisher%E2%80%99s-Global-Proliferation.pdf>.
- 26 "Thinking About ISIS and its Cyber Capabilities: Somewhere Between Blue Skies and Falling Ones." Center for Internet and Society. November 29, 2015. Accessed January 1, 2017. <http://cyberlaw.stanford.edu/blog/2015/11/thinking-about-isis-and-its-cyber-capabilities-somewhere-between-blue-skies-and-falling>.
- 27 Watson, Ben. "The Drones of ISIS." *Defense One*. January 12, 2017. Accessed March 22, 2017. <http://www.defenseone.com/technology/2017/01/drones-isis/134542/>.
- 28 "Ford Targets Fully Autonomous Vehicle for Ride Sharing in 2021; Invests in New Tech Companies, Doubles Silicon Valley Team | Ford Media Center." Ford. August 16, 2016. Accessed March 22, 2017. <https://media.ford.com/content/fordmedia/ina/us/en/news/2016/08/16/ford-targets-fully-autonomous-vehicle-for-ride-sharing-in-2021.html>.
- 29 Markoff, John. "The iPad in Your Hand: As Fast as a Supercomputer of Yore." *The New York Times*. May 09, 2011. Accessed March 22, 2017. https://bits.blogs.nytimes.com/2011/05/09/the-ipad-in-your-hand-as-fast-as-a-supercomputer-of-yore/?_r=0.
- 30 Sengupta, Biswa, and Martin B. Stemmler. "Power Consumption During Neuronal Computation." *Proceedings of the IEEE* 102, no. 5 (2014): 738-50.
- 31 Conversation with Admiral Michael Rogers during his visit to Harvard Kennedy School. Interview by author. October 5, 2016.
- 32 Schneier, Bruce. *Data and Goliath: the hidden battles to collect your data and control your world*. New York, NY: W.W. Norton & Company, 2016. Page 27
- 33 "Interview with John Launchbury, Director, DARPA's Information Innovation Office (I2O)." Interview by author. October 21, 2016.
- 34 Conversation with Admiral Michael Rogers during his visit to Harvard Kennedy School. Interview by author. October 5, 2016.
- 35 Interview with John Launchbury, Director, DARPA's Information Innovation Office (I2O)." Interview by author. October 21, 2016.
- 36 Roff, Heather M. *Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons*. Report. Global Security Initiative. Arizona State University. Geneva, 2016.
- 37 Cuthbertson, Anthony. "An algorithm powered by this \$35 computer just beat a human fighter pilot." *Newsweek*. July 03, 2016. Accessed January 1, 2017. <http://www.newsweek.com/artificial-intelligence-raspberry-pi-pilot-ai-475291>.
- 38 Eshel, Tamir. "Russian Military to Test Combat Robots in 2015." *Defense Update*. December 31, 2015. Accessed January 1, 2017. http://defense-update.com/20151231_russian-combat-robots.html.
- 39 Bendert, Samuel. "Get Ready, NATO: Russia's New Killer Robots are Nearly Ready for War." *The National Interest*. March 7, 2017. Accessed March 1, 2017. <http://nationalinterest.org/blog/the-buzz/russias-new-killer-robots-are-nearly-ready-war-19698>.
- 40 Axe, David. "In Case You Forgot, America's Got a Ton of Warplanes." *War is Boring*. June 27, 2013. Accessed March 24, 2017. <https://warisboring.com/in-case-you-forgot-americas-got-a-ton-of-war-planes-229b88710602#.c5z9w1iou>.
- 41 Society, National Geographic. "Canada Geese, Canada Goose Pictures, Canada Goose Facts." National Geographic. Accessed January 1, 2017. <http://animals.nationalgeographic.com/animals/birds/canada-geese/>.

- 42 Zastrow, Mark. "South Korea trumpets \$860-million AI fund after AlphaGo 'shock'" Nature News. March 23, 2016. Accessed March 25, 2017. <http://www.nature.com/news/south-korea-trumpets-860-million-ai-fund-after-alphago-shock-1.19595>.
- 43 "Interview with Dr. Sungwan Kim, Professor, Seoul National University." E-mail interview by author. March 26, 2017.
- 44 Greenberg, Andy. "Hackers Remotely Kill a Jeep on the Highway-With Me in It." Wired. July 21, 2015. Accessed January 1, 2017. <https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway>.
- 45 Gertz, Bill. "NSA: Nation State Cyberattack Included Virtual 'Hand-to-Hand Combat'" Washington Free Beacon. March 22, 2017. Accessed March 24, 2017. <http://freebeacon.com/national-security/nsa-nation-state-cyberattack-included-virtual-hand-hand-combat/>.
- 46 Ibid.
- 47 Work, Robert. "CNAS Defense Forum." U.S. DEPARTMENT OF DEFENSE. December 14, 2015. Accessed May 30, 2017. <https://www.defense.gov/News/Speeches/Speech-View/Article/634214/cnas-defense-forum/>.
- 48 Pisani, Bob. "What Caused the Flash Crash? CFTC, DOJ weigh in." CNBC. April 21, 2015. Accessed January 1, 2017. <http://www.cnbc.com/2015/04/21/what-caused-the-flash-crash-cftc-doj-weigh-in.html>.
- 49 Nguyen, Anh, Jason Yosinski, and Jeff Clune. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images." *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, December 5, 2014. Accessed January 1, 2017. doi:10.1109/cvpr.2015.7298640.
- 50 Allen, Greg. "America's plan for stopping cyberattacks is dangerously weak." Vox. March 27, 2017. Accessed April 30, 2017. <http://www.vox.com/the-big-idea/2017/3/27/15052422/cyber-war-diplomacy-russia-us-wikileaks>.
- 51 Alexander, David. "Theft of F-35 design data is helping U.S. adversaries - Pentagon." Reuters. June 19, 2013. Accessed April 30, 2017. <http://www.reuters.com/article/usa-fighter-hacking-idUSL2N0EV0T320130619>.
- 52 Thompson, Iain. "Script kiddies pwn 1000s of Windows boxes using leaked NSA hack tools." The Register. April 21, 2017. Accessed April 30, 2017. https://www.theregister.co.uk/2017/04/21/windows_hacked_nsa_shadow_brokers/.
- 53 Scharre, Paul. *Autonomous Weapons and Operational Risk*. Report. 20YY Future of Warfare Initiative. Center for a New American Security. February 2016. https://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf.
- 54 Domingos, Pedro. *The master algorithm: how the quest for the ultimate learning machine will re-make our world*. London: Penguin Books, 2015. Page 281
- 55 *IDC Digital Universe Study. Big Data, Bigger Digital Shadows and Biggest Growth in the Far East*. Report. EMC Corporation. December 11, 2012. Accessed January 1, 2017. file:///Users/gregoryallen/Downloads/Media_Presentation_2012_DigiUniverseFINAL1.pdf.
- 56 Domingos, Pedro. *The master algorithm. how the quest for the ultimate learning machine will re-make our world*. London: Penguin Books, 2015. Page 19
- 57 Johnson, R. Colin. "Microsoft, Google Beat Humans at Image Recognition | EE Times." EETimes. February 2, 2015. Accessed January 1, 2017. http://www.eetimes.com/document.asp?doc_id=1325712.
- 58 Simonite, Tom. "Why Amazon and the CIA want algorithms to understand satellite photos." MIT Technology Review. August 25, 2016. Accessed January 1, 2017. <https://www.technologyreview.com/s/602239/amazon-and-the-cia-want-to-teach-ai-to-watch-from-space/>.
- 59 Manella, Morgan. "88 satellites being launched will image the entire Earth daily." AOL.com. February 13, 2017. Accessed March 1, 2017. <https://www.aol.com/article/news/2017/02/13/88-satellites-being-launched-will-image-the-entire-earth-daily/21712844/>.
- 60 Thies, Justus, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. "Face2Face: Real-time Face Capture and Reenactment of RGB Videos." *Face2Face: Real-time Face Capture and Reenactment of RGB Videos*. June 2016. Accessed January 1, 2017. <http://www.graphics.stanford.edu/~niessner/thies2016face.html>.
- 61 Stewart, Craig. "Adobe prototypes 'Photoshop for audio'" Creative Bloq. November 03, 2016. Accessed January 1, 2017. <http://www.creativebloq.com/news/adobe-prototypes-photoshop-for-audio>.

- 62 Nguyen A, Yosinski J, Bengio Y, Dosovitskiy A, Clune J (2016). Plug & Play Generative Networks: Conditional Iterative Generation of Images in Latent Space. In Computer Vision and Pattern Recognition (CVPR '17), 2017. <https://arxiv.org/abs/1612.00005>
- 63 Keohane, Joe. "What News-Writing Bots Mean for the Future of Journalism." *Wired*. February 16, 2017. Accessed March 1, 2017. <https://www.wired.com/2017/02/robots-wrote-this-story/>.
- 64 Saito, Shunsuke, Lingyu Wei, Liwen Hu, Koki Nagano, and Hao Li. "Photorealistic Facial Texture Inference Using Deep Neural Networks." *Photorealistic Facial Texture Inference Using Deep Neural Networks*. December 02, 2016. Accessed January 1, 2017. <https://arxiv.org/pdf/1612.00523.pdf>.
- 65 Owens, Andrew, Phillip Isoia, Josh McDermott, Antonio Torralba, Edward H. Adelson, and William T. Freeman. "Visually Indicated Sounds." *Visually Indicated Sounds*. April 30, 2016. Accessed March 1, 2017. <https://arxiv.org/abs/1512.03512>.
- 66 Kenreck, Todd. "'Star Wars: Rogue One' Is Brilliant, That CGI Though." *Forbes*. December 19, 2016. Accessed March 1, 2017. <https://www.forbes.com/sites/toddkenreck/2016/12/19/star-wars-rogue-one-is-brilliant-that-cgi-though/>.
- 67 Ferreira, Becky. "Robots Can Smell Now." *Motherboard*. July 24, 2014. Accessed January 1, 2017. https://motherboard.vice.com/en_us/article/robots-can-smell-now.
- 68 Daniels, Matt (Study Chair) Department of Defense Office of Net Assessment: "Summer Study: (Artificial) Intelligence: What questions should DoD be asking" July 2016
- 69 Fisher, Max. "Syrian hackers claim AP hack that tipped stock market by \$136 billion. Is it terrorism?" *The Washington Post*. April 23, 2013. Accessed January 1, 2017. https://www.washingtonpost.com/news/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/?utm_term=.762eb0c09357.
- 70 Wang, Lu, Whitney Kusing, and Eric Lam. "Fake Post Erasing \$136 Billion Shows Markets Need Humans." *Bloomberg.com*. April 23, 2013. Accessed January 1, 2017. <https://www.bloomberg.com/news/articles/2013-04-23/fake-report-erasing-136-billion-shows-market-s-fragility>.
- 71 Timberg, Craig. "Russian propaganda effort helped spread 'fake news' during election, experts say." *The Washington Post*. November 24, 2016. Accessed March 1, 2017. https://www.washingtonpost.com/business/economy/russian-propaganda-effort-helped-spread-fake-news-during-election-experts-say/2016/11/24/793903b6-8a40-4ca9-b712-716af66098fe_story.html?utm_term=.5cbd2b78536e.
- 72 Trending, BBC. "#BBCTrending: Are #GazaUnderAttack images accurate?" *BBC News*. July 08, 2014. Accessed January 1, 2017. http://www.bbc.com/news/blogs-trending-28198622?ocid=social-flow_twitter.
- 73 Domingos, Pedro. *The master algorithm: how the quest for the ultimate learning machine will remake our world*. London: Penguin Books, 2015. Page 13
- 74 King, Ross D., Jem Rowland, Stephen G. Oliver, Michael Young, Wayne Aubrey, Emma Byrne, Maria Liakata, Magdalena Markham, Pinar Pir, Larisa N. Soldatova, Andrew Sparkes, Kenneth E. Wheilan, and Amanda Clare. "The Automation of Science." *Science* 324, no. 5923 (2009): 85-89. doi:10.1126/science.1165620.
- 75 Hinchliffe, Emma. "IBM's Watson supercomputer discovers 5 new genes linked to ALS." *Mashable*. December 14, 2016. Accessed January 1, 2017. <http://mashable.com/2016/12/14/ibm-watson-als-research/#oKfRVPG3C8qI>.
- 76 Tayaran-N., Mohammad-H., Xin Yao, and Hongming Xu. "Meta-Heuristic Algorithms in Car Engine Design: A Literature Survey." *IEEE Transactions on Evolutionary Computation* 19, no. 5 (2015): 609-29. doi:10.1109/tevc.2014.2355174.
- 77 Furman, Jason. "Artificial Intelligence, Automation, and the Economy." *White House Council of Economic Advisers*. December 2016. Accessed March 25, 2017. <https://cbamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy>.
- 78 "Farm Population Lowest Since 1850's." *The New York Times*. July 19, 1988. Accessed January 1, 2017. <http://www.nytimes.com/1988/07/20/us/farm-population-lowest-since-1850-s.html>.
- 79 "Difference Engine: Luddite legacy." *The Economist*. November 04, 2011. Accessed January 1, 2017. <http://www.economist.com/blogs/babbage/2011/11/artificial-intelligence>.
- 80 Matthews, Christopher. "Summers. Automation is the middle class' worst enemy." *Axios*. June 05, 2017. Accessed June 07, 2017. <https://www.axios.com/automation-is-already-the-middle-class-worst-enemy-2413151019.html>.
- 81 Dimsdale, Nicholas H., Nicholas Horsewood, and Arthur Van Riel. "Unemployment in Interwar Germany: An Analysis of the Labor Market, 1927-1936." *The Journal of Economic History* 66, no. 03 (September 2006): 778-808. doi:10.1017/s0022050706000325.

- 82 Brynjolfsson, Erik, and Andrew McAfee. "Will Humans Go the Way of Horses?" *Foreign Affairs*. September 05, 2016. Accessed January 1, 2017. <https://www.foreignaffairs.com/articles/2015-06-16/will-humans-go-way-horses>.
- 83 Ibid
- 84 Chatham House 2015—Chatham House Research Paper "The Resource Curse Revisited" Paul Stevens, Glada Lahn and Jaakko Kooroshy *Energy, Environment and Resources* | August 2015
- 85 This possibility was raised by Daniels, Matt (Study Chair) Department of Defense Office of Net Assessment: "Summer Study: (Artificial) Intelligence: What questions should DoD be asking" July 2016
- 86 Borger, Julian. "Tapes reveal Enron's secret role in California's power blackouts." *The Guardian*. February 04, 2005. Accessed January 1, 2017. <https://www.theguardian.com/business/2005/feb/05/enron.usnews>.
- 87 Bostrom, Nick. *Superintelligence: paths, dangers, strategies*. Oxford: Oxford University Press, 2014.
- 88 Weiserstein, Alex. "No one can stop President Trump from using nuclear weapons. That's by design." *The Washington Post*. December 01, 2016. Accessed January 1, 2017. https://www.washingtonpost.com/posteverything/wp/2016/12/01/no-one-can-stop-president-trump-from-using-nuclear-weapons-thats-by-design/?utm_term=.6ac66f4e18a.
- 89 Schiosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 79
- 90 Ibid Page 76.
- 91 Bostrom, Nick. *Superintelligence: paths, dangers, strategies*. Oxford: Oxford University Press, 2014.
- 92 Kennett, Lee B. *A history of strategic bombing*. New York, NY: Scribner, 1983 Page 10.
- 93 Ibid Page 15
- 94 Ibid Page 34.
- 95 Conversation with Admiral Michael Rogers during his visit to Harvard Kennedy School. Interview by author. October 5, 2016.
- 96 Joiner, Stephen. "The Jet that Shocked the West." *Air & Space Magazine*. December 2013. Accessed January 1, 2017. <http://www.airspacemag.com/military-aviation/the-jet-that-shocked-the-west-180947758/>.
- 97 Allen, Greg. "Thank Goodness Nukes Are So Expensive and Complicated." *Wired*. March 04, 2017. Accessed March 22, 2017. <https://www.wired.com/2017/03/thank-goodness-nukes-expensive-complicated/>.
- 98 "Preparing for the Future of Artificial Intelligence." *The Administration's Report on the Future of Artificial Intelligence*. Executive Office of the President—National Science and Technology Council. October 2016. Accessed March 1, 2017. <https://obamawhitehouse.archives.gov/blog/2016/10/12/administrations-report-future-artificial-intelligence>. Page 25
- 99 By Justin Fox. "Amazon and Google Change the R&D Race." *Bloomberg.com* August 09, 2016. Accessed January 1, 2017. <https://www.bloomberg.com/view/articles/2016-08-09/amazon-and-google-change-the-r-d-race>.
- 100 "Interview with Sean Legassick, Policy Advisor, Google DeepMind." Interview by author. December 15, 2016.
- 101 Kafka, Peter. "Google wants out of the creepy military robot business." *Recode*. March 17, 2016. Accessed January 1, 2017. <http://www.recode.net/2016/3/17/11597060/google-wants-out-of-the-creepy-military-robot-business>.
- 102 Schiosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 466.
- 103 McNamara, Robert. "Apocalypse Soon." *Foreign Policy*. October 21, 2009. Accessed February 24, 2017. <http://foreignpolicy.com/2009/10/21/apocalypse-soon/>.
- 104 McNamara, Robert, and Errol Morris. "Film Transcript: The Fog of War." Errol Morris: Film Transcript. 2003. Accessed May 1, 2017. http://www.errolmorriss.com/film/fow_transcript.html.
- 105 Guillemin, Jeanne. *Biological weapons: from the invention of state-sponsored programs to contemporary bioterrorism*. New York: Columbia University Press, 2006. Page 73.
- 106 Ibid page 93.
- 107 Quoted in Ibid page 123

- 108 Crickmore, Paul F., and Alison J. Crickmore. *Nighthawk F-117: stealth fighter*. St. Paul, MN: Motorbooks, 2003. Page 9.
- 109 Ufimtsev, P. Ya. "DTIC Translation - Method of Edge Waves in the Physical Theory of Diffraction." Defense Technical Information Center. September 07, 1971. Accessed April 02, 2017. <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0733203>.
- 110 *US Stealth Programs and Technology: Soviet Exploitation of the Western Press*. Report. Washington, DC: Central Intelligence Agency - Directorate of Intelligence, 1981. Accessed January 1, 2017. http://nsarchive.gwu.edu/NSAEBB/NSAEBB443/docs/area51_44.PDF.
- 111 "Department of Defense Announces Successful Micro-Drone Demonstration." U.S. Department of Defense. January 9, 2017. Accessed April 1, 2017. <https://www.defense.gov/News/News-Releases/News-Release-View/Article/1044811/departement-of-defense-announces-successful-micro-drone-demonstration/>.
- 112 Christensen, Clayton M., and Michael E. Raynor. *The innovator's solution: creating and sustaining successful growth*. Boston, MA: Harvard Business Review Press, 2013. Page 32.
- 113 Ibid.
- 114 Ibid page 57.
- 115 Mukunda, Gautam. "We Cannot Go On: Disruptive Innovation and the First World War Royal Navy." *Security Studies* 19, no. 1 (2010): 124-59. doi:10.1080/09636410903546731.
- 116 Wilson, J. R. "IED hunters adapt to sophisticated threats." *Military and Aerospace Electronics*. June 11, 2015. Accessed January 1, 2017. <http://www.militaryaerospace.com/articles/print/volume-26/issue-6/special-report/ied-hunters-adapt-to-sophisticated-threats.html>.
- 117 Nguyen, Anh, Jason Yosinski, and Jeff Clune. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images." *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. December 5, 2014. Accessed January 1, 2017. doi:10.1109/cvpr.2015.7298640.
- 118 Loizos, Connie. "It isn't just Uber: Carnegie Mellon's computer science dean on its poaching problem." *TechCrunch*. April 26, 2016. Accessed January 1, 2017. <https://techcrunch.com/2016/04/26/it-isnt-just-uber-carnegie-mellons-computer-science-dean-on-its-poaching-problem/>.
- 119 Paletta, Damian. "The CIA's Venture-Capital Firm, Like Its Sponsor, Operates in the Shadows." *The Wall Street Journal*. August 30, 2016. Accessed January 1, 2017. <https://www.wsj.com/articles/the-cias-venture-capital-firm-like-its-sponsor-operates-in-the-shadows-1472587352>.
- 120 "2015 Annual U.S. Venture Industry Report." *PitchBook News*. Accessed January 1, 2017. <http://pitchbook.com/news/reports/2015-annual-us-venture-industry-report>
- 121 National Science and Technology Council: "The National Artificial Intelligence Research and Development Strategic Plan" October 2016. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf.
- 122 Sagan, Scott D. *The limits of safety: organizations, accidents, and nuclear weapons*. Princeton (NJ): Princeton University Press, 1995.
- 123 Wells, Herbert George. *The World Set Free*. February 11, 2016. Accessed January 1, 2017. Full text available online at <http://www.gutenberg.org/files/1059/1059-h/1059-h.htm>.
- 124 Schlosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 37.
- 125 "Manhattan Project." *Manhattan Project: CTBTO Preparatory Commission*. 2007. Accessed January 1, 2017. <https://www.ctbto.org/nuclear-testing/history-of-nuclear-testing/manhattan-project/>.
- 126 Schwartz, Stephen I., ed. *Atomic audit: the costs and consequences of U.S nuclear weapons since 1940*. Washington (D.C.): Brookings institution, 1998. Page 8.
- 127 "Federal Government: National Defense Consumption Expenditures and Gross Investment." *FRED*---Saint Louis Federal Reserve. January 27, 2017. Accessed February 23, 2017. <https://fred.stlouisfed.org/series/FDEFX>.
- 128 Adapted from Schwartz, Stephen I., ed. *Atomic audit: the costs and consequences of U.S nuclear weapons since 1940*. Washington (D.C.): Brookings institution, 1998. Pages 4-5
- 129 Schlosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 98.
- 130 Ibid. page 126
- 131 Kristensen, Hans M., and Robert S. Norris. "Global nuclear weapons inventories, 1945–2013." *Bulletin of the Atomic Scientists* 66, no. 4 (November 2, 2016): 77-83. doi:10.2968/066004008.

- 132 Stephen J., ed. *Atomic audit: the costs and consequences of U.S. nuclear weapons since 1940*. Washington (D.C.): Brookings institution, 1998. Pages 611-613
- 133 "Interview with David Kay." PBS. Accessed February 23, 2017. <http://www.pbs.org/wgbh/pages/frontline/shows/unscom/interviews/kay.html>.
- 134 Pearson, James, and Ju-min Park. "North Korea overcomes poverty, sanctions with cut-price nukes." Reuters. January 11, 2016. Accessed January 1, 2017. <http://www.reuters.com/article/us-northkorea-nuclear-money-idUSKCN0UP1G820160111>.
- 135 "Sharing the Bomb among Friends: The Dilemmas of Sino-Soviet Strategic Cooperation." Wilson Center. November 18, 2015. Accessed January 1, 2017. <https://www.wilsoncenter.org/publication/sharing-the-bomb-among-friends-the-dilemmas-sino-soviet-strategic-cooperation>.
- 136 "Nuclear Terrorism FAQ." Harvard Kennedy School Belfer Center for Science and International Affairs. September 2007. Accessed January 1, 2017. <http://www.belfercenter.org/publication/nuclear-terrorism-faq>.
- 137 Covington, N. "Medical isotope production and nuclear terrorism." *Canadian Medical Association Journal* 179, no. 1 (2008): 54-55. doi:10.1503/cmaj.1080059.
- 138 "The nuke detectives." *The Economist*. September 05, 2015. Accessed January 1, 2017. <http://www.economist.com/news/technology-quarterly/21662652-clandestine-weapons-new-ways-detect-covert-nuclear-weapons-are-being-developed>.
- 139 Davis, Christopher M. *9/11 Commission recommendations: Joint Committee on Atomic Energy—a model for congressional oversight?* Washington, D.C.: Congressional Research Service, Library of Congress, 2004.
- 140 Schiosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 187
- 141 Drew, Christopher. "A Sad Record of Submarine Disasters." *The New York Times*. August 15, 2000. Accessed January 1, 2017. <http://www.nytimes.com/2000/08/16/world/a-sad-record-of-submarine-disasters.html>.
- 142 Conca, James. "America's Navy the Unsung Heroes of Nuclear Energy." *Forbes*. October 30, 2014. Accessed January 1, 2017. <https://www.forbes.com/sites/jamesconca/2014/10/28/americas-navy-the-unsung-heroes-of-nuclear-energy/#60a2bfd53eeb>.
- 143 Schiosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 85.
- 144 Ibid page 86
- 145 Weil, Martin. "John A. Walker Jr., who led family spy ring, dies at 77." *The Washington Post*. August 30, 2014. Accessed January 1, 2017. https://www.washingtonpost.com/national/john-a-walker-who-led-family-spy-ring-dies/2014/08/30/db041a56-2f9c-11e4-bb9b-997ae96fad33_story.html?utm_term=.3a9f5d696ddd.
- 146 McNamara, Robert. "Apocalypse Soon." *Foreign Policy*. October 21, 2009. Accessed February 24, 2017. <http://foreignpolicy.com/2009/10/21/apocalypse-soon/>.
- 147 Garwin, Richard L., and Georges Charpak. *Megawatts and megatons: the future of nuclear power and nuclear weapons*. Chicago, IL: The University of Chicago Press, 2002. Page 109
- 148 Schiosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 465.
- 149 Garwin, Richard L., and Georges Charpak. *Megawatts and megatons: the future of nuclear power and nuclear weapons*. Chicago, IL: The University of Chicago Press, 2002. Page 109
- 150 "U.S. Energy Information Administration - EIA - Independent Statistics and Analysis." *History of energy consumption in the United States, 1775–2009 - Today in Energy - U.S. Energy Information Administration (EIA)*. Accessed February 24, 2017. <http://www.eia.gov/todayinenergy/detail.php?id=10>.
- 151 Weilock, Thomas. "Too Cheap to Meter: A History of the Phrase." U.S. Nuclear Regulatory Commission. June 03, 2015. Accessed January 1, 2017. <https://public-blog.nrc-gateway.gov/2016/06/03/too-cheap-to-meter-a-history-of-the-phrase/>.
- 152 "Canadian-Indian Reactor, U.S." Nuclear Threat Initiative. September 1, 2003. Accessed February 24, 2017. <http://www.nti.org/learn/facilities/832/>.
- 153 Schiosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 327-328.
- 154 McNamara, Robert. "Apocalypse Soon." *Foreign Policy*. October 21, 2009. Accessed February 24, 2017. <http://foreignpolicy.com/2009/10/21/apocalypse-soon/>.

- 155 Jacobsen, Annie. *The Pentagon's brain: an uncensored history of DARPA, America's top secret military research agency*. New York, NY: Back Bay Books/Little, Brown and Company, 2016.
- 156 Schlosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 167.
- 157 Ibid. 466
- 158 Sagan, Scott D. *The limits of safety: organizations, accidents, and nuclear weapons*. Princeton (NJ): Princeton University Press, 1995. Page 57
- 159 Ibid. page 80
- 160 "Close Calls with Nuclear Weapons (2015)." Union of Concerned Scientists. April 2015. Accessed February 24, 2017. <http://www.uccusa.org/nuclear-weapons/hair-trigger-alert/close-calls>.
- 161 Schlosser, Eric. *Command and control: nuclear weapons, the Damascus Accident, and the illusion of safety*. New York, NY: Penguin Books, 2014. Page 256.
- 162 Ibid.
- 163 Kennett, Lee B. *A history of strategic bombing*. New York, NY: Scribner, 1983. Page 7.
- 164 Ibid. pages 8-9
- 165 Ibid. 34
- 166 Singh, Jasjit. *Air power in modern warfare*. New Delhi: Lancer International, 1988. Page 59
- 167 O'Brien, Phillips Payson. *How the war was won: air-sea power and Allied victory in World War II*. Cambridge, United Kingdom: Cambridge University Press, 2015. Page 48.
- 168 Kennett, Lee B. *A history of strategic bombing*. New York, NY: Scribner, 1983. Page 62-63
- 169 Birstein, Roger E. *The American aerospace industry: from workshop to global enterprise*. New York, NY: Twayne, 1996. Page 15
- 170 Ibid. Page 37.
- 171 Ibid. Page 225.
- 172 Ibid. Page 22.
- 173 Ibid. Page 225.
- 174 Ibid. Page 26.
- 175 Spencer, Glen C. "2012 - The safest year for air travel again | Flight Ascend Consultancy." Ascend - FG Advisory. February 8, 2013. Accessed February 24, 2017. <http://www.ascendworldwide.com/2013/02/2012---the-safest-year-for-air-travel-again.html>.
- 176 Nakashima, Ellen, and Steven Mufson. "U.S., China vow not to engage in economic cyberespionage." *The Washington Post*. September 25, 2015. Accessed January 1, 2017. https://www.washingtonpost.com/national/us-china-vow-not-to-engage-in-economic-cyberespionage/2015/09/25/90e74b6a-63b9-11e5-8e9e-dce8a2a2a679_story.html?utm_term=.5a3bab2cc59d.
- 177 Abramson, Larry. "Sputnik Left Legacy for U.S. Science Education." NPR. September 30, 2007. Accessed February 24, 2017. <http://www.npr.org/templates/story/story.php?storyId=14829195>.
- 178 Birstein, Roger E. *The American aerospace industry: from workshop to global enterprise*. New York, NY: Twayne, 1996. Page 135
- 179 Singer, Peter W., and Allan Friedman. *Cybersecurity and cyberwar: what everyone needs to know*. New York ; Oxford, NY: Oxford University Press, 2014. Page 13.
- 180 Singer, Peter W., and Allan Friedman. *Cybersecurity and cyberwar: what everyone needs to know*. New York; Oxford, NY: Oxford University Press, 2014. Page 126.
- 181 Blair, Bruce G. "Why Our Nuclear Weapons Can Be Hacked." *The New York Times*. March 14, 2017. Accessed May 1, 2017. <https://www.nytimes.com/2017/03/14/opinion/why-our-nuclear-weapons-can-be-hacked.html?mcubz=1&r=0>.
- 182 Dewey, Caitlin. "The U.S. weapons systems that experts say were hacked by the Chinese." *The Washington Post*. May 28, 2013. Accessed January 1, 2017. https://www.washingtonpost.com/news/worldviews/wp/2013/05/28/the-u-s-weapons-systems-that-experts-say-were-hacked-by-the-chinese/?utm_term=.866cb359bf98.
- 183 Allen, Greg. "America's plan for stopping cyberattacks is dangerously weak." *Vox*. March 27, 2017. Accessed March 27, 2017. <http://www.vox.com/the-big-idea/2017/3/27/15052422/cyber-war-diplomacy-russia-us-wikileaks>.

- 184 Fung, Brian. "How Many Cyberattacks Hit the United States Last Year?" Nextgov. March 8, 2013. Accessed February 24, 2017. <http://www.nextgov.com/cybersecurity/2013/03/how-many-cyberattacks-hit-united-states-last-year/61775/>.
- 185 Voiz, Dustin, and Mark Hosenball. "Concerned by cyber threat, Obama seeks big increase in funding." Reuters. February 10, 2016. Accessed January 1, 2017. <http://www.reuters.com/article/us-obama-budget-cyber-idUSKCN0VI0R1>.
- 186 Schneier, Bruce. *Data and Goliath: the hidden battles to collect your data and control your world*. New York, NY: W.W. Norton & Company, 2016. Page 27
- 187 Zetter, Kim. "Suite of Sophisticated Nation-State Attack Tools Found with Connection to Stuxnet." Wired. February 16, 2015. Accessed January 1, 2017. <https://www.wired.com/2015/02/kapersky-discovers-equation-group/>.
- 188 Singer, Peter W., and Allan Friedman. *Cybersecurity and cyberwar: what everyone needs to know*. New York; Oxford, NY: Oxford University Press, 2014. Page 78.
- 189 Nye, Joseph S., Jr. "Nuclear Lessons for Cyber Security." *Strategic Studies Quarterly*, December 2011. Accessed January 1, 2017. <http://www.au.af.mil/au/ssq/2011/winter/nye.pdf>.
- 190 Sternstein, Aliya. "Suddenly, Everyone Wants the NSA's Cyber Defense Tech." *Defense One*. July 13, 2015. Accessed February 24, 2017. <http://www.defenseone.com/technology/2015/07/uddenly-everyone-wants-nsas-cyber-defense-tech/117673/>.
- 191 Brenner, Joel, and Jon R. Lindsay. "Correspondence: Debating the Chinese Cyber Threat." *International Security* 40, no. 1 (2015): 191-95. doi:10.1162/isec_c_00208.
- 192 Scherer, F. M. "Semiconductors." *Industry Structure, Strategy, and Public Policy*. New York. Harper-Collins College, 1996. Page 202.
- 193 Ibid page 230.
- 194 Singer, Peter W., and Allan Friedman. *Cybersecurity and cyberwar: what everyone needs to know*. New York; Oxford, NY: Oxford University Press, 2014, page 18.
- 195 Ibid.
- 196 Schneier, Bruce. *Data and Goliath: the hidden battles to collect your data and control your world*. New York, NY: W.W. Norton & Company, 2016. Page 140
- 197 Schneier, Bruce. *Data and Goliath: the hidden battles to collect your data and control your world*. New York, NY: W.W. Norton & Company, 2016. Page 76
- 198 Brenner, Joel, and Jon R. Lindsay. "Correspondence: Debating the Chinese Cyber Threat." *International Security* 40, no. 1 (2015): 191-95. doi:10.1162/isec_c_00208.
- 199 Swisher, Kara. "White House. Red Chair. Obama Meets Swisher." *Recode*. February 15, 2015. Accessed February 24, 2017. <http://www.recode.net/2015/2/15/11559056/white-house-red-chair-obama-meets-swisher>.
- 200 Schneier, Bruce. *Data and Goliath: the hidden battles to collect your data and control your world*. New York, NY: W.W. Norton & Company, 2016. Page 140
- 201 Essers, Loek. "Facebook must comply with German data protection law, court rules." *PCWorld*. February 18, 2014. Accessed February 24, 2017. <http://www.pcworld.com/article/2098720/facebook-must-comply-with-german-data-protection-law-court-rules.html>.
- 202 Singer, Peter W., and Allan Friedman. *Cybersecurity and cyberwar: what everyone needs to know*. New York; Oxford, NY: Oxford University Press, 2014. Page 151.
- 203 Swisher, Kara. "White House. Red Chair. Obama Meets Swisher." *Recode*. February 15, 2015. Accessed February 24, 2017. <http://www.recode.net/2015/2/15/11559056/white-house-red-chair-obama-meets-swisher>.
- 204 Government Accountability Office. *Cybersecurity: actions needed to strengthen U.S. capabilities*. By Gregory C. Wilshusen. 2017.
- 205 Philipkoski, Kristen. "Black Death's Gene Code Cracked." *Wired*. October 03, 2001. Accessed January 1, 2017. <https://www.wired.com/2001/10/black-deaths-gene-code-cracked/>.
- 206 Wheelis, Mark. "Biological Warfare at the 1346 Siege of Caffa - Volume 8, Number 9-September 2002 - Emerging Infectious Disease journal - CDC." *Centers for Disease Control and Prevention*. July 16, 2010. Accessed January 1, 2017. https://wwwnc.cdc.gov/eid/article/8/9/01-0536_article.
- 207 Guillemin, Jeanne. *Biological weapons: from the invention of state-sponsored programs to contemporary bioterrorism*. New York. Columbia University Press, 2006. Page 3.
- 208 Graham, Thomas, Jr. *Cornerstones of Security Arms Control Treaties in the Nuclear Era*. Seattle. University of Washington Press, 2011.

- 209 Taubenberger, Jeffery K., and David M. Morens. "1918 influenza: the Mother of All Pandemics." *Emerging Infectious Diseases* 12, no. 1 (2006): 15-22. doi:10.3201/eid1209.05-0979.
- 210 Guillemin, Jeanne. *Biological weapons: from the invention of state-sponsored programs to contemporary bioterrorism*. New York: Columbia University Press, 2006. Page 4.
- 211 Ibid Page 25.
- 212 Ibid page 60.
- 213 Ibid page 81.
- 214 Ibid page 98.
- 215 Ibid page 107.
- 216 Ibid page 100.
- 217 Ibid page 125.
- 218 Tucker, Jonathan B., ed. *Innovation, dual use, and security: managing the risks of emerging biological and chemical technologies*. Cambridge: MIT Press, 2012. Page 9.
- 219 Ibid page 24.
- 220 Guillemin, Jeanne. *Biological weapons: from the invention of state-sponsored programs to contemporary bioterrorism*. New York: Columbia University Press, 2006. Page 69.
- 221 Ibid page 73.
- 222 Ibid page 93.
- 223 Ibid page 159.
- 224 Yuki, Hidemi, Lloyd Hough, Marc Sageman, Richard Danzig, Rui Kotani, and Terrence Leighton. "Aum Shinrikyo: Insights into How Terrorists Develop Biological and Chemical Weapons." Center for a New American Security. July 20, 2011. Accessed May 1, 2017. <https://www.cnas.org/publications/reports/aum-shinrikyo-insights-into-how-terrorists-develop-biological-and-chemical-weapons>.
- 225 Ibid page 93.
- 226 "Anthrax - Basic Information." *Centers for Disease Control and Prevention*. Centers for Disease Control and Prevention, 01 Sept. 2015. Web. 1 Jan. 2017.
- 227 Tucker, Jonathan B., ed. *Innovation, dual use, and security: managing the risks of emerging biological and chemical technologies*. Cambridge: MIT Press, 2012. Page 125.
- 228 Guillemin, Jeanne. *Biological weapons: from the invention of state-sponsored programs to contemporary bioterrorism*. New York: Columbia University Press, 2006. Page 123.
- 229 Ibid page 142.
- 230 Weiner, Tim. "Soviet Defector Warns of Biological Weapons." *The New York Times*. February 24, 1998. Accessed May 1, 2017. <http://www.nytimes.com/1998/02/25/world/soviet-defector-warns-of-biological-weapons.html?mcubz=1>.
- 231 Ibid page 154.
- 232 Powell, Colin. "Full text of Colin Powell's speech." *The Guardian*. Guardian News and Media, 05 Feb. 2003. Web. 1 Jan. 2017.
- 233 "Nuclear Posture Review." U.S. DEPARTMENT OF DEFENSE. April 2010. Accessed January 1, 2017. <https://www.defense.gov/News/Special-Reports/NPR>.
- 234 "Riegler Report." Riegler Report - Wikisource, the free online library. 1994. Accessed January 1, 2017. https://en.wikisource.org/wiki/Riegler_Report.
- 235 Tucker, Jonathan B., ed. *Innovation, dual use, and security: managing the risks of emerging biological and chemical technologies*. Cambridge: MIT Press, 2012. Page 19.
- 236 Federation of American Scientists. 2014. "Recombinant DNA: From Vaccines to Weapons." Federation of American Scientists. December 1. Accessed January 1, 2017. http://fas.org/biosecurity/education/dualuse/FAS_Jackson/1_B.html.
- 237 Gronvall, Gigi Kwik. "US Competitiveness in Synthetic Biology." *Health Security*. Mary Ann Liebert, Inc., 01 Dec. 2015. Web. 1 Jan. 2017.
- 238 Sell, Tara Kirk, and Matthew Watson. "Federal Agency Biodefense Funding, FY2013-FY2014." *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*. September 2013. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3778993/>.

Belfer Center for Science and International Affairs
Harvard Kennedy School
79 John F. Kennedy Street
Cambridge, MA 02138

www.belfercenter.org

[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI)
Sent: Tuesday, September 04, 2018 8:04 AM
To: [redacted] (CYD) (FBI)
Cc: [redacted] (CYD) (FBI)
Subject: RE: article

b6
b7C

Thanks. We are meeting with a company on Friday that claims to have a solutions. We'll see...

Regards,

[redacted]
FBI Cyber Division

b6
b7C
b7E

[redacted] (desk)
[redacted] (mobile)

-----Original Message-----

From: [redacted] (CYD) (FBI)
Sent: Tuesday, September 04, 2018 7:57 AM
To: [redacted] (CYD) (FBI) [redacted]
Subject: article

b6
b7C
b7E

<https://money.cnn.com/2018/08/08/technology/deepfakes-countermeasures-facebook-twitter-youtube/index.html>

Thanks, [redacted]

[redacted]
Chief, Technology Cyber Intelligence Unit Cyber Engagement & Intelligence Section Cyber Division Federal Bureau of Investigation

b6
b7C
b7E

[redacted] o)
[redacted] c)

[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI)
Sent: Friday, September 14, 2018 8:46 AM
To: [redacted] (CYD) (FBI)
Subject: RE: Article + previous convo

b6
b7C
b7E

Thanks!

[redacted]
Federal Bureau of Investigation || Cyber Division
Desk [redacted]

-----Original Message-----
From: [redacted] (CYD) (FBI)
Sent: Friday, September 14, 2018 8:33 AM
To: [redacted] (CYD) (FBI) <[redacted]>
Subject: FW: Article + previous convo

FYI - I am pinging my [redacted] contacts as well so we can be involved in the response.

b6
b7C
b7E

Regards,

[redacted]
FBI Cyber Division

[redacted] (desk)
[redacted] (mobile)

-----Original Message-----
From: [redacted] (CYD) (FBI)
Sent: Friday, September 14, 2018 7:55 AM
To: [redacted] (CYD) (FBI) <[redacted]>
Subject: FW: Article + previous convo

b6
b7C
b7E

fysa

Thanks, [redacted]

-----Original Message-----
From: [redacted] (CYD) (FBI)
Sent: Friday, September 14, 2018 7:52 AM
To: [redacted] (CYD) (FBI) <[redacted]>
Cc: Karl, Larry D. (CYD) (FBI) <[redacted]>
Subject: Article + previous convo

b6
b7C
b7E

[redacted] sorry if I'm bugging you but I wasn't sure if you were or weren't in the office. If not, let me know who to deal with as it seems [redacted] is out for an extended time.

I saw this article last night and I remember talking to you about if we could get this topic in front of congressional affairs for a staffer's brief. This article calls for info that my unit has, ready to go [redacted]. Thoughts?

U.S. lawmakers call for deepfakes counter measures <https://venturebeat.com/2018/09/13/u-s-lawmakers-call-for-deepfakes-counter-measures/>

Thanks, [redacted]

b6
b7C
b7E

[redacted]

Chief, Technology Cyber Intelligence Unit Cyber Engagement & Intelligence Section Cyber Division Federal Bureau of Investigation

[redacted] (o)
[redacted] (c)

[redacted] (IMD) (CON)

b6
b7C

From: [redacted] (CYD) (FBI)
Sent: Thursday, September 27, 2018 12:04 PM
To: [redacted] (CYD) (FBI)
Subject: Re: Deepfakes meeting

Awesome, thanks!

Regards,

[redacted]

b6
b7C

FBI Cyber Division

[redacted] (desk)
[redacted] (mobile)

----- Original message -----

From: [redacted] (CYD) (FBI)" [redacted] >
Date: 9/27/18 11:09 (GMT-05:00)
To: [redacted] (CYD) (FBI)" [redacted] }
Subject: Deepfakes meeting

b6
b7C
b7E

Hi [redacted]

The Deepfakes meeting that was originally scheduled for today has been rescheduled for next Friday, October 5th, from 1000 to 1100 at [redacted] I've reserved a Bu car and requested the EZPass (so as to avoid most of the fun morning traffic).

[redacted]
Federal Bureau of Investigation || Cyber Division
Desk: [redacted]

b3
b6
b7C
b7E

NotSoFakeApp: Determining the Provenance of AI Created Video Artifacts

Sean M. Futch
Information Security Institute
Johns Hopkins University
Baltimore, MD 21218
sfutch1@jhu.edu

Timothy R. Leschke
Information Security Institute
Johns Hopkins University
Baltimore, MD 21218
tleschk1@jhu.edu

Abstract

The newly released DeepFake technology has problematized the field of digital forensics, specifically in the realm of video and still image analysis. Easy to use and readily available, advanced face-swapping technology based upon AI and machine learning has the potential capability to reduce faith and trust in digital artifacts, as well as greatly hinder forensic validation of said items. By repurposing current analysis programs and using them in tandem, we show that this new avenue of video forgery, while sophisticated, remains susceptible to current forensic techniques.

Keywords

FakeApp, DeepFake, Artificial Intelligence, Forensic Image Processing, ELA, Noise Reduction, Machine Learning, Neural Network

1. Introduction

The recent advent and release of FakeApp, DeepFake and similar machine learning based face-altering programs has created a new issue for the forensic examination of videos: establishing the provenance and authenticity of persons portrayed in them. This is particularly important to digital forensic examiners establishing video authenticity. The FakeApp program allows minimally trained users to quickly face-swap the actors in a video with another person. Unlike previous methods, such as OpenCV, this process alters videos, not just pre-existing images, and utilizes neural network processing to morph a person's face into another video while preserving the original facial expression [1].

This process is enabled by "deep learning", in which the software scans user-selected videos and images, creating datasets for use in the training of Artificial Intelligence (AI). AI in this case refers to advanced machine learning focused upon scanning the selected images to fully map and encode the facial structure of the subject over the target in the final video. Once complete, the program performs an iterative process, encoding a warped version of the image to be stitched into the resultant video, in effect "capturing the essence of a face" that mimics the facial expressions of the target [2]. The software then transposes and overlays the reconstructed image from the datasets onto the face of an individual in a video, stitching and converting the original video into a new artifact.

When done correctly, assuming enough images exist of the subject to create a facial map, the results can be difficult to detect with the naked eye.



Figure 1: Facial Reconstruction of Actress Gal Gadot repurposed for use in pornography video. Image taken from <https://www.madmy.com/fake-porn-video-gal-gadot-highlights-dangers-machine-learning>

2. Related Work

Numerous tools have been developed enabling the forensic analysis of videos and still images. These include open-source/free video forensic tools Video Cleaner, CodecVisa, Forevid, and Kinovea. Forensically and FotoForensics are free web based image forensic programs offering multiple analysis tools.

None of the above tools were created to detect videos created by FakeApp or similar programs, but our approach utilized them to determine whether altered videos contain forensic markers indicating the presence of alterations. Out of these tools, Videocleaner, Forensically, and FotoForensics stood out, allowing a frame-by-frame detailed examination of the

video file, and further analysis of captured still images from the altered movie.

Our choice to use freely available tools was due to time constraints as well as approaching the problem set from the lowest common denominator: cost. Current proprietary tools and hosted services are expensive, may require additional time to learn, and are not always an option for every forensic examiner. With this approach, we intend to enable every analyst with the resources to determine video provenance and authenticity.

3. Forensic Analysis on Videos and Pictures

The forensic examination of digital images has become a commonplace phenomenon. Forensic Image Processing (FIP) is utilized to extract specific information from videos or pictures that are typically incomplete, noisy, or under/over exposed [3]. Two common techniques in FIP are Error Level Analysis and Image Noise Analysis.

Error Level Analysis (ELA) is the procedure of examining the error levels in an image to detect any digital manipulation. Error levels are introduced in the saving of images (both moving and still). In an unaltered image, error levels should be uniform; significantly different error rates show possible manipulation. Image noise is an inherent artifact of the image capture process when producing video or still images. It may come from physical sources or from the conversion of image information from electrical signals to digital data [4].

Image noise analysis seeks to find the variances that occur when an image is modified. Applying a noise reduction filter assists in discerning alterations in a visually identifiable manner.

4. Test Data and Experiment

For this experiment, Dr. Sven Charleer provided a sample of altered and unaltered videos, which we utilized as an initial data set, to use in determining forensic differences[5]. We then compiled a small number of altered videos posted to YouTube as a test sample to verify results.

The original sample consisted of 3 short videos (1 Anne Hathaway interview on the Tonight Show, 2 clips of Anne Hathaway in the movie Get Smart), and the resultant, edited videos. The test sample consisted of two altered videos posted to YouTube, specifically Nicolas Cage replacing Amy Adams in a scene from Man of Steel, and Nicolas Cage replacing Harrison Ford in a scene from Raiders of the Lost Ark.

We divided the forensic examination tools into two sets: video and still image. For video, we used Videocleaner 5.0. This freely available instrument offers numerous means and

methods to collect forensic data and perform video authenticity analysis. Of particular use was the ability to capture video frames in a TIFF format, preserving the quality of the image for further analysis. For the experiment, we captured still images of specific frames from each video to conduct a side-by-side comparison in order to detect variances in the modified versions.

To analyze the captured TIFF images, we used two online resources: Fotoforensics and Forensically Beta. Both tools provided ELA but only Forensically provided Noise Analysis and other forensic tools, as well as scalable settings to fine-tune results.

To ensure consistency, all video comparisons and test data sets were sampled at the same resolution, and after experimenting with variable settings to determine forensic markers in FakeApp videos, all variable ELA and Noise Amplitude settings were standardized. Forensically Settings: **Error Level Analysis** - JPEG Quality 100/100, Error Scale 20/100, Opacity 1.0. **Noise Analysis** - Noise Amplitude 20/100, Equalize histogram checked, Magnifier Enhancement None, Opacity 1.0.

5. Results

Viewed as a cohesive whole, the quality of the FakeApp videos depend greatly upon two variables: the familiarity of the user with the program and the number and quality of the images that are used to create the modified face in the target video.

Though the AI learning and facial reconstruction tools are impressive, they are not infallible. Slowing or stopping the video and conducting a frame by frame analysis highlighted markers that can be used by an investigator to determine authenticity.

Variances in physical movement is particularly noteworthy. The program will make every attempt to follow the target video's facial and body actions, but in both sample and test data sets, the results were imperfect. Additionally, when viewed as a still image, it is possible to identify alterations when compared to an original.

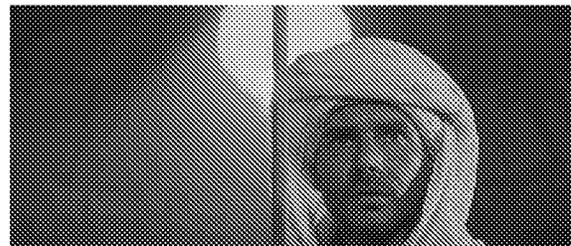


Figure 2: Original image from Raiders of the Lost Ark.

The facial structure, though close to the target's, is softened and shows fuzzing around the edges. Lighting effects, such as sweat or light reflections, are not faithfully rendered.

Additionally, the video/image quality near the outside edges of the facial structure is diffused and has a blurred quality when compared to surrounding artifacts in the image. This is particularly noticeable when viewing as a video.

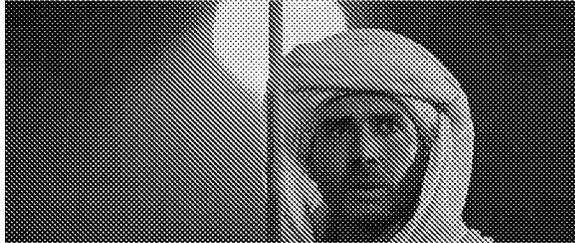


Figure 3: FakeApp recreation of Figure 2. Note the softening of the edges around the boundaries of the facial structure and lack of sweat/light reflection.

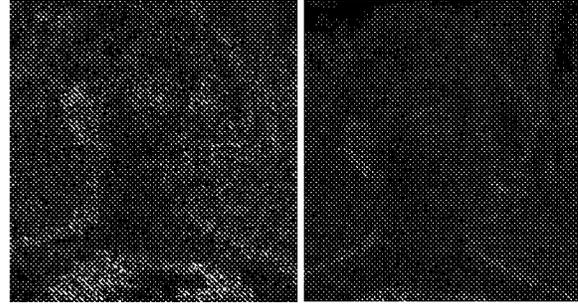


Figure 4: Original image from Get Smart.



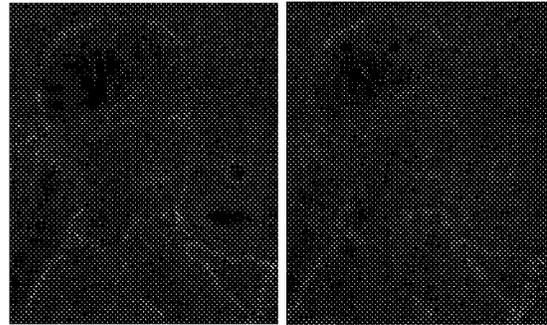
Figure 5: FakeApp recreation of Figure 4. Note similar softening facial structure as in previous forged image and inexact expression.

Moving into Error Level Analysis, more differences become apparent. ELA visualization in the video was problematic due to the difficulty in highlighting levels in moving objects, forcing the capture of still images to compare. Once complete, the modifications become more apparent.



Figures 6 & 7: ELA analysis of original (left) and modified (right) still images taken from Anne Hathaway Tonight Show interview.

In the altered image, there is a notable lack of definition in facial structure, as well as a flattened, almost unnoticeable quality for the nose and mouth. This quality was consistent throughout further ELA analysis.



Figures 8 & 9: ELA analysis of original (left) and modified (right) still images taken Figures 4 and 5, respectively.

Measuring the image noise also proved fruitful. Analyzing the original image for noise quality showed standard and consistent results. When compared to FakeApp created content, the results were notable, with similar results previously shown in ELA.

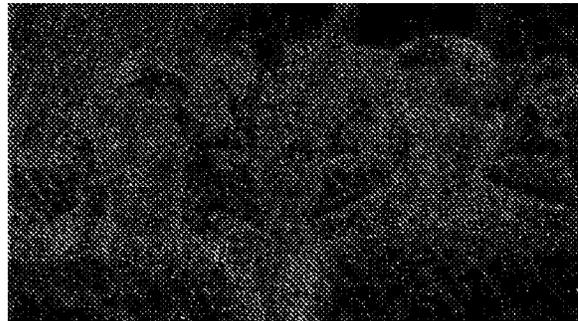


Figure 10: Noise level analysis of still image taken from Get Smart. Note the consistent level of noise throughout image.

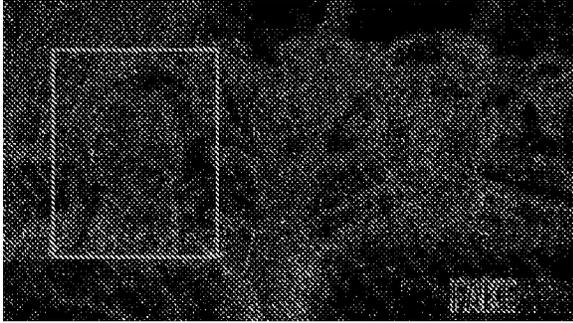


Figure 11: Noise level analysis of still image taken from FakeApp version of Figure 10. Note the inconsistent level of noise (dark areas) shown in highlighted area.

Previous research into image forensics and noise level analysis have shown that Image forgeries with computer graphics can have unusual noise, or no noise [4]. This phenomenon remained consistent throughout the testing of this new video forgery technology.

6. Conclusion

Using freely available technology we identified and delineated noticeable forensic markers that are inherent in the application of new “deep learning” facial reconstruction and forgery applications. As this technology becomes more sophisticated and commonplace, our research will help digital forensic examiners successfully identify altered videos more efficiently.

FakeApp and other DeepFake technology, though effective, appear to lack both the ability to detect mistakes that occur during the stitching process [1] and the error correction necessary to fix them. This was shown in the fuzzing of areas near the target face as well as the inexact facial expression matching seen in altered videos.

This effect likely occurs because the images used to train the AI were taken in a different context than the target video, causing the program to extrapolate imperfectly. As the typical forensic examiner will most likely lack an original, unaltered video for comparison, the facial fuzzing, ELA variances, and noise level analysis will provide the best possible indicators that the video under examination has been altered.

Though this technique has proven effective in this experiment, it must be noted that the investigator’s ability to discern a video forgery can depend on the same two variables with which the video producer must contend: user familiarity with the program, and the number and quality of the images used to create the modified face in the target video.

7. Future Work

Our future goals and recommendations for this branch of forensic investigation include the development of

video analysis software capable of detecting, tracking, and analyzing the facial fuzzing indicative of altered videos, as well as the ability to create frame-by-frame stills that automatically contain ELA and noise analysis.

Acknowledgements

We would like to thank Dr. Sven Charleer for his generosity in sharing his work and data sets, without which this paper would not have been written.

8. References

- [1] Zucconi, A. (2018). *An introduction to DeepFakes*. Retrieved 25 March 2018, from <https://www.alanzucconi.com/2018/03/14/introduction-to-deepfakes/>.
- [2] Oberoi, G. (2018). *Exploring DeepFakes*. Retrieved 20 March 2018, from <https://hackemoon.com/exploring-deepfakes-20c9947c22d9>.
- [3] Borengasser, M. (2014). *Introduction to forensic image processing*. Retrieved 08 April 2018, from <https://www.forensicmag.com/article/2014/06/introduction-forensic-image-processing>.
- [4] Julliard, T., Nozick, V. & Talbot, H. (Unknown). *Image noise and digital image forensics*. Retrieved 08 April 2018, from <https://pdfs.semanticscholar.org/2618/03c53b6d3409bae532d686f3655c97f017ad.pdf>.
- [5] Charleer, S. (2018). *Family fun with deepfakes, or how I got my wife onto the Tonight Show*. Retrieved 20 March 2018, from <https://towardsdatascience.com/family-fun-with-deepfakes-or-how-i-got-my-wife-onto-the-tonight-show-a4454775c011>.

[redacted] **IMD) (CON)**

From: [redacted] (OTD) (FBI) b6
Sent: Tuesday, October 09, 2018 3:36 PM b7C
To: [redacted] (CYD) (FBI) b7E
Cc: [redacted] (OTD) (FBI); [redacted] (OTD) (FBI)
Subject: FW: DeepFakes Congressional Request [redacted] -- ~~SECRET//NOFORN~~
Attachments: [redacted]

Classification: ~~SECRET//NOFORN~~

Classified By: [redacted]
Derived From: Multiple Sources
Declassify On: 50X1-HUM
=====

b6
b7C

[redacted]
[redacted] reached out to my UC [redacted] about working this jointly with OTD, and it was assigned to me. Could you fill me in on what all has been done, where things stand, and how/if we can assist or help? Also, I wasn't sure if [redacted] had been brought into this? Let me know what you need. Thanks.

b6
b7C

[redacted]

From: [redacted] (OTD) (FBI)
Sent: Tuesday, October 09, 2018 3:06 PM
To: [redacted] (OTD) (FBI); [redacted]
Cc: [redacted] (OTD) (FBI); [redacted]
Subject: FW: DeepFakes Congressional Request [redacted] --- ~~SECRET//NOFORN~~

b6
b7C
b7E

Classification: ~~SECRET//NOFORN~~

Classified By: [redacted]
Derived From: Multiple Sources
Declassify On: 50X1-HUM
=====

b6
b7C

[redacted]

Can you please plan to work this jointly with [redacted]? You will need to work closely with [redacted] on it.

Thank you!

[redacted]

b6
b7C
b7E

[redacted]

Unit Chief - Technical Intelligence Unit
FBI Operational Technology Division

[redacted]
[redacted]
NSTG [redacted]

From: [redacted] (CYD) (FBI)
Sent: Tuesday, October 09, 2018 2:07 PM
To: [redacted] (OTD) (FBI); [redacted]
Cc: [redacted] (OTD) (FBI); [redacted]
Subject: FW: DeepFakes Congressional Request [redacted] --- ~~SECRET//NOFORN~~

b6
b7C
b7E

Classification: ~~SECRET//NOFORN~~

Classified By: [redacted]
Derived From: Multiple Sources
Declassify On: 50X1-HUM

b6
b7C

[redacted] per conversation, here is the write up from the Deep Fakes meeting on Friday. My IA POC for this is [redacted] No issues here working together on the Bu's part of the response.

b6
b7C

Thanks, [redacted]

From: [redacted] (CYD) (FBI)
Sent: Tuesday, October 09, 2018 12:02 PM
To: [redacted] (CYD) (FBI); [redacted] (TD) (FBI); [redacted]
Cc: [redacted] (CYD) (FBI); [redacted]
Subject: DeepFakes Congressional Request [redacted] -- ~~SECRET//NOFORN~~

b6
b7C
b7E

Classification: ~~SECRET//NOFORN~~

Classified By: [redacted]
Derived From: Multiple Sources
Declassify On: 50X1-HUM

b6
b7C

=====
TRANSITORY RECORD

[redacted]

b6
b7C
b7E

Per our discussion this morning.

Regards,

[redacted]
Technology Cyber Intelligence Unit
FBI Cyber Division

Open: [redacted]
Secure: [redacted]

=====
Classification: ~~SECRET//NOFORN~~

=====
Classification: ~~SECRET//NOFORN~~

=====
Classification: ~~SECRET~~//~~NOFORN~~

=====
Classification: ~~SECRET~~//~~NOFORN~~

[redacted] (MD) (CON)

b6
b7C

From: Karl, Larry D. (CYD) (FBI)
Sent: Tuesday, August 28, 2018 2:25 PM
To: [redacted] (CYD) (FBI); Mckinsey, William G. (CJIS) (FBI)
Cc: [redacted] (CYD) (FBI); [redacted] (CYD) (FBI)
Subject: RE: FACE SWAPPING

Great collaboration. Please add me to the lists as well. Thx!

Larry Karl
Section Chief
FBI Cyber Division
Cyber Engagement and Intelligence Section (CEIS)
Desk: [redacted]
Cell: [redacted]
Unclass email: [redacted]

b7E

From: [redacted] (CYD) (FBI)
Sent: Monday, August 20, 2018 11:52 AM
To: Mckinsey, William G. (CJIS) (FBI) [redacted]
Cc: [redacted] (CYD) (FBI) [redacted]; Karl, Larry D. (CYD) (FBI) [redacted]; [redacted] (CYD) (FBI)
[redacted]
Subject: RE: FACE SWAPPING

b6
b7C
b7E

Will do, sir. Thank you.

Regards,

[redacted]
FBI Cyber Division

b6
b7C
b7E

[redacted] (desk)
[redacted] (mobile)

From: Mckinsey, William G. (CJIS) (FBI)
Sent: Monday, August 20, 2018 9:54 AM
To: [redacted] (CYD) (FBI) [redacted]
Cc: [redacted] (CYD) (FBI) [redacted] (OTD) (FBI) [redacted]; [redacted] (OTD) (FBI) [redacted] [redacted] (CJIS)
[redacted] (FBI) [redacted]
Subject: RE: FACE SWAPPING

b6
b7C
b7E

[redacted]

CJIS is very extremely interested in this subject. We have a big effort underway to [redacted]
[redacted]

b7E

We'll add you to our communication lists; pls add us (Myself, [redacted]) to yours. Thank you.

b6
b7C
b7E

MCK
William G. McKinsey
Section Chief, FBI/CJIS
[O]
[M]

From: [redacted] (OTD) (FBI)
Sent: Monday, August 20, 2018 8:04 AM
To: [redacted] (CYD) (FBI); [redacted]; [redacted]; Mckinsey, William G. (CJIS) (FBI); [redacted] (CJIS) (FBI); [redacted]
Cc: [redacted] (CYD) (FBI); [redacted]; [redacted] (OTD) (FBI); [redacted]
Subject: Re: FACE SWAPPING

b6
b7C
b7E

Thank you, [redacted]

I am adding [redacted] to this thread, as he is Chief of the OTD Technical Intelligence Unit and has been looking at Digital identity-related issues with me for several years.

b6
b7C
b7E

[redacted]
FBI - OTD - TODB
Building 27958A, Pod E
Quantico, VA 22135
[redacted]

From: [redacted] (CYD) (FBI)
Sent: Monday, August 20, 2018 7:51 AM
To: [redacted] (OTD) (FBI); [redacted] Mckinsey, William G. (CJIS) (FBI); [redacted] (CJIS) (FBI)
Cc: [redacted] (CYD) (FBI)
Subject: RE: FACE SWAPPING

b6
b7C

Good morning all,

[redacted] thank you for including me in the discussion. For everyone's awareness, I have been [redacted] on this technology for the last 9 months or so in a series of intelligence products. Most recently [redacted] to detail where the [redacted]

b6
b7C
b7E

I would love to include [redacted] in that report, to demonstrate how FBI in particular is maintaining awareness, or planning a capability build-out for this issue. Happy to share the Bulletin and discuss on other systems at everyone's convenience.

Regards,

[redacted]

b6
b7C

FBI Cyber Division

[redacted] (desk)
[redacted] (mobile)

b7E

From: [redacted] (OTD) (FBI)
Sent: Monday, August 20, 2018 7:41 AM
To: [redacted] <[redacted]>; Mckinsey, William G. (CJIS) (FBI) <[redacted]>; [redacted] <[redacted]> (CJIS) (FBI)
Cc: [redacted] (CYD) (FBI) <[redacted]>
Subject: Re: FACE SWAPPING

b6
b7C
b7E

As a further point regarding this issue...

[redacted] of Cyber Division (copied) has also been tracking this issue and was also at the DARPA MEDIFOR PI meeting last month.

[redacted] is not alone in being concerned with this threat, so to the extent that we decide to formally track it and develop responses, we should be sure to include Cyber (and others) in the discussion.

b6
b7C
b7E

[redacted]
FBI - OTD - TODB
Building 27958A, Pod E
Quantico, VA 22135
[redacted]

From: [redacted] <[redacted]>
Sent: Monday, August 20, 2018 7:28 AM
To: Mckinsey, William G. (CJIS) (FBI); [redacted] (CJIS) (FBI); [redacted] (OTD) (FBI)
Subject: Re: FACE SWAPPING

b6
b7C
b7E

- o DEEPFAKES: Fake America great again (MIT Tech Review, 8/17) Inside the race to catch the worryingly real fakes that can be made using AI. Perhaps the greatest risk is that the technology will further undermine truth and objectivity.

On 08/16/18 01:33 PM, [redacted] <[redacted]> wrote:

b6
b7C
b7E

MIT is working on detection...

This algorithm automatically spots "face swaps" in videos (MIT Tech Review, 4/10) But the same system can be used to make better fake videos that are harder to detect.

Very good, in-depth article on the national security/democracy implications!

Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy? (Lawfare, 2/21)

On 08/16/18 01:25 PM, [redacted] [redacted] wrote:

b6
b7C

looks like its still in its infancy... but is getting more prevalent. "Experts" are saying it's 1-2 years away from being really good. DARPA is already working on detection software. I was afraid to click on too many articles - since it's being used mostly for porn. these were ok.

Good basic video explaining the tech...

Face-swapping videos could lead to more 'fake news' (Business Insider, 4/13/18)

I never said that! High-tech deception of 'deepfake' videos (Phys.org, 7/2) This technology uses facial mapping and artificial intelligence to produce videos that appear so genuine it's hard to spot the phonies...Realizing the implications of the technology, the U.S. Defense Advanced Research Projects Agency is already two years into a four-year program to develop technologies that can detect fake images and videos. ..it's easy to foresee a nation state using them for nefarious activities against the U.S....

In An Era of Fake News, Advancing Face-Swap Apps Blur More Lines (NPR, 2/3/18)

Fake video news is coming, and this clip of Obama 'insulting' Trump shows how dangerous it could be (CNBC, 4/17) A BuzzFeed PSA seems to show former President Barack Obama saying disparaging things about President Donald Trump, but it's actually a PSA to show how easy it is to manipulate video and spread misinformation.

How to identify if an online video is fake (New Statesman, 2/14) As "deep fakes" raise concerns, everyone needs to equip themselves with the knowledge to spot a fraudulent video.

The deepest fake: how new tech will test our belief in what we see (Sydney Morning Herald (AU), 5/4)

On 08/16/18 09:34 AM, "Mckinsey, William G. (CJIS) (FBI)" [redacted] wrote:

b6
b7C
b7E

GUYS,

Pls follow this topic. This could require urgent action our part if it is real. Is anyone working on prevention or detection. I though the reaction of the Privacy Group was something else – Shut down Face Book and face recognition.

Let's discuss what if anything we should be doing about this challenge. [redacted] will help us track the topic in the media.

MCK
William G. McKinsey
Section Chief, FBI/CJIS

[O]
[M] [redacted]

b6
b7C
b7E

From: [redacted] (CJIS) (FBI)
Sent: Thursday, August 16, 2018 6:53 AM
To: Mckinsey, William G. (CJIS) (FBI) <[redacted]>
Subject: RE: face-swapping article fyi

Fake videos? Computer program generates eerily realistic fake footage

Set to be unveiled at a computer animation festival in Vancouver, the software can also tweak head and torso poses, eye movements and background details to create more convincing fakes

<http://www.foxnews.com/tech/2018/08/15/fake-videos-computer-program-generates-eerily-realistic-fake-footage.html>

From: [redacted] (CJIS) (FBI)
Sent: Tuesday, March 13, 2018 7:38 AM
To: Mckinsey, William G. (CJIS) (FBI) <[redacted]>
Subject: RE: face-swapping article fyi

b6
b7C
b7E

How AI-generated videos could be the next big thing in fake news (Fox News) It's difficult to assess the national security risk or potential for disruption that is presented by the threat of AI-built fake videos.

From: Mckinsey, William G. (CJIS) (FBI)
Sent: Sunday, January 28, 2018 12:12 PM
To: [redacted] (CJIS) (FBI) <[redacted]>
Subject: Re: face-swapping article fyi

b6
b7C
b7E

[redacted] - I googled face swapping and learned a lot. I have [redacted] and [redacted] working on it. Thanks.

Pls follow [redacted] closely. It could put us out of business.

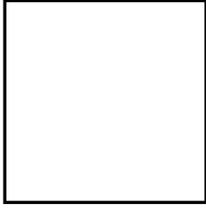
MCK

----- Original message -----
From: [redacted] (CJIS) (FBI)" <[redacted]>
Date: 1/25/18 6:41 AM (GMT-07:00)
To: "Mckinsey, William G. (CJIS) (FBI)" <[redacted]>
Subject: face-swapping article fyi

b6
b7C
b7E

FYI

I didn't want to click on the full article... This is a blurb on Slashdot.com that I took a screenshot of so the links can't be clicked. It has enough info in it.
Made me think that if they are doing this for trivial crap, then what is being done to surveillance video or other facial recognition images by others with better tools.



From: [redacted] <[redacted]>
Sent: Monday, October 15, 2018 11:20 AM
To: Mckinsey, William G. (CJIS) (FBI)
Cc: [redacted] (CYD) (FBI); [redacted] (OTD) (FBI); [redacted] (OTD) (FBI); [redacted] (CJIS) (FBI); [redacted] (CJIS) (FBI); [redacted] (CYD) (FBI)
Subject: RE: FACE SWAPPING

[Magic Leap's new AI assistant looks alarmingly human \(CNN\)](#) ... could have far-reaching implications for society... This breakthrough raises ethical questions, too.
[Deepfake Videos Are Ruining Lives. Is Democracy Next? \(WSJ\)](#) Moving Upstream explores the dark side of sophisticated video fakery. Researchers have developed forensic methods to detect fakes.

On 10/12/18 01:27 PM, [redacted] <[redacted]> wrote:
[Sen. Hassan Questions FBI Director on 'Deepfakes' \(NHPR, 10/10\)](#) Senator Maggie Hassan (D-NH) wants to make sure the FBI has the authority and tools it needs to crack down on so-called "deepfakes." Wray said the FBI already has a # of its S&T folks burrowing in on this issue.

[Deepfakes helped Charli XCX imitate the Spice Girls in her latest music video \(The Verge, 10/11\)](#) A first for AI face-swapping algorithms – as a special effect.

On 10/09/18 11:03 AM, [redacted] <[redacted]> wrote:
 related side article
[Chinese investment into computer vision tech and AR surges as US funding dries up \(CTOvision, 10/8\)](#)

b6
b7C
b7E

On 10/04/18 11:07 AM, [redacted] <[redacted]> wrote:
 related article that ties in with fake vids...
 • [PERSPECTIVE: Shadow Banning and Astroturfing -- Understanding the New War on Ideas \(HS Today\)](#) How Automation Suppresses or Promotes Information; How Analysts Can Stay Updated with Emerging Manipulative Techniques. The key to understanding many of these techniques is that they are not new at all; they've only become automated. For analysts, it's important to understand new and emerging manipulation techniques, regardless of the format, in order to mitigate their effects on information collection, research, and analysis.

On 09/25/18 10:42 AM, [redacted] <[redacted]> wrote:
 related article...
[Machine Learning Confronts the Elephant in the Room \(Quanta Mag\)](#) A visual prank exposes an Achilles' heel of computer vision systems: Unlike humans, they can't do a double take. The result takes place in the field of computer vision...

On 09/24/18 11:54 AM, [redacted] <[redacted]> wrote:

b6
b7C
b7E

just one article of possible related interest... or possibly not...
8 New Technologies Changing Video Production (eWeek) Technologies transforming data centers in other sectors are enabling tv/video teams to produce, edit, finish and deliver clearer, crisper and more lifelike content faster and at lower cost.

On 09/21/18 10:03 AM, [redacted] <[redacted]> wrote:

Amnesty International toils to tell real videos from fakes (HSNW, 9/20)
When it launched a probe this year into police crackdowns against Russian protesters, one of its research methods was to collect and verify videos posted on social media from across Russia since 2012... created its Digital Verification Corps (DVC).

RELATED: Someone Watching? Tenable Says Hackers Can Access, Alter Surveillance Footage (MeriTalk, 9/18) Zero day in popular video surveillance tech goes public, unpatched (Cyber Scoop, 9/17)

On 09/18/18 11:15 AM, [redacted] <[redacted]> wrote:

House Lawmakers Urge IC to Look Into 'Deep Fake' Tech (Exec Gov, 9/17) The signatories asked that the report be submitted by 12/14 at the latest.

Congress wants the IC to weigh in on how to counter 'deepfakes' (Fed Scoop, 9/17)

DHS Can Neither Confirm Nor Deny It Has Records on Deepfakes (Motherboard, 9/18 – use Firefox to open) As DARPA researchers work on identifying manipulated videos, and lawmakers call for an IC report, the DHS is staying tight-lipped on deepfakes.

b6
b7C
b7E

On 09/14/18 10:26 AM, [redacted] <[redacted]> wrote:

Lawmakers want US intelligence assessment on fake videos (AP, 9/13)

Bipartisan trio asks US intelligence to investigate 'deepfakes' (The Hill, 9/13)

Congress seeks probe of deepfakes (AXIOS, 9/13)

On 09/13/18 12:41 PM, '[redacted]' wrote:

Researchers Come Out with Yet Another Unnerving, New Deepfake Method (Gizmodo, 9/11) Carnegie Mellon researchers have figured out a way to automatically transfer the "style" of one person to another. *Facial Expressions*

On 09/06/18 12:02 PM, '[redacted]' wrote:

Can AI Get a Bead on Faked Fingerprints? (MeriTalk, 9/6) ...The FBI wants to use artificial intelligence to add a new layer to its NGI system, specifically to counteract the increasingly common practice of criminals altering their fingerprints...AI techniques in machine learning and deep learning have made significant improvements in recent years at performing tasks they were specifically trained to do, but still struggle to learn from examples and take the next steps on their own...

b6
b7C
b7E

On 08/28/18 01:11 PM, '[redacted]' wrote:

related article

Semantic cache for AI-enabled image analysis (phys.org, 8/28)

...Edge computing, as this is known, not only reduces the strain on bandwidth but also reduces latency of obtaining intelligence from raw data. However, availability of resources at the edge is limited due to the lack of economies of scale that make cloud infrastructure cost-

effective to manage and offer...

The potential of edge computing is nowhere more obvious than with video analytics (surveillance)...

In our Hot Edge 2018 Conference Paper "Shadow Puppets: Cloud-level Accurate AI Inference at the Speed and Economy of Edge," our team at IBM Research – Ireland experimentally evaluated the performance of one such AI workload, object classification, using commercially available cloud-hosted services. The best result we could secure was a classification output of 2 frames per second which is far below the standard video production rate of 24 frames per second. Executing a similar experiment on a representative edge device (NVIDIA Jetson TK1) achieved the latency requirements but used up most of the resources available on the device in this process...

On 08/23/18 02:06 PM,

[Redacted]
[Redacted]

wrote:

vendors to possibly be aware of/watch what they are doing...mostly in India, with some in India and the US.

The 10 "computer vision" startups you

b6
b7C
b7E

need to watch out for
(Your Story, 8/23)

On 08/21/18 11:39

AM, [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] wrote:

All,

Below are links to OSINT media sources to scan... just for your awareness, in case you haven't seen them. I was researching a little this morning.

Anything I find will be from the public domain (nothing classified). It's not all directly related to face swapping, altered fingerprints, or AI & Biometrics, but could be gleaned for related ideas/information. I hope a few items help in your strategic planning.

•

↑

b6
b7C
b7E

I
N
B
I
O
M
E
T
R
I
C
S
A
N
D
S
E
C
U
R
I
T
Y
=

Z
C
U
R
I
E
N
T
B
U
S
I
N
E
S
S
A
P
P
L
I
C
A
T
I
O
N

s (t e c h e m e r g e n c e , 8 / 2 0)

•

19 THE TECHNICAL ASSISTANCE RECURRING

U
N
D
E
R
S
T
A
N
D
I
N
G
(
d
z
o
n
e
,
8
/
1
7
)

8
B
I
O
M
E
T
R
I
C
S
...
3
V
R
,
A
f
f
e
c
t
i
v
a

,
A
g
n
i
t
i
o
,
F
a
c
e
F
i
r
s
t
,
S
e
n
s
o
r
y
,
S
y
n
q
e
r
a
,
a
n
d
T
a
h
z
o
o
a
r
e
a
l
l
b

i
o
m
e
t
r
i
c
s
c
o
;
s
w
o
r
k
i
n
g
h
a
r
d
t
o
d
e
v
e
l
o
p
t
h
i
s
a
r
e
a
.

1
8
I
M
A
G
E

R
E
C
O
G
N
I
T
I
O
N
.
.
l
i
c
e
n
s
e
p
l
a
t
e
s
a
n
d
f
a
c
e
s
.

*These articles
all relate to
IBM
weaponizing
AI-enabled
malware using
facial
recognition.*

•
M
O
W
W
e
r

D
O
N
I
Z
E
M
A
I
L
E
T
E
R
E
A
T
E
S
A
N
D
N
E
W
B
I
R
E
D
O
P
C
Y
C
L
E
T
A
T
T
A
C
K
S
(
T
e
c
h
R
e
p
u
b
l
i

c
,
8
/
1
6
)
I
B
M
s
e
c
u
r
i
t
y
r
e
s
e
a
r
c
h
e
r
s
d
i
s
c
o
v
e
r
e
d
i
n
v
a
s
i
v
e
a
n
d
t

a
r
g
e
t
e
d
A
I
-
p
o
w
e
r
e
d
c
y
b
e
r
-
a
t
t
a
c
k
s
t
r
i
g
g
e
r
e
d
b
y
g
e
o
l
o
c
a
t
i
o

T
e
c
h
R
e
p
u
b
l
i
c
,
8
/
1
5
)
I
B
M
s
e
c
u
r
i
t
y
r
e
s
e
a
r
c
h
e
r
s
d
e
m
o
n
s
t
r
a
t
e

h
o
w
n
e
w
A
I
-
p
o
w
e
r
e
d
F
R
T
c
a
n
t
r
i
g
g
e
r
m
a
l
w
a
r
e
l
u
r
k
i
n
g
w
/
i
n
c
o
m
m

A
R
T
I
C
L
E
S
:
A
I
I
n
t
e
r
f
a
c
e
(
B
i
o
m
e
t
r
i
c
U
p
d
a
t
e
:
8
/
9
)

These articles deal more with AI combined with facial recognition, than with fingerprinting, but still could provide useful insight.

•

T
O
M
O
R
O
W
.
.
s
I
n
t
r
e
m
i
s
e
n
t
M
a
l
w
a
r
e
W
.
i
l
l
A
t
t
a
c
k
W
h
e
n
t
S
e
e
s
Y
O

U
R
R
E
C
E
(
D
e
f
e
n
s
e
O
n
e
,
8
/
1
4
)
I
B
M
r
e
s
e
a
r
c
h
e
r
s
h
a
v
e
i
n
j
e
c
t
e
d
v
i

r
u
s
e
s
w
i
t
h
n
e
u
r
a
l
n
e
t
s
,
m
a
k
i
n
g
t
h
e
m
s
t
e
a
l
t
h
i
e
r
a
n
d
p
r
e
c
i
s
e
l

y t a r g e t a b l e



M
i
n
f
o
r
m
a
t
i
o
n
a
b
o
u
t
t
h
e
a
b
o
v
e
r
l
a
p
p
i
n
g
o
f
t
h
e
t
w
o
s
e
t
s
i
s
d
e
s
c
r
i
b
e
d
i
n
t
h
e
p
r
e
v
i
o
u
s
p
a
r
a
g
r
a
p
h

I
O
R
T
O
P
I
C
O
F
I
N
T
E
R
N
E
T
I
N
F
O
R
M
A
T
I
O
N
S
(
B
I
O
M
E
T
R
I
C
U
P
D
A
T
E
,
5
/
2
9
)

This main article listed out examples of biometrics (face) and AI (bullets below):

BR
IEF
:
Art
ific
ial
Int
elli
gen
ce
Be
co
me
s_a
Sm
art
Bet
(M
obi
le
ID
Wo
rld,
7/2
7)



I
N
T
E
R
F
O
L
D
e
i
v
e
s
i
n
t
a
i
.

Reprints (Mobile ID World, 7/13) They've ended
lived into district

i
b
u
t
e
d
l
e
d
g
e
r
t
e
c
h
n
o
l
o
g
y
(
b
l
o
c
k
c
h
a
i
n
)
...
a
n
d
f
a
c
i
a
l
r
e
c
o
g
n
i

t
i
o
n
...
I
N
T
E
R
P
O
L
s
a
y
s
p
a
r
t
i
c
i
p
a
n
t
s
c
a
l
l
e
d
f
o
r
a
f
o
l
l
o
w
-
u
p
m
e
e

t
i
n
g
t
o
f
o
c
u
s
o
n
t
h
e
s
e
i
s
s
u
e
s



Q
u
e
r
i
e
s
t
i
o
n
s
r
e
l
a
t
e
d
t
o
t
h
e
s
e
i
s
s
u
e
s

l
i
s
t
r
e
s
s
e
s
W
o
r
l
d
-
s
e
M
o
s
t
V
a
l
u
a
b
l
e
A
l
l
i
n
f
o
r
m
a
t
i
o
n
(
F
i
n
d
B
i
o
m
e
t
r
i
c
s
,
4
/

i
c
U
p
d
a
t
e
:
4
/
2
7
)
T
h
e
B
o
a
r
d
w
i
l
l
m
e
e
t
t
w
i
c
e
a
y
e
a
r
t
o
d
i
s
c
u
s
s
e
t

h
i
c
a
l
i
m
p
l
i
c
a
t
i
o
n
s
o
f
A
I
-
p
o
w
e
r
e
d
t
e
c
h
n
o
l
o
g
i
e
s
b
e
i
n
g
d
e
v
e
l

o
p
e
d
b
y
A
x
o
n
a
n
d
w
i
l
l
b
e
c
o
m
p
o
s
e
d
o
f
e
x
p
e
r
t
s
f
r
o
m
v
a
r
y
i
n
g
f
i
e
l

d
s
i
n
c
l
u
d
i
n
g
A
I
:
c
o
m
p
u
t
e
r
s
c
i
e
n
c
e
:
p
r
i
v
a
c
y
:
l
a
w
e
n
f
o
r
c
e
m
e
n

t
a
(
B
i
o
m
e
t
r
i
c
U
p
d
a
t
e
,
6
/
1
3
)

This article focuses more on the thousands of tiny brushstrokes in art than fingerprints (the art world has “gone CSI”), but could still be gleaned for useful intelligence or to potentially reach out to researchers (at Rutgers U)?

•
I
b
e
n

.....

8
8
8
8
(
T
h
e
G
u
a
r
d
i
a
n
,
8
/
6
)

On 08/20/18
09:54 AM,
"Mckinsey,
William G.
(CJIS) (FBI)"

[Redacted]

[Redacted]

wrote:

[Redacted]

b6
b7C
b7E

CJIS is very extremely interested in this subject. We have a big effort underway to [Redacted]

[Redacted]

We'll add you to our communication lists; pls add us (Myself, [Redacted]
[Redacted]) to yours. Thank you.

[Redacted]

From: [redacted] (CYD) (FBI)
Sent: Monday, August 20, 2018 7:51 AM
To: [redacted] (OTD) (FBI); [redacted] Mckinsey, William G. (CJIS) (FBI); [redacted] (CJIS) (FBI)
Cc: [redacted] (CYD) (FBI)
Subject: RE: FACE SWAPPING

b6
b7C

Good morning all,

[redacted] thank you for including me in the discussion. For everyone's awareness, I have been [redacted] on this technology for the last 9 months or so in a series of intelligence products. Most recently [redacted] [redacted] to detail where [redacted] [redacted]

b6
b7C
b7E

I would love to include [redacted] in that report, to demonstrate how FBI in particular is maintaining awareness, or planning a capability build-out for this issue. Happy to share the Bulletin and discuss on other systems at everyone's convenience.

Regards,

[redacted]

FBI Cyber Division

b6
b7C
b7E

[redacted] (desk)
[redacted] (mobile)

From: [redacted] (OTD) (FBI)
Sent: Monday, August 20, 2018 7:41 AM
To: [redacted]; Mckinsey, William G. (CJIS) (FBI); [redacted]; [redacted] (CJIS) (FBI); [redacted]
Cc: [redacted] (CYD) (FBI); [redacted]
Subject: Re: FACE SWAPPING

b6
b7C
b7E

As a further point regarding this issue...

[redacted] of Cyber Division (copied) has also been tracking this issue and was also at the DARPA MEDIFOR PI meeting last month.

[redacted] is not alone in being concerned with this threat, so to the extent that we decide to formally track it and develop responses, we should be sure to include Cyber (and others) in the discussion.

b6
b7C
b7E

[redacted]

FBI - OTD - TODB

Building 27958A, Pod E

Quantico, VA 22135

[redacted]

From: [redacted] <[redacted]>
Sent: Monday, August 20, 2018 7:28 AM
To: Mckinsey, William G. (CJIS) (FBI); [redacted] (CJIS) (FBI); [redacted] (OTD) (FBI)
Subject: Re: FACE SWAPPING

b6
b7C
b7E

- DEEPPAKES: Fake America great again (MIT Tech Review, 8/17) Inside the race to catch the worryingly real fakes that can be made using AI. Perhaps the greatest risk is that the technology will further undermine truth and objectivity.

On 08/16/18 01:33 PM, [redacted] [redacted] wrote:

b6
b7C
b7E

MIT is working on detection...

This algorithm automatically spots "face swaps" in videos (MIT Tech Review, 4/10) But the same system can be used to make better fake videos that are harder to detect.

Very good, in-depth article on the national security/democracy implications!

Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy? (Lawfare, 2/21)

On 08/16/18 01:25 PM, [redacted] [redacted] wrote:

b6
b7C
b7E

looks like its still in its infancy... but is getting more prevalent. "Experts" are saying it's 1-2 years away from being really good. DARPA is already working on detection software.

I was afraid to click on too many articles - since it's being used mostly for porn.

these were ok.

Good basic video explaining the tech...

Face-swapping videos could lead to more 'fake news' (Business Insider, 4/13/18)

I never said that! High-tech deception of 'deepfake' videos (Phys.org, 7/2) This technology uses facial mapping and artificial intelligence to produce videos that appear so genuine it's hard to spot the phonies...Realizing the implications of the technology, the U.S. Defense Advanced Research Projects Agency is already two years into a four-year program to develop technologies that can detect fake images and videos. ..it's easy to foresee a nation state using them for nefarious activities against the U.S....

In An Era of Fake News, Advancing Face-Swap Apps Blur More Lines (NPR, 2/3/18)

Fake video news is coming, and this clip of Obama 'insulting' Trump shows how dangerous it could be (CNBC, 4/17) A BuzzFeed PSA seems to show former President Barack Obama saying disparaging things about President Donald Trump, but it's actually a PSA to show how easy it is to manipulate video and spread misinformation.

How to identify if an online video is fake (New Statesman, 2/14) As "deep fakes" raise concerns, everyone needs to equip themselves with the knowledge to spot a fraudulent video.

The deepest fake: how new tech will test our belief in what we see (Sydney Morning Herald (AU), 5/4)

On 08/16/18 09:34 AM, "Mckinsey, William G. (CJIS) (FBI)" wrote:

GUYS,

Pls follow this topic. This could require urgent action our part if it is real. Is anyone working on prevention or detection. I though the reaction of the Privacy Group was something else – Shut down Face Book and face recognition.

Let's discuss what if anything we should be doing about this challenge. will help us track the topic in the media.

b6
b7C
b7E

MCK

William G. McKinsey

Section Chief, FBI/CJIS

[O]
[M]

b7E

From: (CJIS) (FBI)
Sent: Thursday, August 16, 2018 6:53 AM
To: Mckinsey, William G. (CJIS) (FBI) < >
Subject: RE: face-swapping article fyi

b6
b7C
b7E

Fake videos? Computer program generates eerily realistic fake footage

Set to be unveiled at a computer animation festival in Vancouver, the software can also tweak head and torso poses, eye movements and background details to create more convincing fakes

<http://www.foxnews.com/tech/2018/08/15/fake-videos-computer-program-generates-eerily-realistic-fake-footage.html>

From: (CJIS) (FBI)
Sent: Tuesday, March 13, 2018 7:38 AM
To: Mckinsey, William G. (CJIS) (FBI) < >
Subject: RE: face-swapping article fyi

b6
b7C
b7E

How AI-generated videos could be the next big thing in fake news (Fox News) It's difficult to assess the national security risk or potential for disruption that is presented by the threat of AI-built fake videos.

From: Mckinsey, William G. (CJIS) (FBI)
Sent: Sunday, January 28, 2018 12:12 PM

To: [redacted] (CJIS) (FBI) [redacted] >
Subject: Re: face-swapping article fyi

b6
b7C
b7E

[redacted] I googled face swapping and learned a lot. I have [redacted] and [redacted] working on it. Thanks.

Pls follow [redacted] closely. It could put us out of business.

MCK

----- Original message -----

From: "[redacted] (CJIS) (FBI)" <[redacted]>

Date: 1/25/18 6:41 AM (GMT-07:00)

To: "Mckinsey, William G. (CJIS) (FBI)" <[redacted]>

Subject: face-swapping article fyi

b6
b7C
b7E

FYI

I didn't want to click on the full article.... This is a blurb on Slashdot.com that I took a screenshot of so the links can't be clicked. It has enough info in it.

Made me think that if they are doing this for trivial crap, then what is being done to surveillance video or other facial recognition images by others with better tools.



Cc: HQ-DIV19-DIRECTORS-BRIEFING-BOOKS [redacted]

b7E

Subject: FBI External Products Feedback Reload --- ~~SECRET//NOFORN//LES~~

Classification: ~~SECRET//NOFORN//LES~~

Classified By: [redacted]

b6

Derived From: FBI NSIC CG

b7C

Declassify On: 20431231

=====

Greetings,

The Director's Daily Briefing Unit is pleased to provide this weekly FBI External Products Feedback Reload. This document summarizes the daily external feedback with a visual representation of the number of FBI external intelligence products delivered to US Policymakers, confirmed number of principals briefed, and noteworthy feedback provided to the FBI last week.

The Director's Daily Briefing Unit welcomes your feedback.
Please send any feedback to the unit's group E-Mail account at [redacted]

b6
b7C
b7E

UC [redacted]
Director's Daily Briefing Unit
[redacted]

=====
Classification: ~~SECRET//NOFORN//LES~~

=====
Classification: ~~SECRET//NOFORN//LES~~

=====
Classification: ~~SECRET//NOFORN//LES~~

[redacted] (IMD) (CON)

b6
b7C

From: [redacted] (CYD) (FBI)
Sent: Thursday, August 02, 2018 2:56 PM
To: [redacted]
Subject: FW: Adversarial Sample Generation - Making Machine Learning Systems Robust for Security

[redacted]

FYI -

Regards,

b6
b7C
b7E

[redacted]
FBI Cyber Division

[redacted] (desk)
[redacted] (mobile)

From: noreply+feedproxy@google.com [mailto:noreply+feedproxy@google.com]
Sent: Thursday, August 02, 2018 9:47 AM
To: [redacted] (CYD) (FBI) [redacted]
Subject: [BULK] TrendLabs Security Intelligence Blog - by Trend Micro

b6
b7C
b7E

TrendLabs Security Intelligence Blog - by Trend Micro



Adversarial Sample Generation: Making Machine Learning Systems Robust for Security

Posted: 02 Aug 2018 05:00 AM PDT

The history of antimalware security solutions has shown that malware detection is like a cat-and-mouse game: every new detection technique, there's a new evasion method. When signature detection was invented, cybercriminals used packers, compressors, metamorphism, polymorphism, and obfuscation to evade it. Meanwhile, API hooking and code injection methods were developed to evade behavior detection. By the time machine learning (ML) was used in security solutions, it was already expected that cybercriminals would develop new tricks to evade ML.

To be one step ahead of cybercriminals, one method of enhancing an ML system to counter evasion tactics is generating *adversarial samples*, which are input data modified to cause an ML system to incorrectly classify. Interestingly, while adversarial samples can be *designed* to cause ML systems to malfunction, they can also result, be used to improve the efficiency of ML systems.

Making machine learning systems more robust via adversarial samples

Adversarial samples can help identify weaknesses in an ML model, which, in turn, can be used to gain valuable insights on how to enhance the model. By using a huge number of handcrafted samples modified from original malware, it is possible to repeatedly probe the capability of an ML system. This way, adversarial samples can retrain an ML system to make it more robust.

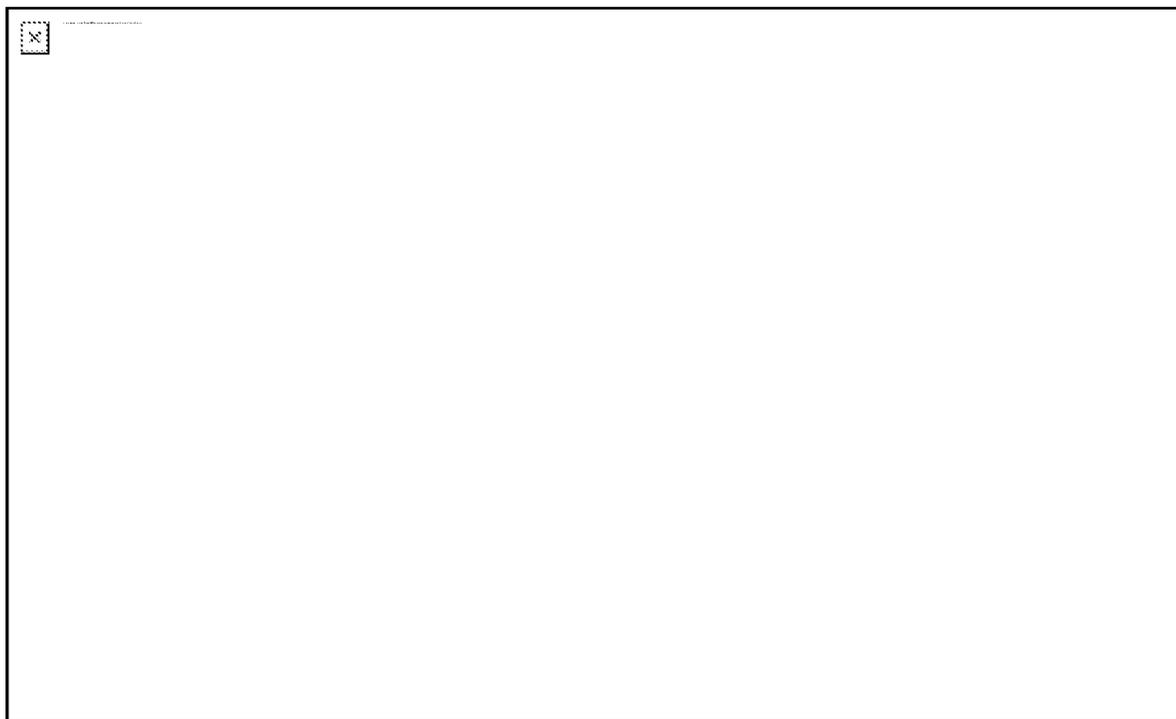


Figure 1. Using adversarial samples and AI to make an ML system more robust

At the onset of our research on a system to generate adversarial samples, we saw high probability scores. If a sample is detected with a high score, it means it has more similarities to the malware samples in our ML training set. Our goal is to gradually reduce the high probability score by modifying the malware sample until it becomes undetected. If successful, it means we have identified a weakness in the ML system and we may consider a number of activities to mitigate this weakness, such as identifying new features, do searches for related malware, or other components to identify such variants.

We selected a malware sample as seed, and defined it as m , a value signifying a certain number of possible changes (for example, 10, 20, 32, and 64). In our research, m is 32, which means we pre-defined 32 possible ways to modify the malware file. Through a genetic algorithm (GA), we found the combinations of changes we implement to the malware for it to evade detection. Here are the steps we took:

1. Generate a batch of new files with random n of m changes on the seed file.

2. Get ML prediction (detected or undetected) and gradient information (probability) on the new generated files.
3. If it reaches N loops (for example, 200), collect all undetected files from the whole procedure, and then exit.
4. Choose X (certain number) files as new seeds, which are undetected or detected, but with the lowest probability score.
5. Generate another batch of files by implementing a random combination of changes in the seeds and random new changes (optional).
6. Repeat step 2. The changes may damage and render the portable executable (PE) file unable to run. Also use a sandbox technology to validate if a newly generated file is still executable.



Figure 2. How we generated adversarial samples using genetic algorithm (GA)

In our findings, we observed that the probability output can be a security hole that the attackers can exploit easily probe an ML system's capability. Therefore, this number should be hidden in security products. With no probability output as a guide, we got curious whether a *brute force method* can be used to generate adversarial samples. We discovered that it still worked, but instead of producing one sample (in an undetected and undamaged state) in every 60 samples (when GA is used), we were able to produce only one in every 500 samples using brute force method.

The modification success rate of 0.2 percent (= 1/500) for the brute force method can still be considered a successful rate for generating adversarial samples when taking into account the significant and fundamental changes to the file structure. In our experience, approximately 3 percent of the generated samples were undamaged even after undergoing changes, and 7 percent of the samples were undetected. However, when

one (out of 500) adversarial sample is used as a seed in the next phase where we generate another batch of samples, the success rate can increase back to 1.5 percent. The generation rate of undamaged samples will be at 3 percent, but around half of the samples will be undetected.

There are two main factors to consider when generating adversarial samples: First, figuring out how to safely modify a PE file without damaging it, and second, finding a method to generate undetected samples efficiently. For the second point, AI can be used for choosing the right file features to modify and map the changes to the features and the numerous potential changes to the PE files. It takes a lot of time and effort to come up with many possible combinations of changes to a sample and to test them in a system to produce all possible adversarial samples. ML can help quickly choose the most useful changes or combinations that can decrease gradient information (i.e., probability) — therefore making adversarial sample generation more efficient.

Protecting ML systems from potential evasion methods and other attacks

While using adversarial samples to enhance an ML system can be effective, security holes may still appear for cybercriminals to exploit. For example, in the same way that we were trying to add normal characteristics to a malware sample for it to seem benign and become undetectable, attackers could find ways to evade detection by infecting a benign PE file or compiling a benign source code with malicious code or injecting binary code. These methods can make a malware appear benign to an ML system when its structure still comprises mostly that of the original benign file. This can bring challenges to an ML system: If this situation is not carefully accounted for, some ML systems might detect the compromised file as more similar to the original benign file it originated from.

ML training set poisoning is another issue to watch for. When an ML system's training set includes malware samples similar to benign files, it will be prone to false positives. Example: the *PTCH_NOPLE* malware, a part of the *PTCH* family that modifies the *dnsapi.dll* file, which is a module that assists the DNS client service in the Windows® operating system. Some ML systems in the industry have higher false positive rates because of benign *dnsapi.dll* files infected with *PTCH_NOPLE*.

To counter evasion methods and other types of attacks against machine learning in security solutions, we can come up with mitigation techniques.

1. Set up a defense at the infrastructure level by reducing the attack surface of the ML system. Some techniques to achieve this include the following:
 - Not exposing the system to probing or making the system less susceptible to probing. An attacker can stealthily modify samples to probe an ML system by using a free tool that has a local ML model for use. A cloud-based system can prevent this, as all predictions by the ML system can be recorded at the backend. That way, details on who is attempting to probe the model and where and when the attempt happened can be tracked. Distribution and usage of such tools should be limited.

- Use cloud-based solutions, such as products with Trend Micro XGen security, to detect and block malicious probing. If an attempt is detected by the solution, it will show fake results to the attacker or it can terminate the product or service associated with the account the attacker is using.
 - Use security products armed with a combination of detection technologies. By doing this, the attacker cannot exactly know which will be the only sample detected by the ML system.
 - Hiding the real gradient information (probability score) of an ML system.
2. Make the ML system more robust, first, by identifying potential vulnerabilities early on in its design phase and making it accurate for every parameter. Second, generate adversarial samples and use them to retrain the ML model. It could be done via black box testing using GA or brute force computation or white box testing. These two methods should be implemented continuously throughout the ML system's whole lifecycle.
 3. Consider using generative adversarial network (GAN). GAN has two types of AI: one generates data instances, and the other evaluates them for authenticity. The two AI types can train each other to evolve. We also used GAN to find better ways to generate adversarial samples (automatically) as well as to find ways to secure them.
 4. To reduce false positives caused by threats such as PTCH_NOPLE, use security solutions that not only utilize ML for detection but also for whitelisting. Trend Micro XGen security uses the Trend Micro Locality Sensitive Hash (LSH), an approach that generates a hash value which can then be analyzed for similarities. Since collecting all file versions and adding them for whitelisting is difficult, a similar version of a file that is known and legitimate can be used to compare to a wrongly detected file. If the LSH values are similar and they have the same signature chain, false positives can be reduced. Therefore, we also encourage application developers to sign their files to reduce the risk of files being misclassified by antimalware products.

Enhancing a machine learning system fortifies overall cyberdefense

An efficient ML system should detect not only existing malware but also adversarial samples. Using GANs, GA, and brute force methods, among other strategies, can enable an ML system to perform such a task. This capability can give an ML system a wider coverage for threats and lower false positive rates, which in turn, can help an ML system detect and counter evasion techniques when coupled with an ML-based whitelisting method. Countermeasures for ML evasion methods will be one of the key features in ML in cybersecurity in the future. Looking out for evasion samples in the wild is important because in the game of evasion versus anti-evasion, it will be difficult to detect what can't be seen.

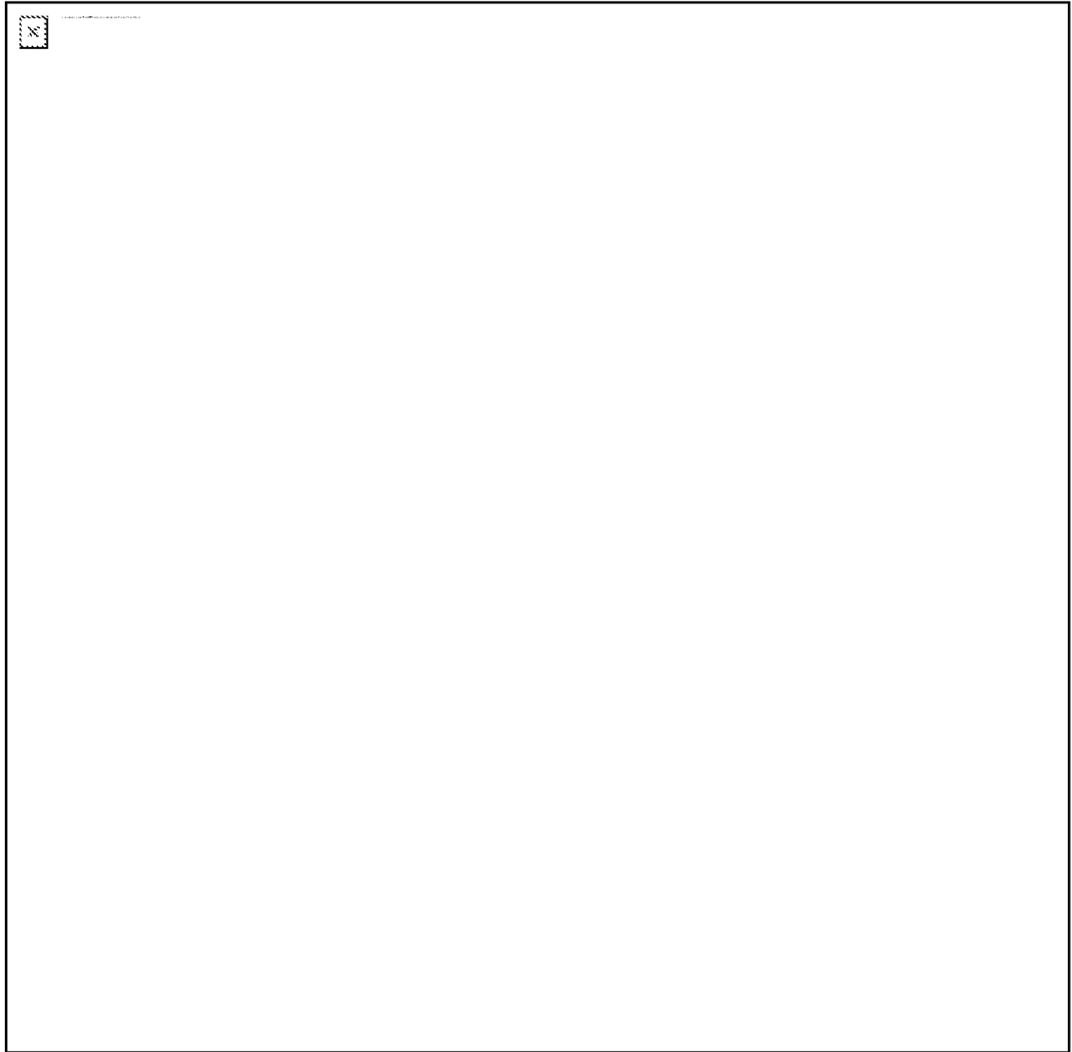
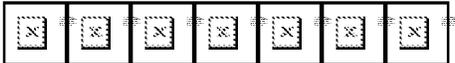


Figure 3. Diagram of an efficient ML system that is capable of detecting and blocking threats and adversarial samples

However, while an enhanced machine learning system certainly improves detection and block rates, it isn't the all and end-all in cybersecurity. Since cybercriminals are also always on the lookout for security gaps, a multilayered defense is still most effective at defending users and enterprises against different kinds of threats. Trend Micro XGen security is equipped with a cross-generational blend of threat defense techniques, including machine learning, web/URL filtering, behavioral analysis, and custom sandboxing, and defends data centers, cloud environments, networks, and endpoints against a full range of threats.

The post [Adversarial Sample Generation: Making Machine Learning Systems Robust for Security](#) appeared first on [...](#)



You are subscribed to email updates from [a feed](#).
To stop receiving these emails, you may [unsubscribe now](#).

Email delivery powered by Google

Google, 1600 Amphitheatre Parkway, Mountain View, CA 94043, United States

[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI)
Sent: Thursday, August 02, 2018 2:40 PM
To: [redacted]
Subject: FW: [BULK] TrendLabs Security Intelligence Blog - by Trend Micro

b6
b7C

FYI -

Regards,

[redacted]
FBI Cyber Division

b6
b7C

[redacted] (desk)
[redacted] (mobile)

From: noreply+feedproxy@google.com [mailto:noreply+feedproxy@google.com]
Sent: Thursday, August 02, 2018 9:47 AM
To: [redacted] (CYD) (FBI) [redacted]
Subject: [BULK] TrendLabs Security Intelligence Blog - by Trend Micro

b6
b7C
b7E

TrendLabs Security Intelligence Blog - by Trend Micro



Adversarial Sample Generation: Making Machine Learning Systems Robust for Security

Posted: 02 Aug 2018 05:00 AM PDT

The history of antimalware security solutions has shown that malware detection is like a cat-and-mouse game. For every new detection technique, there's a new evasion method. When signature detection was invented, cybercriminals used packers, compressors, metamorphism, polymorphism, and obfuscation to evade it. Meanwhile, API hooking and code injection methods were developed to evade behavior detection. By the time machine learning (ML) was used in security solutions, it was already expected that cybercriminals would develop new tricks to evade ML.

To be one step ahead of cybercriminals, one method of enhancing an ML system to counter evasion tactics is generating *adversarial samples*, which are input data modified to cause an ML system to incorrectly classify. Interestingly, while adversarial samples can be *designed* to cause ML systems to malfunction, they can also result, be used to improve the efficiency of ML systems.

Making machine learning systems more robust via adversarial samples

Adversarial samples can help identify weaknesses in an ML model, which, in turn, can be used to gain valuable insights on how to enhance the model. By using a huge number of handcrafted samples modified from original

malware, it is possible to repeatedly probe the capability of an ML system. This way, adversarial samples can retrain an ML system to make it more robust.

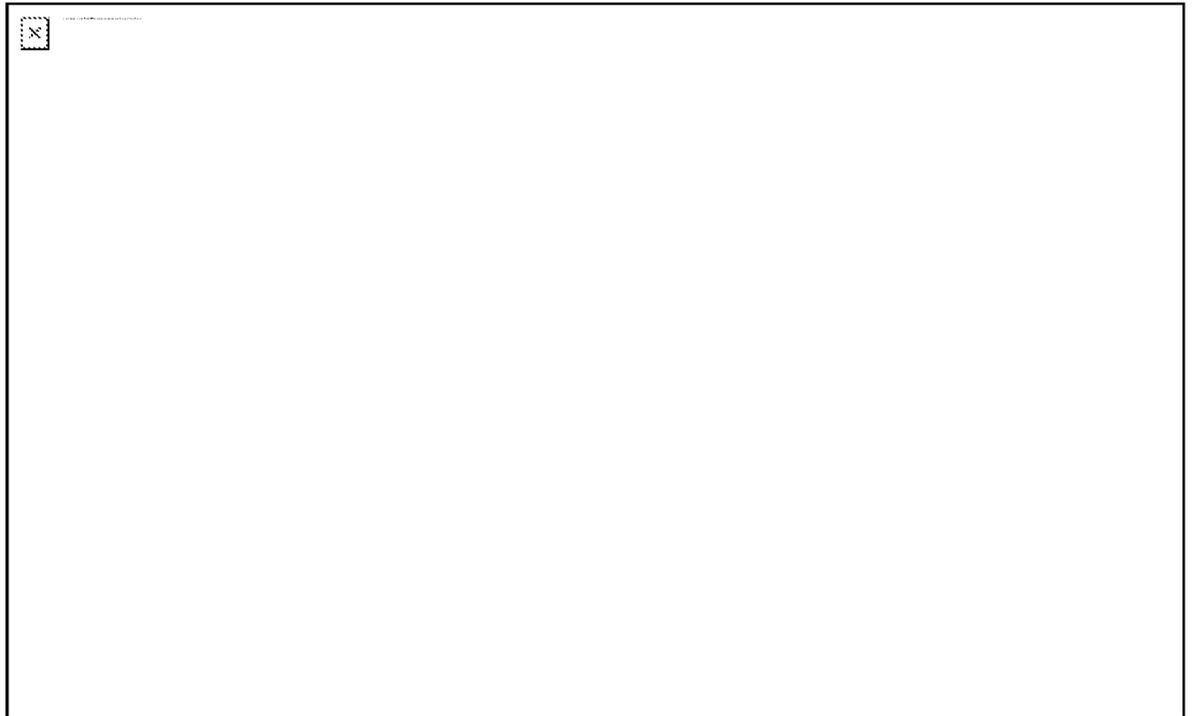


Figure 1. Using adversarial samples and AI to make an ML system more robust

At the onset of our research on a system to generate adversarial samples, we saw high probability scores. If is detected with a high score, it means it has more similarities to the malware samples in our ML training set. The goal is to gradually reduce the high probability score by modifying the malware sample until it becomes undetected. If successful, it means we have identified a weakness in the ML system and we may consider a number of activities to mitigate this weakness, such as identifying new features, do searches for related malware, or other components to identify such variants.

We selected a malware sample as seed, and defined it as m , a value signifying a certain number of possible changes (for example, 10, 20, 32, and 64). In our research, m is 32, which means we pre-defined 32 possible ways to modify the malware file. Through a genetic algorithm (GA), we found the combinations of changes we implement to the malware for it to evade detection. Here are the steps we took:

1. Generate a batch of new files with random n of m changes on the seed file.
2. Get ML prediction (detected or undetected) and gradient information (probability) on the new generated files.
3. If it reaches N loops (for example, 200), collect all undetected files from the whole procedure, and then exit.

4. Choose X (certain number) files as new seeds, which are undetected or detected, but with the lower probability score.
5. Generate another batch of files by implementing a random combination of changes in the seeds and random new changes (optional).
6. Repeat step 2. The changes may damage and render the portable executable (PE) file unable to run. Also use a sandbox technology to validate if a newly generated file is still executable.

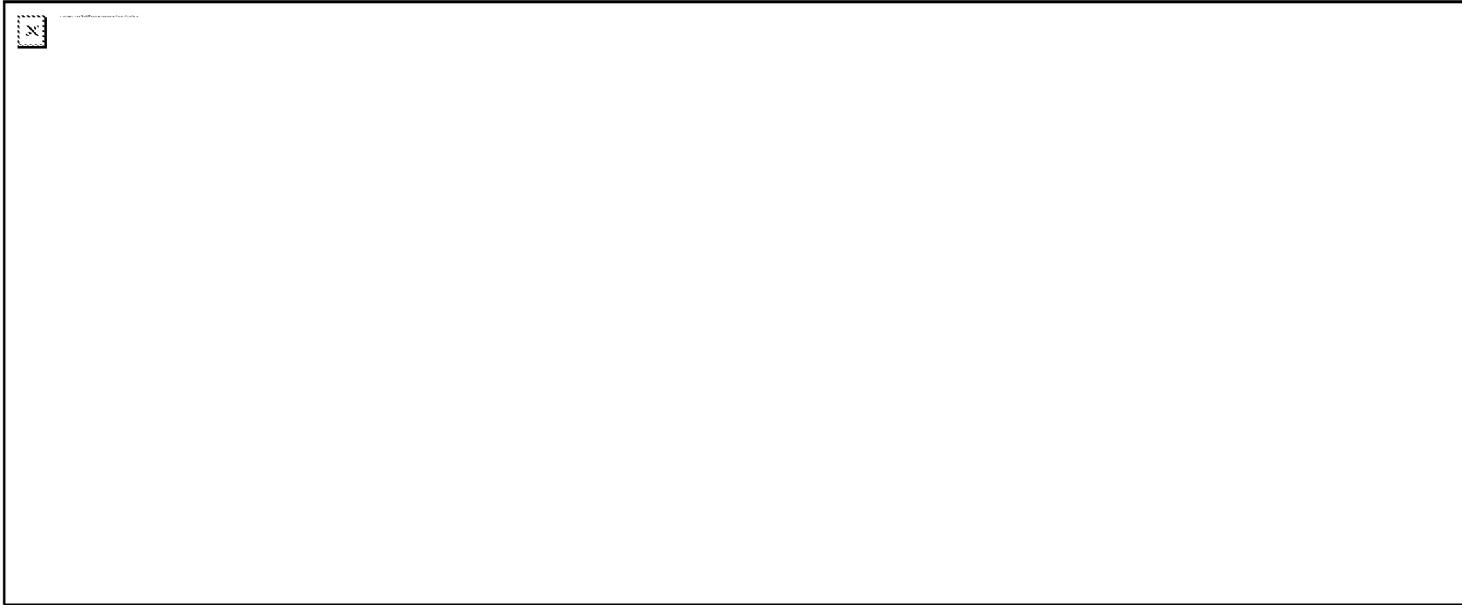


Figure 2. How we generated adversarial samples using genetic algorithm (GA)

In our findings, we observed that the probability output can be a security hole that the attackers can exploit easily probe an ML system's capability. Therefore, this number should be hidden in security products. With no probability output as a guide, we got curious whether a *brute force method* can be used to generate adversarial samples. We discovered that it still worked, but instead of producing one sample (in an undetected and undamaged state) in every 60 samples (when GA is used), we were able to produce only one in every 500 samples using brute force method.

The modification success rate of 0.2 percent ($= 1/500$) for the brute force method can still be considered a successful rate for generating adversarial samples when taking into account the significant and fundamental changes to the file structure. In our experience, approximately 3 percent of the generated samples were undamaged even after undergoing changes, and 7 percent of the samples were undetected. However, when one (out of 500) adversarial sample is used as a seed in the next phase where we generate another batch of samples, the success rate can increase back to 1.5 percent. The generation rate of undamaged samples will be at 3 percent, but around half of the samples will be undetected.

There are two main factors to consider when generating adversarial samples: First, figuring out how to safely modify a PE file without damaging it, and second, finding a method to generate undetected samples efficiently.

For the second point, AI can be used for choosing the right file features to modify and map the changes to the features and the numerous potential changes to the PE files. It takes a lot of time and effort to come up with many possible combinations of changes to a sample and to test them in a system to produce all possible adversarial samples. ML can help quickly choose the most useful changes or combinations that can decrease gradient information (i.e., probability) — therefore making adversarial sample generation more efficient.

Protecting ML systems from potential evasion methods and other attacks

While using adversarial samples to enhance an ML system can be effective, security holes may still appear for cybercriminals to exploit. For example, in the same way that we were trying to add normal characteristics to a malware sample for it to seem benign and become undetectable, attackers could find ways to evade detection by infecting a benign PE file or compiling a benign source code with malicious code or injecting binary code. These methods can make a malware appear benign to an ML system when its structure still comprises mostly that of the original benign file. This can bring challenges to an ML system: If this situation is not carefully accounted for, some ML systems might detect the compromised file as more similar to the original benign file it originated from.

ML training set poisoning is another issue to watch for. When an ML system's training set includes malware samples similar to benign files, it will be prone to false positives. Example: the `PTCH_NOPLE` malware, a part of the `PTCH` family that modifies the `dnsapi.dll` file, which is a module that assists the DNS client service in the Windows® operating system. Some ML systems in the industry have higher false positive rates because of benign `dnsapi.dll` files infected with `PTCH_NOPLE`.

To counter evasion methods and other types of attacks against machine learning in security solutions, we can come up with mitigation techniques.

1. Set up a defense at the infrastructure level by reducing the attack surface of the ML system. Some techniques to achieve this include the following:
 - Not exposing the system to probing or making the system less susceptible to probing. An attacker can stealthily modify samples to probe an ML system by using a free tool that has a local ML model for use. A cloud-based system can prevent this, as all predictions by the ML system can be recorded at the backend. That way, details on who is attempting to probe the model and where and when the attempt happened can be tracked. Distribution and usage of such tools should be limited.
 - Use cloud-based solutions, such as products with Trend Micro  security, to detect and block malicious probing. If an attempt is detected by the solution, it will show fake results to the attacker or it can terminate the product or service associated with the account the attacker is using.
 - Use security products armed with a combination of detection technologies. By doing this, the attacker cannot exactly know which will be the only sample detected by the ML system.

- Hiding the real gradient information (probability score) of an ML system.
2. Make the ML system more robust, first, by identifying potential vulnerabilities early on in its design phase and making it accurate for every parameter. Second, generate adversarial samples and use them to retrain the ML model. It could be done via black box testing using GA or brute force computation or white box testing. These two methods should be implemented continuously throughout the ML system's whole lifecycle.
 3. Consider using *generative adversarial network (GAN)*. GAN has two types of AI: one generates data instances, and the other evaluates them for authenticity. The two AI types can train each other and evolve. We also used GAN to find better ways to generate adversarial samples (automatically) as well as to find ways to secure them.
 4. To reduce false positives caused by threats such as PTCH_NOPLÉ, use security solutions that not only utilize ML for detection but also for whitelisting. Trend Micro XGen security uses the *Trend Micro Locality Sensitive Hash (TLSH)*, an approach that generates a hash value which can then be analyzed for similarities. Since collecting all file versions and adding them for whitelisting is difficult, a similar version of a file that is known and legitimate can be used to compare to a wrongly detected file. If the TLSH values are similar and they have the same signature chain, false positives can be reduced. Therefore, we also encourage application developers to sign their files to reduce the risk of files being misclassified by antimalware products.

Enhancing a machine learning system fortifies overall cyberdefense

An efficient ML system should detect not only existing malware but also adversarial samples. Using GANs, GA, and brute force methods, among other strategies, can enable an ML system to perform such a task. This capability can give an ML system a wider coverage for threats and lower false positive rates, which in turn, can help an ML system detect and counter evasion techniques when coupled with an ML-based whitelisting method. Countermeasures for ML evasion methods will be one of the key features in ML in cybersecurity in the future. Looking out for evasion samples in the wild is important because in the game of evasion versus anti-evasion, it will be difficult to detect what can't be seen.

[redacted] (IMD) (CON)

b6
b7C

From: [redacted] (CYD) (FBI)
Sent: Thursday, October 04, 2018 8:55 AM
To: [redacted] (TD) (FBI); [redacted] (CYD) (FBI)
Subject: FW: Deep Fakes POC

Thanks, [redacted]

From: [redacted] (CYD) (FBI)
Sent: Thursday, October 04, 2018 8:43 AM
To: [redacted] (CYD) (FBI); [redacted]; Karl, Larry D. (CYD) (FBI); [redacted]
Subject: RE: Deep Fakes POC

b6
b7C
b7E

Works for me.

Thanks, [redacted]

From: [redacted] (CYD) (FBI)
Sent: Thursday, October 04, 2018 8:42 AM
To: [redacted] (CYD) (FBI); [redacted]; Karl, Larry D. (CYD) (FBI); [redacted]
Subject: Re: Deep Fakes POC

b6
b7C
b7E

So, shall I reply that you are POC and will continue [redacted] on this initiative as you have been?

----- Original message -----

From: [redacted] (CYD) (FBI)" [redacted]
Date: 10/4/18 8:39 AM (GMT-05:00)
To: "Karl, Larry D. (CYD) (FBI)" [redacted], [redacted] (CYD) (FBI)" [redacted]
Subject: FW: Deep Fakes POC

b6
b7C
b7E

Good morning again. I wanted to provide some background on this topic and [redacted] specifically [redacted]

In the past sixteen months, [redacted] All of which were coordinated with [redacted] and Mr. [redacted] No response was ever received from [redacted] on any of the 4 products send down for coordination. Not even a one line, "we concur". Mr. [redacted] has been very helpful and provided input to all four pieces.

b6
b7C
b7E

Whether they jump on board for the [redacted] or not, I wanted to give you a clearer picture of our outreach to them regarding this issue. I do not think the below message shows our outreach to them re this topic in the correct light.

Thanks, - [redacted]

b6
b7C
b7E

From [redacted] (OTD) (FBI)
Sent: Thursday, October 04, 2018 8:08 AM
To [redacted] (CYD) (FBI); [redacted] (CYD) (FBI); [redacted]; [redacted]
[redacted] (OTD) (FBI); [redacted]
Cc [redacted] (OTD) (FBI); [redacted]; Ferensic, Susan (SC) (FBI); [redacted]; [redacted] (OTD)
(FBI) [redacted]
Subject: Fwd: Deep Fakes POC

b6
b7C
b7E

[redacted]

Are you the one included in the list below?

I would like to suggest that [redacted] deep fakes query. The Digital Section has been conducting authentication work for years and we are engaged with the research community, like your colleague, [redacted]. Likewise, our [redacted] has a responsibility to maintain awareness of adversary capabilities like this.

b6
b7C
b7E

Even if the request from [redacted] was of a nontechnical nature, it would benefit the Bureau to make sure we have covered the response across the board.

Im tied up all day at a video analytics event in NGA, so limited access to email today, but I'd be happy to touch base tomorrow.

FWIW - the Deep Fakes issue was a key aspect of this conference yesterday, with [redacted] who started this thread involved...

[redacted]

FBI - OTD
Building 27958A
Pod E
Quantico, VA 22135

(O) [redacted]
(M) [redacted]

b6
b7C
b7E

[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI) b6
b7C
Sent: Wednesday, August 01, 2018 9:40 AM
To: [redacted] (CYD) (FBI)
Cc: [redacted] (CYD) (FBI); KARL, LARRY D. JR. (CYD) (FBI); [redacted]
[redacted] (CYD) (FBI); [redacted] (CYD) (FBI)
Subject: FW: GANs - OCA Proposal --- ~~SECRET//NOFORN~~
Attachments: Congressional Briefing Proposal_GANs.docx

Classification: ~~SECRET//NOFORN~~

Classified By: [redacted] b6
b7C
Derived From: Multiple Sources
Declassify On: 50X1-HUM
=====

[redacted] per our UNET conversation, please see briefing overview, attached.

Thanks, [redacted]

[redacted]
Chief, Technology Cyber Intelligence Unit
Cyber Engagement & Intelligence Section - CyD
Mailstop: M-Ridg/A Cyber Rm-502

- Desk
- Cell
- SIPR
- NSTS

[redacted] (NIPR)
[redacted] (SIPR)
[redacted] (SCION/JWICS)

b6
b7C
b7E

1995: Every object in your home has a clock & it is blinking 12:00
2025: Every object in your home has an IP address & the password is admin

From: [redacted] (CYD) (FBI)
Sent: Wednesday, August 01, 2018 9:32 AM
To: [redacted] (CYD) (FBI); [redacted]; [redacted] (CYD) (FBI)
[redacted]
Subject: GANs - OCA Proposal --- ~~SECRET//NOFORN~~ b6
b7C
b7E

Classification: ~~SECRET//NOFORN~~

Classified By: [redacted] b6
b7C
Derived From: Multiple Sources
Declassify On: 50X1-HUM
=====

TRANSITORY RECORD

As requested.

Regards,

[Redacted]

Technology Cyber Intelligence Unit
FBI Cyber Division

b6
b7C
b7E

Open: [Redacted]

Secure: [Redacted]

=====
Classification: ~~SECRET~~//NOFORN

=====
Classification: ~~SECRET~~//NOFORN

[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI)
Sent: Tuesday, October 16, 2018 2:09 PM
To: [redacted] (CYD) (FBI)
Subject: guess who

Just got a [redacted]

Thanks, [redacted]

b6
b7C
b7E

[redacted]
Chief, Technology Cyber Intelligence Unit Cyber Engagement & Intelligence Section Cyber Division Federal Bureau of Investigation

[redacted] (o)
[redacted] (c)

[redacted] IMD) (CON)

b6
b7C

From: [redacted] (CYD) (FBI)
Sent: Wednesday, August 01, 2018 9:39 AM
To: [redacted] (CYD) (FBI)
Subject: FW: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

FYI – We are working with ESU to push a request to Congressional Affairs for a [redacted]

Regards,

[redacted]
FBI Cyber Division

b6
b7C
b7E

[redacted] (desk)
[redacted] (mobile)

From: [redacted] (CYD) (FBI)
Sent: Tuesday, July 31, 2018 7:50 AM
To: [redacted]

[redacted]

b6
b7C
b7E

Subject: FW: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

Gans

Thanks, [redacted]

From: The Washington Post [mailto:email@washingtonpost.com]

Sent: Tuesday, July 31, 2018 7:45 AM

To: [redacted] (CYD) (FBI) [redacted]

Subject: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

b6
b7C
b7E

[redacted] [VIEW ON WEB >](#)

[redacted]

[redacted] [redacted]

Decoding cybersecurity news in one morning tipsheet. Not on the list? [Sign up here.](#)



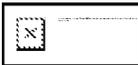
Hack your day

 Share

 Share

 Tips/Feedback

Doctored videos could send fake news crisis into overdrive, lawmakers warn



BY DEREK HAWKINS
with Bastien Inzaurrealde

THE KEY



Sen. Marco Rubio (R-Fla.) on Capitol Hill in Washington on March 14. (Jacquelyn Martin/AP)

Two lawmakers are warning that the country is woefully unprepared for the rise of deepfakes, alarmingly realistic videos that appear to show people doing things they didn't do.

Sens. Mark R. Warner (D-Va.) and Marco Rubio (R-Fla.) are exploring ways to curb the trend of doctored videos before it becomes too widespread, saying they could wreak havoc if used in disinformation campaigns like the one conducted by the Russian government in 2016. In a wide-ranging technology policy paper Monday, Warner floated the idea of holding social media platforms liable for failure to take down deepfakes. And Rubio in a recent speech called on government and political leaders to treat them as a national security threat.

The attention from lawmakers means deepfakes are no longer a fringe issue but a more serious front in the fight against fake news,

and tech companies may soon feel pressure to get ahead of them. **But any policy solution would have to balance the harm to potential victims against free-speech rights for people who use deepfakes for creative or satirical purposes.**

Warner said the easily accessible technology used to make the videos could “usher in an unprecedented wave of false and defamatory content.” In his policy paper, he wrote, **“Just as we’re trying to sort through the disinformation playbook used in the 2016 election and as we prepare for additional attacks in 2018, a new set of tools is being developed that are poised to exacerbate these problems.”**

Software to create deepfakes is available for free online, and it doesn’t require advanced production skills to use. It works by feeding hundreds of pictures of a person’s face into a machine learning algorithm that then maps them onto video of another person’s body. Anything the person in the video does or says can be made to look like it’s coming from the victim. The results are sometimes so seamless that it’s difficult to tell with the naked eye that the videos are fraudulent.

Lawmakers caution that it’s a tool that could send the fake news crisis into overdrive. Think about it: Realistic-looking videos appearing to show politicians meeting taking bribes or uttering inflammatory statements could be used to try to sway an election. Or doctored footage purporting to show officials announcing military action could trigger a national security crisis.

“This all sounds fantastic, it all sounds exaggerated, it all sounds hyperbolic. **But the capability to do all of this is real and exists now,**

the willingness exists now, all that's missing is the execution. And we are not ready for it," Rubio said in a speech earlier this month at the right-leaning Heritage Foundation. "I know for a fact that the Russian Federation at the command of Vladimir Putin tried to sow instability and chaos in American politics in 2016," he said. "They did that through Twitter bots and they did that through a couple of other measures that will increasingly come to light. But they didn't use this. **Imagine using this. Imagine injecting this in an election."**

To chip away at the problem, Warner has proposed is amending the Communications Decency Act to hold social media platforms liable under state law if they don't take down deepfakes and other manipulated content shown in court to be defamatory. Right now, the law provides immunity for platforms in such cases.

"Currently the onus is on victims to exhaustively search for, and report, this content to platforms — who frequently take months to respond and who are under no obligation thereafter to proactively prevent the same content from being re-uploaded in the future," Warner wrote in his policy proposal. The platforms, he said, were "in the best place to identify and prevent this kind of content from being propagated."

Legislation to do this would almost certainly run into opposition from civil liberties groups. This year, organizations such as the Electronic Frontier Foundation lobbied unsuccessfully against a similar carve-out in the Communications Decency Act that sought to hold media platforms liable for sex trafficking. The groups said the move, while well-intended, was so broadly written that it criminalized protected speech.

“Any effort on this front would need to address the challenge of distinguishing true deepfakes aimed at spreading disinformation from satire or other legitimate forms of entertainment or parody,” Warner wrote. “Attempting to distinguish between true disinformation and legitimate satire could prove difficult,” he said, but “courts already must make distinction between satire and defamation/libel.”

Deepfakes started cropping up last year on Reddit after a user superimposed the faces of Gal Gadot, Taylor Swift and other celebrities onto the faces of actors in pornographic videos. They’ve also been used to lampoon President Trump by pasting his face over Russian President Vladimir Putin and German Chancellor Angela Merkel. And the comedian Jordan Peele used the technology to graft President Barack Obama’s face over his own in a widely-circulated public service announcement warning of the dangers of deepfakes.

“It’s only a matter of time until ‘deepfake’ videos become a household term,” Rubio told me in an email.

Rubio hasn’t offered any concrete policy proposals yet. For now, he told me, he’s simply trying to sound the alarm in hopes of bringing new ideas to the table.

“I’m working to raise awareness,” he said, “and find ways to address this threat from foreign actors and criminals and defend our elections this fall and in the future.”



CONTENT FROM NORTHROP GRUMMAN

TRUSTED PROTECTION ACROSS EVERY DOMAIN.

In today's conflicts, traditional systems aren't the only ones targeted. At Northrop Grumman, we create full-spectrum cyber solutions to actively combat these threats. Learn more.



You are reading **The Cybersecurity 202**, our must-read newsletter on cybersecurity policy news.

Not a regular subscriber?

[SIGN UP NOW](#)

PINGED, PATCHED, PWNEED



Sen. Mark R. Warner (D-Va.) on Capitol Hill in Washington on July 25. (Al Drago/Getty Images)

PINGED: Warner's deepfakes proposal is one of 20 ideas he proposed to overhaul the rules that govern tech companies. In his policy paper, Warner also proposes “to give users ownership of their data and require their consent before a third party can access that information, and to commit new funding to the Federal Trade Commission and media literacy campaigns,” The Washington Post's Karoun Demirjian reported. However, it is far more certain that Warner would be able to garner support from Republican senators for his measures, especially as the midterm elections approach, my colleague reported.

“Some of Warner’s proposals reflect demands that have been voiced elsewhere around Congress, such as his calls to improve national defenses against cyber intrusions and establish a ‘deterrence doctrine’ to specify what steps the United States will take in response to cyber attacks,” Demirjian wrote. “But others envision a **new legal conceptualization of social media companies, as entities with a fiduciary duty to their users, and only temporary custodians of content and information that users could have the right to take with them from platform to platform**, much like the portability of telephone numbers from company to company. **Warner imagines laws that would allow for audits of social media companies’ algorithms**, as well as ‘public interest’ laws that would let experts and academics scrutinize how companies are using the data they collect.”



Sen. Jeanne Shaheen (D-N.H.) on Capitol Hill in Washington Jan. 27, 2016. (Alex Brandon/AP)

PATCHED: A man claiming to be a Latvian official emailed and called the office of Sen. Jeanne Shaheen (D-N.H.) last year to seek information on U.S. sanctions against Russia, the Daily Beast’s Andrew Desiderio and Kevin Poulsen reported Monday. The man offered to set up a phone call between Shaheen and Latvia’s foreign minister to discuss sanctions as well as the Russian anti-virus company Kaspersky Lab. Desiderio and Poulsen noted that Shaheen had pushed for a measure requiring the federal government to rid its networks of Kaspersky software. **The attempt was thwarted after Shaheen’s staff spoke with the Latvian Embassy and realized the operation was not legitimate.**

“Ryan Nickel, a spokesman for Shaheen, told the Daily Beast that staffers in her Senate office frequently receive hoax emails and phishing attempts

on their official email accounts,” Desiderio and Poulsen wrote. “They shared the more troubling ones, including the approach by the fake Latvian, with law enforcement officials.” However, there are no indications yet that Russian authorities are to blame for the operation against Shaheen. **“No malware was attached to the emails, and the fake foreign ministry official did not try to send Shaheen’s staff to a malicious website,”** Desiderio and Poulsen wrote. “An Internet IP address in the e-mail headers traces back to a hosting company in Amsterdam.”



A laptop in North Andover, Mass., on June 19, 2017. (Elise Amendola/AP)

PWNED: “One of Iowa’s main hospital and clinic systems has notified about 1.4 million patients that their personal information might have been breached,” the Des Moines Register’s Tony Leys reported on Monday. “UnityPoint Health officials said hackers used ‘phishing’ techniques to break into the company’s email system. The company, based in West Des Moines, said the hackers could have obtained medical information, such as diagnoses and types of care, that was included in emails.” In a notice posted on its website, UnityPoint Health said it discovered the cyberattack on May 31, reported it to law enforcement and launched a forensics investigation.

The company said some employees gave away their log-in credentials after receiving the phishing emails, which were crafted to look as if a “trusted executive” of the company had sent them. “Some of the compromised accounts included emails or attachments to emails, such as standard reports related to healthcare operations, containing protected health information and/or personal information for certain patients,” according to the company’s notice. “While unauthorized access

to patient information may have occurred, **no known or attempted misuse of patient information has been reported at this time.**” The company also said it is “more likely” that hackers carried out the cyberattack to ultimately steal money rather than to seize patients’ information.

— **More cybersecurity news:**

DHS Forms New Cyber Hub to Protect Critical U.S. Infrastructure

The Department of Homeland Security will announce on Tuesday the creation of a center aimed at guarding the nation’s banks, energy companies and other industries from major cyberattacks, agency officials said.



The Wall Street Journal • [Read more »](#)

U.S. spy agencies: North Korea is working on new missiles

Weeks after the Trump-Kim summit, factories are still producing intercontinental ballistic missiles and enriched uranium.



Ellen Nakashima and Joby Warrick • [Read more »](#)

Paul Manafort made more than \$60 million in Ukraine, prosecutors say

The special counsel’s team defended the inclusion of details of Manafort’s Ukraine work in his trial on bank- and tax-fraud charges.



Rachel Weiner • [Read more »](#)





Chinese telecom company ZTE's Beijing research and development center on June 13. (Jason Lee/Reuters)

— **The federal government’s ambition to contain Chinese telecom giants ZTE and Huawei out of concern that they may threaten national security could in return hamper efforts to develop 5G technology in the United States**, according to CyberScoop’s Ryan Duffy. “The quest to upend China’s surveillance capabilities may be hurting America’s competitiveness in the race to develop and roll out 5G wireless technology,” Duffy reported Monday. “The dilemma presents the latest — and perhaps fiercest — technological showdown between Washington and Beijing to date.”

— **“The U.S. Department of Defense will for the first time be using large-scale artificial intelligence systems that could automate mundane tasks and augment the work of military members as a result of an \$885 million five-year contract**, said Josh Sullivan, senior vice president at government consulting firm Booz Allen Hamilton,” the Wall Street Journal’s Sara Castellanos reported Monday. “The technology will allow the Defense Department to better compete with nations including China and Russia, said Mr. Sullivan, who leads the analytics business for Booz Allen.”

— **More cybersecurity news about the public sector:**

Supply Chain Cybersecurity a Major Legislative Priority for House Homeland

The committee wants to broaden DHS' authority to kick questionable contractors off government networks.

Nextgov • [Read more »](#)



House GOP intends to seek Comey interview after August recess

House Republicans are planning to seek an interview with former FBI Director James Comey in September to discuss his decisionmaking during the 2016 election, The Hill has learned.

The Hill • [Read more »](#)



Russian Jamming Poses a Growing Threat to U.S. Troops in Syria

But this type of warfare also gives the United States a chance to learn about the latest Russian technology.

Foreign Policy • [Read more »](#)



PRIVATE KEY

Amazon Promises “Unwavering” Commitment to Police, Military Clients Using AI Technology

As employees protest the use of Amazon artificial intelligence by U.S. enforcement agencies, executive Teresa Carlson affirms commitment to the government.

The Intercept • [Read more »](#)



Symantec: Financial cyberattacks are on the rise

Hackers targeting both large financial institutions and individuals for profit have increased over the past year, according to security software company Symantec.

Fifth Domain • [Read more »](#)



SECURITY FAILS

— Law enforcement authorities have caught a hacker who allegedly carried out SIM hijacking schemes against cryptocurrency investors, Motherboard's Lorenzo Franceschi-Bicchierai reported Monday. "On July 12, police in California arrested a college student accused of being part of a group of criminals who hacked dozens of cellphone numbers to steal more than \$5 million in cryptocurrency," Franceschi-Bicchierai wrote. "Joel Ortiz, a 20-year-old from Boston, allegedly hacked around 40 victims with the help of still unnamed accomplices, according to court documents obtained by Motherboard." Here is how the scam works, according to Motherboard: "SIM swapping consists of tricking a provider like AT&T or T-Mobile into transferring the target's phone number to a SIM card controlled by the criminal. Once they get the phone number, fraudsters can leverage it to reset the victims' passwords and break into their online accounts (cryptocurrency accounts are common targets.) In some cases, this works even if the accounts are protected by two-factor authentication."

— More news about security breaches:

Hackers find creative way to steal \$7.7 million without being detected

Thieves obtain platform's private key, use it to destroy coins, then create new ones.
[Ars Technica](#) • [Read more »](#)

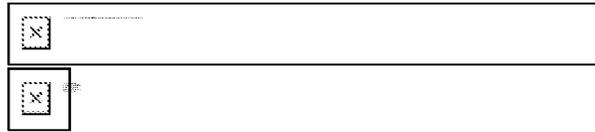
THE NEW WILD WEST

UK Group Threatens to Sue Facebook Over Cambridge Analytica

Lawyers have served Facebook with a letter before claim, the first step for filing a class action lawsuit in the UK.



Wired • [Read more »](#)



ZERO DAYBOOK

Today

- The Department of Homeland Security holds a National Cybersecurity Summit in New York.
- Senate Commerce subcommittee hearing on "global Internet governance."

Coming soon

- Senate Intelligence Committee hearing on foreign influence operations on social media tomorrow.
- Black Hat USA security conference on Aug. 8 through Aug. 9 in Las Vegas.
- DEF CON security conference on Aug. 9 through Aug. 12 in Las Vegas.

EASTER EGGS

San Antonio shark miraculously rescued after being stolen from aquarium:



San Antonio shark miraculously rescued after being stolen from aquarium

States sue government over 3-D printed guns:



States sue government over 3-D printed guns

How Bruce Lee changed Hollywood:



How Bruce Lee changed Hollywood

Share The Cybersecurity 202:  Twitter  Facebook

Trouble reading? [Click here](#) to view in your browser.

You received this email because you signed up for The Cybersecurity 202 or because it is included in your subscription.

[Manage my email newsletters and alerts](#) | [Unsubscribe from The Cybersecurity 202](#)

[Privacy Policy](#) | [Help](#)

©2018 The Washington Post | 1301 K St NW, Washington DC 20071



Democracy Dies in Darkness



[redacted] (IMD) (CON)

b6
b7C

From: [redacted] (OTD) (FBI)
Sent: Tuesday, July 31, 2018 3:24 PM
To: [redacted] (OTD) (FBI)
Cc: [redacted] (CYD) (FBI)
Subject: RE: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

Copy. Thanks, let me know if you need anything!

From: [redacted] (OTD) (FBI)
Sent: Tuesday, July 31, 2018 3:23 PM
To: [redacted] (OTD) (FBI) [redacted]
Cc: [redacted] (CYD) (FBI) [redacted]
Subject: Re: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

b6
b7C
b7E

No.

That is one of the reasons that our colleague, [redacted], and I are monitoring the DARPA Medifor program...that is our best current USG research effort to address this problem.

[redacted]
FBI - OTD
Building 27958A
Pod E
Quantico, VA 22135
(O) [redacted]
(M) [redacted]

b6
b7C
b7E

----- Original message -----

From: [redacted] (OTD) (FBI)" <[redacted]>
Date: 7/31/18 2:09 PM (GMT-06:00)
To: [redacted] (OTD) (FBI)" <[redacted]>
Subject: FW: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

b6
b7C
b7E

[redacted]
Do we have the ability to effectively detect this?

Thanks
[redacted]

[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI)
Sent: Wednesday, August 01, 2018 10:12 AM
To: [redacted] (CYD) (FBI)
Cc: [redacted] (CYD) (FBI); Karl, Larry D. (CYD) (FBI); [redacted] (CYD) (FBI)
Subject: RE: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

b6
b7C
b7E

[redacted]

Thank you!

[redacted]
Executive Staff Unit
Cyber Division
Desk: [redacted]

From: [redacted] (CYD) (FBI)
Sent: Wednesday, August 01, 2018 7:49 AM
To: [redacted] (CYD) (FBI); [redacted]
Cc: [redacted] (CYD) (FBI); [redacted]; Karl, Larry D. (CYD) (FBI); [redacted]; [redacted] (CYD) (FBI); [redacted]
Subject: RE: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

b6
b7C
b7E

Should have something later this morning, most likely red or yellow enclave.

Thanks, [redacted]

From: [redacted] (CYD) (FBI)
Sent: Tuesday, July 31, 2018 10:49 AM
To: [redacted] (CYD) (FBI); [redacted]
Cc: [redacted] (CYD) (FBI); [redacted]
Subject: FW: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

[redacted]

I hope all is well. Do you mind providing a short overview of the briefing you will provide? I will pass on this information to OCA, and they can float interest to the Committees.

b6
b7C
b7E

Thank you for your help!

[redacted]
Executive Staff Unit
Cyber Division
Desk: [redacted]

From: [redacted] (CYD) (FBI)
Sent: Tuesday, July 31, 2018 9:52 AM
To: [redacted] (CYD) (FBI) [redacted]
Cc: Karl, Larry D. (CYD) (FBI) [redacted]
Subject: RE: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

b6
b7C
b7E

[redacted]

I know A/AD Welling and former AD [redacted] wanted to be more proactive with briefings on the hill. This could be a good opportunity as you point out [redacted] and I will reach out to OCA to discuss the opportunity and follow up shortly.

Thanks,

[redacted]

From: [redacted] (CYD) (FBI)
Sent: Tuesday, July 31, 2018 8:49 AM
To: [redacted] (CYD) (FBI) [redacted]
Cc: Karl, Larry D. (CYD) (FBI) [redacted]
Subject: FW: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

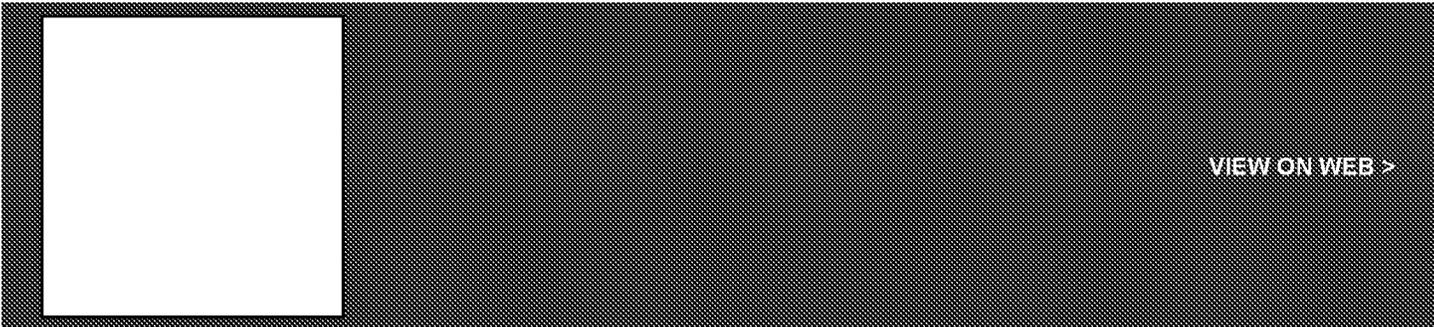
b6
b7C
b7E

[redacted] I wanted to get your thoughts and gauge the front office's interest in the main story below. [redacted]
[redacted] My folks have inquired to reaching out to OCA re possible [redacted] on the subject. I can put our pieces together and send up for review again if anyone in EM is interested.

Thanks, [redacted]

From: The Washington Post [mailto:email@washingtonpost.com]
Sent: Tuesday, July 31, 2018 7:45 AM
To: [redacted] (CYD) (FBI) [redacted]
Subject: [MARKETING] The Cybersecurity 202: Doctored videos could send fake news crisis into overdrive, lawmakers warn

b6
b7C
b7E



[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI)
Sent: Monday, July 16, 2018 10:25 AM
To: [redacted] (CYD) (FBI)
Subject: FW: meeting at APL
Attachments: Deepfake Paper v6.pdf

b6
b7C
b7E

[redacted]

FYSA.

[redacted]

[redacted]

FBI/TCIU

Desk: [redacted]

Cell: [redacted]

From: [redacted] [mailto:[redacted]]
Sent: Monday, July 16, 2018 10:19 AM
To: [redacted] (CYD) (FBI) [redacted]
Cc: [redacted]; [redacted]
Subject: RE: meeting at APL

[redacted]

Please see attached (admittedly rough) paper on Deepfake forensics. Hope it is useful.

I have cc'd [redacted] on this as well.

[redacted] this is [redacted] with the FBI, I met him at the FBI fellowship symposium, and he has expressed interest in AR supporting cyber intrusion/defense. Please see below email chain for reference.

b6
b7C
b7E

Have a good one.

V/R,

[redacted]

[redacted]

[redacted]

Graduate Student JHU-ISI / JHU-APL Fellow

[redacted]

De inimico non loquaris sed cogites

From: [redacted] (CYD) (FBI) [redacted]
Sent: Friday, July 13, 2018 5:08 PM

b6
b7C
b7E

To [redacted] [redacted]

Subject: RE: meeting at APL

[redacted]

I am interested in your paper on DeepFake.

Also, please do intro/put me in touch with [redacted] I understand you are at the beginning stages, and as for me I may shift temporarily to other topics, but I still would like to touch base with you both and find out more about uses of AR with respect to computer/network intrusions.

b6
b7C
b7E

Thanks.

[redacted]

[redacted]

FBI/TCIU

Desk: [redacted]
Cell: [redacted]

From: [redacted] [mailto:[redacted]]

Sent: Friday, July 13, 2018 9:06 AM

To: [redacted] (CYD) (FBI) [redacted]

Subject: RE: meeting at APL

[redacted]

Sorry for the confusion, I have been working on a lot of different projects. To clarify:

The additive mfr paper was my capstone, and is the one I and my group are presenting at IEEE.

The social media/cognitive hacking paper was for my intrusion detection class

The DeepFake paper was for my computer forensics class

b6
b7C
b7E

The AR project is what I am working on right now, with [redacted] no paper yet. We are looking into using AR in support of cyber intrusion/intrusion detection. Project is still in its beginning stages. Did you want me to link you up with her?

V/R,

[redacted]

[redacted]

[redacted]

Graduate Student JHU-ISI / JHU-APL Fellow

[redacted]

De inimico non loquaris sed cogites

From: [redacted] (CYD) (FBI) [redacted] >
Sent: Thursday, July 12, 2018 6:31 PM
To: [redacted]. [redacted]
Subject: RE: meeting at APL

[redacted]
It was definitely good to meet you on Tuesday.

I apologize for not getting back to you yesterday. [redacted]
[redacted]

Is the paper the additive manufacturing one, or was it the generative media paper? I think I got confused on Tuesday and thought you had a paper that was AR or generative media related.

If it was AR related, sure, feel free to send me a copy. Please specify handling caveats (I assume it is just for me [redacted] and perhaps a few other people in my immediate unit to see, and PROPIN, but let me know).

b6
b7C
b7E

If it was the additive mfg. paper, thanks for CC'ing me [redacted] is definitely the better person for that topic.

I have a lot of AR stuff that's already going out soon as I said, but follow-up pieces, and AR and generative media in general is definitely of interest.

Thanks for the additive mfg. paper, and please send the AR one if you like, too.

Let's stay in touch for future collaboration.

[redacted]
[redacted]
FBI/TCIU
Desk [redacted]
Cell: [redacted]

From: [redacted] [mailto:[redacted]]
Sent: Tuesday, July 10, 2018 10:38 AM
To: [redacted] (CYD) (FBI) [redacted]
Subject: meeting at APL

[redacted]
Was great to meet you and discuss AR as well as video and image faking tech. if you still want to take a look at my paper, let me know, will send it.

Have a good one.

b6
b7C
b7E

V/R,
[redacted]
[redacted]
[redacted]
Graduate Student JHU-ISI / JHU-APL Fellow
[redacted]



De inimico non loquaris seed cogites

[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI)
Sent: Wednesday, November 29, 2017 1:32 PM
To: [redacted] (NY) (FBI); [redacted] (CYD) (FBI)
Cc: [redacted] (CYD) (FBI); [redacted] (CD) (FBI)
Subject: RE: Meeting on GAN --- UNCLASSIFIED//~~FOUO~~

Classification: UNCLASSIFIED//~~FOUO~~
=====

b6
b7C

That works for me.

[redacted]

From: [redacted] (NY) (FBI)
Sent: Wednesday, November 29, 2017 1:18 PM
To: [redacted] (CYD) (FBI); [redacted]; [redacted] (CYD) (FBI)
[redacted]
Cc: [redacted] (CYD) (FBI); [redacted]; [redacted] (CD) (FBI); [redacted]
Subject: RE: Meeting on GAN --- UNCLASSIFIED//~~FOUO~~

b6
b7C
b7E

Classification: UNCLASSIFIED//~~FOUO~~
=====

Guys, unfortunately won't be able to make Fri at MR now. My apologies. Rather than pushing it off, how about a Lync call instead at the same time, 8 a.m. on Fri?

[redacted]

b6
b7C

From: [redacted] (CYD) (FBI)
Sent: Wednesday, November 29, 2017 7:37 AM
To: [redacted] (CYD) (FBI); [redacted]; [redacted] (NY) (FBI)
[redacted]
Cc: [redacted] (CYD) (FBI); [redacted]; [redacted] (CD) (FBI); [redacted]
Subject: RE: Meeting on GAN --- UNCLASSIFIED//~~FOUO~~

b6
b7C
b7E

Classification: UNCLASSIFIED//~~FOUO~~
=====

That works for me.

Regards,

[redacted]
Technology Cyber Intelligence Unit
FBI Cyber Division

b6
b7C
b7E

Open: [redacted]
Secure: [redacted]

From: [redacted] (CYD) (FBI)
Sent: Wednesday, November 29, 2017 7:35 AM
To: [redacted] (CYD) (FBI); [redacted]; [redacted] (NY) (FBI)
[redacted] >
Cc: [redacted] (CYD) (FBI); [redacted]; [redacted] (CD) (FBI); [redacted] >
Subject: RE: Meeting on GAN --- UNCLASSIFIED//~~FOUO~~

b6
b7C
b7E

Classification: UNCLASSIFIED//~~FOUO~~
=====

I have a team meeting from 0900 to 1000 on Friday, so if we did it at 0800, I'd be good.

[redacted]
[redacted]
Intelligence Analyst
Eurasia Cyber Intelligence Unit || Foreign Influence Task Force
Desk: [redacted]
Mobile: [redacted]

b6
b7C
b7E

From: [redacted] (CYD) (FBI)
Sent: Wednesday, November 29, 2017 6:38 AM
To: [redacted] (NY) (FBI); [redacted]; [redacted] (CYD) (FBI)
[redacted] >
Cc: [redacted] (CYD) (FBI); [redacted]; [redacted] (CD) (FBI); [redacted] >
Subject: RE: Meeting on GAN --- UNCLASSIFIED//~~FOUO~~

b6
b7C
b7E

Classification: UNCLASSIFIED//~~FOUO~~
=====

Hi [redacted]

Unfortunately, I have a meeting with [redacted] Thursday afternoon, so I don't think that would work. I'm at JEH Friday afternoon, so would be free in the morning if that works for everyone else. Probably before 10 if that's okay.
Thanks.

Regards,

b3
b6
b7C
b7E

[Redacted]

Technology Cyber Intelligence Unit
FBI Cyber Division

b6
b7C
b7E

Open: [Redacted]
Secure: [Redacted]

From: [Redacted] (NY) (FBI)
Sent: Tuesday, November 28, 2017 4:33 PM
To: [Redacted] (CYD) (FBI) <[Redacted]>; [Redacted] (CYD) (FBI)
<[Redacted]>
Cc: [Redacted] (CYD) (FBI) <[Redacted]>; [Redacted] (CD) (FBI) <[Redacted]>
Subject: Meeting on GAN --- UNCLASSIFIED//~~FOUO~~

b6
b7C
b7E

Classification: UNCLASSIFIED//~~FOUO~~
=====
TRANSITORY RECORD

[Redacted] = would you be OK moving our meeting on Thursday to the afternoon? A conflict came up – we now have to be at [Redacted] in the am. We should be back at MR by 2 p.m. or so – can you let me know if any time after 2pm on Thur works for you? If not, I can come back to MR on Fri if that works better.

Sorry for the conflict, let me know.

[Redacted]

b3
b6
b7C
b7E

[Redacted]
A/UC, Foreign Influence Task Force
[Redacted] (desk)
[Redacted] (cell)

=====
Classification: UNCLASSIFIED//~~FOUO~~

[redacted] (IMD) (CON)

From: [redacted] (CYD) (FBI)
Sent: Monday, August 27, 2018 2:57 PM
To: [redacted] (OCIO) (OGA)
Cc: [redacted] (OTD) (FBI)
Subject: Re: New Deepfakes

b6
b7C

Thanks, [redacted] images were the big thing with this last year, but now I guess video is gaining popularity.

----- Original message -----

From: [redacted] (OCIO) (OGA)" [redacted] >
Date: 8/27/18 9:21 AM (GMT-08:00)
To: [redacted] (CYD) (FBI)" [redacted] >
Subject: New Deepfakes

b6
b7C
b7E

Berkley put out "Everybody Dance now"
<https://www.youtube.com/watch?v=PCBTZh4IRis&feature=youtu.be>

Basically DeepFake to super impose person into looking like they can dance, from a source video

And related Naughty America (adult video company) is offering DeepFakes as a service.
You provide them with who (target images) you want to insert and/or locations (room in your house for example)

And they will edit a film to make it happen.

Thought you would be interested

[redacted]

b6
b7C

[redacted]
Chief Technology Officer
Office of Chief Information Officer
Federal Bureau of Investigation

[redacted] (IMD) (CON)

b6
b7C

To: [redacted] (DO) (FBI)
Cc: [redacted] (DO) (FBI); [redacted]
Subject: RE: Some interesting recent reads, ICYMI --- ~~SECRET//NOFORN~~

Classification: ~~SECRET//NOFORN~~

~~Classified By: [redacted]
Derived From: Multiple Sources
Declassify On: 50X1-HUM~~

b6
b7C

Thanks [redacted]

[redacted]

Regards,

b6
b7C
b7E

[redacted]
Technology Cyber Intelligence Unit
FBI Cyber Division

Open: [redacted]
Secure: [redacted]

From: [redacted] (DO) (FBI)
Sent: Friday, February 23, 2018 7:20 AM
To: [redacted]
Cc: [redacted] (DO) (FBI); [redacted]
Subject: FW: Some interesting recent reads, ICYMI --- UNCLASSIFIED

Classification: UNCLASSIFIED

b6
b7C
b7E

Some interesting articles on AI and Deep Fakes

[redacted]

MR 410 - [redacted]
FBIHQ 11816 [redacted]
TS [redacted]
SIPR - [redacted]
UNET [redacted]

From [redacted] (DO) (FBI)
Sent: Thursday, February 22, 2018 11:54 AM
To: [redacted] (DO) (FBI) [redacted] [redacted] (DO) (FBI)
[redacted] >
Cc: [redacted] (DO) (FBI) [redacted] >
Subject: FW: Some interesting recent reads, ICYMI --- UNCLASSIFIED

Classification: UNCLASSIFIED
=====
FYI:

b6
b7C
b7E

<< File: AI.xps >> << File: Deep_Fakes.xps >>
[redacted]

=====
Classification: UNCLASSIFIED

=====
Classification: UNCLASSIFIED

=====
Classification: ~~SECRET/NOFORN~~

Good afternoon,

Please see the attached [redacted] regarding assistance with a recent Congressional Request that was passed to ODNI. Please let me know if there are questions.

Regards,

[redacted]

Technology Cyber Intelligence Unit
FBI Cyber Division

b6
b7C
b7E

Open:
Secure

[redacted]

=====
Classification: ~~SECRET~~//~~NOFORN~~

=====
Classification: ~~SECRET~~//~~NOFORN~~

=====
Classification: ~~SECRET~~//~~NOFORN~~

[redacted] (IMD) (CON)

b6
b7C

From: [redacted] <[redacted]>
Sent: Friday, September 14, 2018 6:48 AM
To: [redacted] (CYD) (FBI)
Subject: Fwd: U.S. lawmakers call for deepfakes counter measures

----- Forwarded message -----

From: [redacted] <[redacted]>
Date: Thu, Sep 13, 2018 at 7:50 PM
Subject: U.S. lawmakers call for deepfakes counter measures
To: [redacted]

b6
b7C

U.S. lawmakers call for deepfakes counter measures

<https://venturebeat.com/2018/09/13/u-s-lawmakers-call-for-deepfakes-counter-measures/>