

ENVISIONING CYBER FUTURES WITH A.I.

JANUARY 2024

U.S. and Global
Cybersecurity Groups



ASPEN
DIGITAL

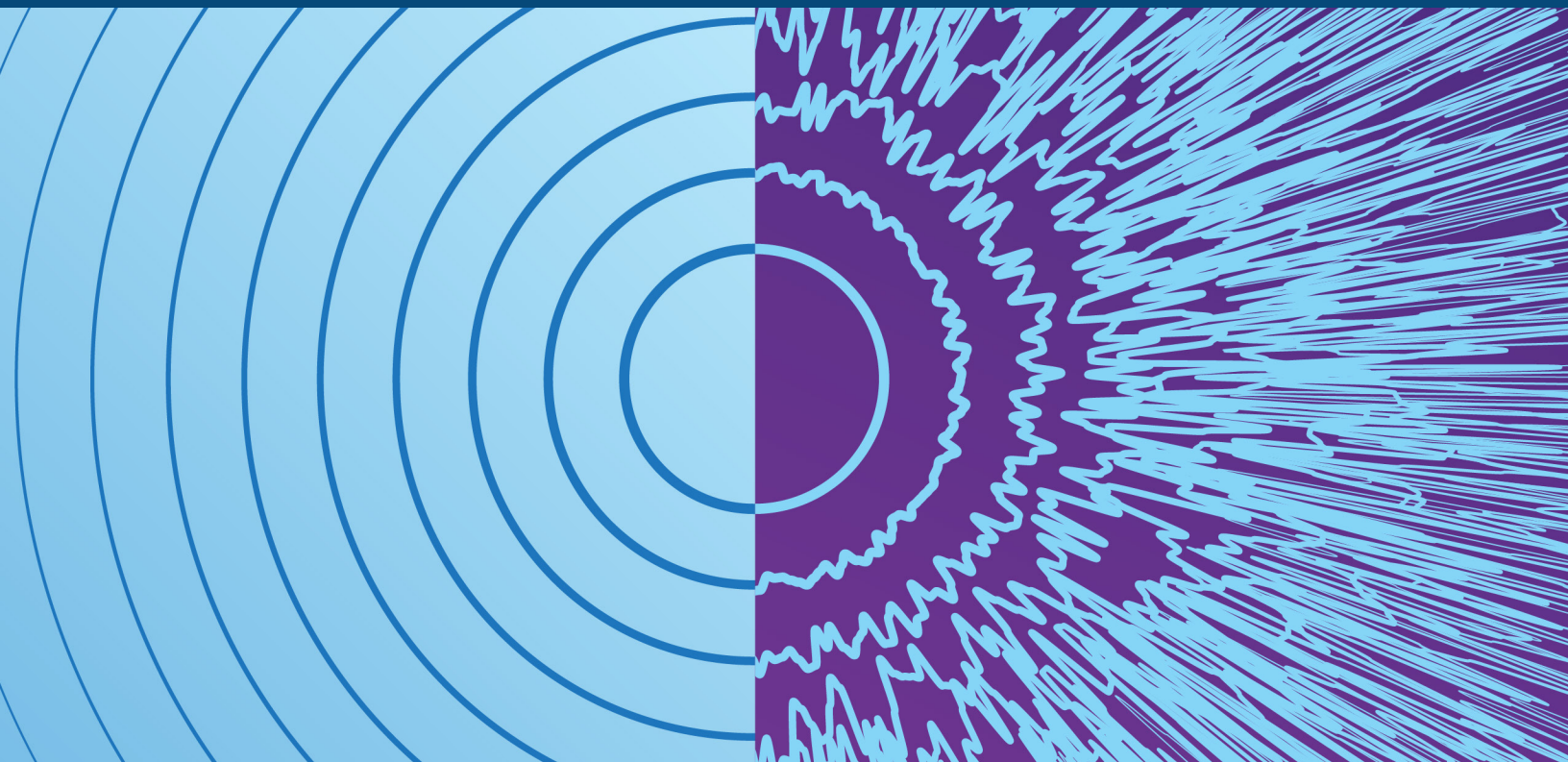


TABLE OF CONTENTS

EXECUTIVE SUMMARY

PAGE 2

INTRODUCTION

PAGE 6

THE SCENARIOS SUMMARIZED

PAGE 8

THE “GOOD PLACE”

THE “BAD PLACE”

PAGE 9

RECOMMENDATIONS

PAGE 11

GENERAL RECOMMENDATIONS

GOVERNMENT-SPECIFIC RECOMMENDATIONS

PAGE 13

INDUSTRY-SPECIFIC RECOMMENDATIONS

PAGE 14

APPENDIX

PAGE 15

“GOOD PLACE” FUTURE SCENARIO

“BAD PLACE” FUTURE SCENARIO

PAGE 21

EXECUTIVE SUMMARY

For more than a year, Artificial Intelligence (AI) has dominated public discourse, from informal conversation to serious consideration of its potential to benefit—or break—society. Discussions about the future of AI often range from the existential risk of killer robots to the potential solution for climate change. But these threats and benefits are still hypothetical, at least as of the writing of this paper. Thus, when assessing how to use AI securely, it is important to examine opportunities and dangers as they exist now or are likely to develop in the near term.

“Artificial Intelligence” in some form is not new. Computers have analyzed data and automated human tasks for decades. What’s changed is the cost of doing so, the quality of the underlying output, and how it can be used. In particular, AI’s ability to create content and automate operations represents a new frontier for the use of AI in many domains, including cybersecurity.

Technology advancements in software and data will have enormous impacts on digital security. AI-based tools, whether used for automation, cognition, or generation, can be used for both good and bad. For example, a tool analyzing network traffic for potential intrusions could also be used for making an intrusion harder to detect. And a tool used to generate human-like content can easily be used for both legitimate and illegitimate purposes.

Organizations are deploying commercial and publicly available AI tools at a rapid pace, yet often lack guidance on how to do so in a manner that enhances, not degrades, their overall security. The goal of this paper is to provide practical assistance to these organizations—the end-users of AI tools. To do so, the Aspen Institute’s Cybersecurity Program (Aspen Cyber) brought together a working group of leaders from across industry, government, and civil society to develop concrete recommendations on how to use AI as safely and effectively as possible.

To develop this guidance, the working group started from the end: defining a “good place” where AI predominantly helps defenders versus a “bad place” where AI predominantly helps

attackers. Establishing these end points gave the working group two poles against which to measure potential actions, as well as what interventions would take an organization (or, at times, broader society) toward that “good” or “bad” future. In the simplest of terms, in the “good” future, cyber defenses are more efficient, defenders collaborate more effectively, and AI tools are accountable and transparent. In the “bad” future, cyber attacks are far easier to develop, criminals share proprietary information and datasets, and there is no recourse when AI tools are misused. The two future scenarios are summarized in the following section and the full text of each is included in the appendices.

So, what actions would move towards that “good place?”
The working group’s key recommendations include:

- **Stay true to cybersecurity principles.** The basics of cybersecurity always apply, especially when using AI.
- **Don’t live in a silo.** AI and cybersecurity practitioners should work together.
- **Proactively manage which decisions AI will be making.** AI tools will be making decisions organizations cannot review, so deploy them with forethought and careful planning; make affirmative choices as to what they will be able to do.
- **Improve logging, log review, and log maintenance.** AI-powered attacks will be increasingly difficult to detect; keeping and reviewing data will be essential.
- **Be intelligently transparent about AI.** Organizations should think about what outcomes they actually hope to achieve through AI transparency and determine how to structure the disclosure accordingly—or even whether disclosure is necessary.
- **Make sure your contracts contain AI rules of engagement.** Even if an organization doesn’t use AI, its partners likely will. Organizations should consider flowing down their own AI policies to partners and third parties.
- **Beware of the bandwagon.** Install AI tools where it makes operational or other sense—it is OK to say no to AI, or to use other technologies if an AI tool is not warranted.

ASPEN US CYBERSECURITY GROUP

CHAIRS

Congresswoman

Yvette Clarke

Co-Chair, U.S House of Representatives

Yasmin Green

CEO, Jigsaw, Google

Christopher Krebs

Co-Chair, Senior Newmark Fellow in Cybersecurity, Aspen Digital

Gary Steele

President & CEO, Splunk

STAFF

John P. Carlin

Strategic Advisor and Chair Emeritus for Cybersecurity, Aspen Digital

Jeff Greene

Senior Director, Cybersecurity Programs, Aspen Digital

Yameen Huq

Director, US Cybersecurity Group, Aspen Digital

MEMBERS

Katherine Adams

Senior Vice President and General Counsel, Apple

Marene Allison

Former CISO

Sara Andrews

Global CISO and Senior Vice President, Pepsico

Monika Bickert

Head of Product Policy and Counterterrorism, Facebook

Geoff Brown

Senior Vice President, Arete

Tom Burt

Corporate Vice President, Customer Security and Trust, Microsoft

Vinton G. Cerf

Chief Internet Evangelist, Google

Dr. Lorrie Cranor

Director, CyLab Security & Privacy Institute

Michael Daniel

President, Cyber Threat Alliance

Noopur Davis

Corporate EVP, Chief Information Security and Product Privacy Officer, Comcast

John Demers

Corporate Secretary, The Boeing Company

Jim Dempsey

Policy Advisor, Stanford Program on Geopolitics, Technology, and Governance

Donald R. Dixon

Co-Founder & Managing Director, ForgePoint Capital

Sue Gordon

Rubenstein Fellow, Duke University

Vishaal Hariprasad

Co-founder and CEO, Resilience

Niloofer Razi Howe

Senior Operating Partner, Energy Impact Partners

Sandra Joyce

Executive Vice President, Mandiant Intelligence & Government Affairs

Sean M. Joyce

Head of Global and US Cybersecurity and Privacy, PwC

Jodie Kautt

Vice President, Cyber Security, Target

Sam King

CEO, Veracode

Dr. Herb Lin

Senior Research Scholar for Cyber Policy and Security, Stanford University

Brad Maiorino

Corporate Vice President and CISO, Raytheon Technologies

Jeanette Manfra

Global Director for Security and Compliance, Google

Chandra McMahon

CISO, CVS Health

Tim Murphy

Chief Administrative Officer, Mastercard

Craig Newmark

Founder, Craig Newmark Philanthropies

Dr. Gregory Rattray

Adjunct Professor, Columbia University SIPA

Nasrin Rezai

Senior Vice President and CISO, Verizon

David Sanger

National Security Correspondent, The New York Times

Dr. Phyllis Schneck

Vice President and CISO, Northrop Grumman Corporation

Bruce Schneier
Fellow, Berkman-Klein
Center & Lecturer,
Harvard Kennedy School

Charley Snyder
Head of Security Policy,
Google

Alex Stamos
Adjunct Professor,
Stanford University

Alissa Starzak
Vice President Global
Head of Public Policy,
Cloudflare

Bobbie Stempfley
Vice President of
Cybersecurity, Dell
Technologies

Scott C. Taylor
Board Member,
Strategic Advisor, and
Former General Counsel

Dr. Hugh Thompson
Managing Partner,
Crosspoint Capital
Partners

Jack Weinstein
Professor, Boston
University

**Dr. Jonathan W.
Welburn**
Researcher, RAND
Corporation

Michelle Zatlyn
Co-founder & COO,
Cloudflare

WORKING GROUP PARTICIPANTS

Lily Arstad
Business Administrator,
Microsoft

William Bartholomew
Director of Public Policy
—Office of Responsible
AI, Microsoft

Archer Batcheller
Cyber Systems Engineer,
Northrop Grumman
Corporation

Robert Brown
Science and Technology
Branch Executive
Assistant Director, FBI

Caroline Budnik
Head of Cyber Policy,
Northrop Grumman
Corporation

Adam Cohn
Vice President of
Worldwide Government
Affairs, Splunk

Jane Courtney
Supervisory Special
Agent, FBI

Lisa Einstein
Senior Advisor for
Artificial Intelligence
and Executive
Director for the CISA
Cybersecurity Advisory
Committee, CISA

Steve Kelly
Chief Trust Officer,
Institute for Security and
Technology

Christine Lai
Cybersecurity R&D,
CISA

Joe Levy,
President, Sophos
Technology Group

Michael Massetti
Senior National
Intelligence Officer for
Emerging Technology,
FBI

Holly Omans
Cyber Security Manager,
Verizon

Chloe Ryan
Special Assistant to EAD
STB, FBI

Sezaneh Seymour
VP and Head of
Regulatory Risk and
Policy, Coalition

Jordana Siegel
Cybersecurity and Data
Protection Policy, AWS

Jono Spring
Cybersecurity Specialist,
CISA

Joseph Szczerba
Cyber Division Section
Chief, FBI

Toni G. Verstandig
Co-Founder and
Executive Director,
Verstandig Family
Foundation

Bryan Vorndran
Cyber Division Assistant
Director, FBI

Justin Williams
Science and Technology
Branch Section Chief,
FBI

Hallie Zimmerman
Global Cyber Policy,
Northrop Grumman
Corporation

**National Institute of
Standards and
Technology**

Note: The U.S. Government does not endorse any product, service, or enterprise. Views expressed in this document do not necessarily represent the views of the U.S. Government or any institution or organization with which the authors or working group participants are affiliated.

INTRODUCTION

The public release of ChatGPT was a rare moment when a new technology immediately dominated the public psyche. In the year since, that conversation has not faded; in fact, the implications of AI tools (generative or otherwise) and how they could change society for better and for worse have become part of a broad range of policy debates, from the impact on the workforce to education to national security. This includes the impact on cybersecurity—will generative AI supercharge new attacks? Will defenders use it to detect malicious activity earlier and faster?

Will AI supercharge new attacks or will defenders use it to detect malicious activity earlier and faster?

AI is not new to cybersecurity. Both defenders and attackers have used machine learning and AI tools for years, but experts agree that the public availability of generative AI will reshape the cybersecurity landscape. However, there is no consensus on how this will occur. Nor is there much guidance on what the end-users of these AI tools (whether companies, governments, individuals, or other organizations) can do right now to maximize AI's utility to defenders and minimize the benefit to attackers. This paper will fill that gap (at least as it stands at the time of writing, January of 2024).

One difficulty in developing guidance for using emerging AI tools securely is that it is still largely speculative; we just don't know how attackers or defenders will use these technologies. As a result, the debate over AI's impact on cybersecurity has largely been theoretical—academic discussions over an undefined potential future. Therein lies the difficulty in developing guidance for what decisions and actions we should take today—as the great philosopher Yogi Berra once said, “If you don't know where you are going, you might wind up someplace else.”

But we cannot wait for certainty—governments, companies, and organizations of all sizes are rushing to adopt AI tools and are making long-term decisions now. They are looking for guidance on what they can do both to maximize the utility of these tools and to limit any future harm.

Governments, companies, and organizations are looking for guidance on what they can do maximize the utility of AI tools and to limit any future harm.

In this paper, Aspen Cyber takes a significant step in providing that guidance. Before thinking about recommendations, however, we needed to know where we wanted to go—to define both the future that we all want to reach and the future we want to avoid. We challenged groups comprised of US Cybersecurity Group members and affiliates to describe their best assessment of two futures: one where generative and other AI tools have given defenders a significant advantage over attackers (the “good place”) and another where the attackers have significant advantages (the “bad place”).

With these possible futures in hand, the group met in person and virtually, and developed the following recommendations that we believe will help steer us toward the good place and away from the bad. Not all of these recommendations will be applicable to every organization, but all should find some that address their needs.

A final note: we placed some constraints around the future scenarios. The groups considered both existing and potential AI tools that could reasonably be available in the coming years. However, they were to not to imagine a future with an AGI—Artificial General Intelligence—with consciousness or reasoning capacity equal to or better than a human. Thus, the good place is not a future where Tony Stark’s sentient Jarvis can protect the world from all forms of evil, and the bad place is not patrolled by Skynet’s terminators. So, while the scenarios below might make for boring science fiction movies, we believe them to be a more realistic prediction of what will come to pass in the near future.

THE SCENARIOS SUMMARIZED

THE “GOOD PLACE”

AI tools will give defenders the edge if they are able to improve security response times, augment human expertise, and improve software and device security. In this world, AI tools sort through an enormous volume of data for a variety of ends: prioritizing vulnerabilities for remediation, detecting data exfiltration, identifying unusual user behavior, and much more.

With these new insights, the tools are tuned to mitigate confirmed threats (such as through isolating endpoints, blocking malicious URLs, or sandboxing malicious operations) and to escalate those that need more analysis. AI also enhances the end-user experience, accurately and efficiently assisting users when they report anomalies or proactively alerting them and helping to address issues that the tools themselves identify. This early and accurate detection reduces response time, minimizes wasted efforts on false positives, and helps flag true threats that otherwise could go undetected. Defenders are thus able to focus their limited resources on investigations that could require human understanding.

Humans are kept at the center and know when they're interacting with an AI system, any potential limitations and risks in the outputs, and in higher-risk scenarios can intervene or override the AI system.

AI would also be key to bringing secure-by-design principles to life. AI tools would write new, secure code and assist in updating existing code by finding and fixing vulnerabilities. It would even rewrite existing applications in more secure languages. The tools would continually update and improve code as attacks evolve or researchers discover new vulnerabilities.

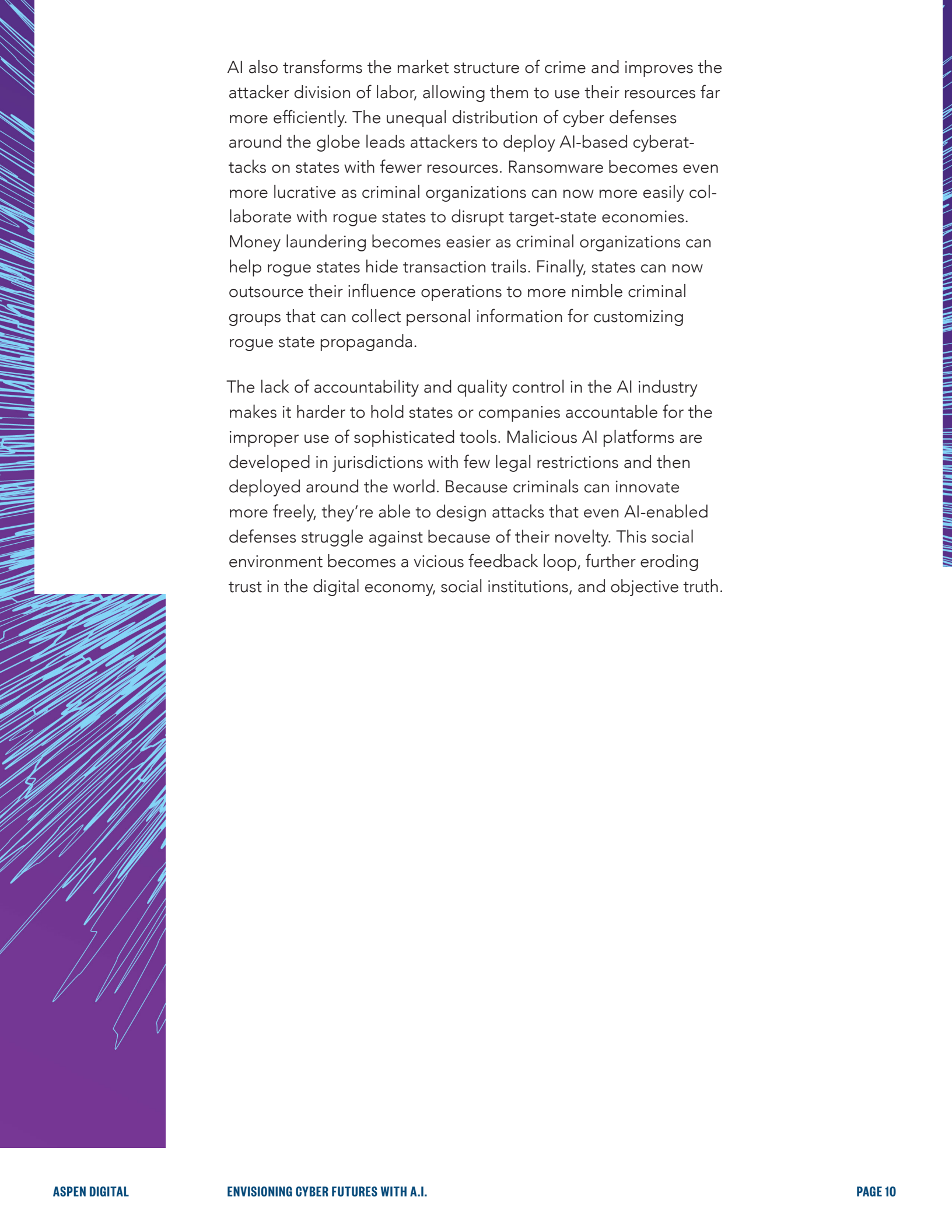
The AI tools themselves would be secure and accountable. They generate high-quality, accurate results and are accountable when they fail to do so. This includes the ability to distinguish between malevolent tampering and so-called “hallucinations” in the tools. Humans are kept at the center and know when they’re interacting with an AI system, are aware of any potential limitations and risks in the outputs, and can intervene or override the AI system in higher-risk scenarios. AI powered cybersecurity tools maintain an edge of attackers because they have access to multiple high-quality sources of data, whereas attackers must rely more on “black-market” data.

THE “BAD PLACE”

AI tools will give attackers the edge if they are able to improve their attack effectiveness, enable criminal collaboration, and learn more quickly than the defenders can adapt and respond. In this world, AI tools reduce the barrier to entry of engaging in crime and make it significantly easier to develop sophisticated social engineering techniques, evade detection, design bespoke malware, and more.

AI also transforms the market structure of crime and improves the attacker division of labor, allowing them to use their resources far more efficiently.

With these new efficiency gains, AI makes cyberattacks both less expensive and more effective, raising the expected payoffs of crime. Using AI tools, criminals can more quickly find vulnerabilities to exploit in existing systems. In addition, they can create personalized spear-phishing campaigns to increase the likelihood of success. Once they gain entry to a system, criminals can use bespoke malware that can be tailored to the specific target. After launching many of these campaigns, they can use machine-learning to see what’s effective and modify their strategies and malware in future campaigns.



AI also transforms the market structure of crime and improves the attacker division of labor, allowing them to use their resources far more efficiently. The unequal distribution of cyber defenses around the globe leads attackers to deploy AI-based cyberattacks on states with fewer resources. Ransomware becomes even more lucrative as criminal organizations can now more easily collaborate with rogue states to disrupt target-state economies. Money laundering becomes easier as criminal organizations can help rogue states hide transaction trails. Finally, states can now outsource their influence operations to more nimble criminal groups that can collect personal information for customizing rogue state propaganda.

The lack of accountability and quality control in the AI industry makes it harder to hold states or companies accountable for the improper use of sophisticated tools. Malicious AI platforms are developed in jurisdictions with few legal restrictions and then deployed around the world. Because criminals can innovate more freely, they're able to design attacks that even AI-enabled defenses struggle against because of their novelty. This social environment becomes a vicious feedback loop, further eroding trust in the digital economy, social institutions, and objective truth.

RECOMMENDATIONS

The recommendations that follow are valid as of the publication date of this paper, January 9, 2024.

GENERAL RECOMMENDATIONS

- **Avoid the hype.** Resist the temptation jump on the AI bandwagon; use an AI tool where it makes operational or other sense.
- **Proactively manage which decisions AI will be making.** AI tools will be making decisions organizations cannot review individually, so deploy them with forethought and careful planning. It is important to make affirmative choices as to what decisions an AI tool will be making and at what level. The factors below can help to: (1) assess the potential benefits and harms of using an AI tool under consideration and (2) identify the actions or processes that must remain in the decision-making loop:
 - **How much human cognition is required?** Is it a repetitive and tactical process or a creative and strategic decision? The latter is more likely to require continued human involvement.
 - **How much quality control or review is required of the action, process, or end result?** As quality becomes paramount, lean towards deliberate human review.
 - **What is the impact or risk from an incorrect decision?** The more severe, the more humans should stay in control.
 - **How frequent are the decisions made and how important is it to make them at speed?** AI excels at making repetitive decisions and at moving at a pace humans could never achieve.
 - **Does the AI tool supplement human decision-making or take its place?** If the latter, weigh the costs and benefits of AI error vs. human error and develop a fail-safe and review mechanisms for mission critical choices.
 - **Is the decision irreversible?** If so, move cautiously and make sure the organization can survive an irreversible bad decision.

- **Know what types of data the AI is using.** Humans can never know all the data that an AI uses, but to the extent possible, organizations should understand what data sources are used. Prioritize stronger security controls for data that is internal or proprietary—the “crown jewels.”
- **Saying “no” is OK.** Before deploying, building, or completing an AI tool, assess whether its apparent benefit will greatly outweigh potential harm. If a tool appears to have significantly more foreseeable harms than benefits, don’t use or build it, or at minimum ensure your organization can effectively control it.
- **Realize that the old rules still apply.** AI tools can seem new, shiny, and powerful, but do not ignore the established basics of information technology security, cybersecurity, and data security. Virtually all the long-standing tools and best practices are applicable to most AI development and use cases.
- **Be intelligently transparent.** Transparency is a good thing, but organizations should avoid turning an important notification into background noise or a meaningless click through—commonly known as “consent fatigue.” Notifications should disclose information relevant to the user and utility of the tool, including privacy concerns or the amount of human oversight. Organizations should avoid oversaturating consumers with disclosures to the point they are meaningless or ignored.
- **Think about social media (because AI is reading it).** Organizations should update social media and communications policies to recognize that large language models (LLMs) are using posts and other communications as training data, and also to account for information that adversaries could exploit.
- **Log, log, and log more.** Good logs are essential to cybersecurity and the potential for AI-driven exploits and attacks only heightens that. Organizations should improve logging, log review, and log maintenance to maximize ability to detect novel, AI-generated attacks and comply with legally authorized reviews as necessary (including through use of AI tools).
- **Keep humans in the code loop.** AI-written code should be more secure than human-written code, but it is still important maintain human and technical review for best practices in vulnerability management.

- **Don't silo AI from other IT, cyber, and other disciplines.** Bring together all relevant teams, such as cybersecurity, legal, data science, program/product teams, and executive leadership, on a regular cadence to collaborate on AI cybersecurity risk. Organizations should consider developing a new role for Chief Artificial Intelligence Officer when AI tools have a significant impact on the organization's goals or operations.
- **Be resilient.** Organizations will need a resilience plan in the event AI tools are disrupted, including training their workforce to perform AI-automated tasks so that they can maintain mission-critical operations.
- **Reflect AI in contractual needs and obligations.** Contracts with vendors, partners, and others may need to include limitations on proprietary data, including:
 - **What data would be provided, especially if it involves an AI vendor.**
 - **How the data will be used.**
 - **How the data will be secured.**
 - **Whether the data will be used for training other models.**
 - **What will happen to that data if the business relationship is concluded.**
- **Create a culture of openness.** AI is already powering phishing emails and other scams that often target junior staff. If staff is afraid to reach out to senior leaders, they are more likely to fail to report risks or fall for scams impersonating executives. Empower staff to reach out to senior leadership to ensure that a communication (and in particular directions to take actions or distribute funds) is legitimate.

GOVERNMENT-SPECIFIC RECOMMENDATIONS

- **Identify high-risk AI tools.** Governments should identify AI tools that could cause extreme harm and monitor their use. In situations where an AI tool has use cases with high risks to society, governments should consider acquiring the intellectual property for the tool and license it for specifically the lower-risk use cases.

- **Promote access to open source cybersecurity tools.** Help organizations below the cyber poverty line access open source cybersecurity tools that help protect against AI-based attacks, review code, and provide training data.
- **Provide educational opportunities.** Support university programs and certifications that integrate AI, data science, and cybersecurity skills.

INDUSTRY-SPECIFIC RECOMMENDATIONS

- **Stick to the basics.** AI tools and models are fundamentally software, and developers and deployers should employ existing cybersecurity, resilience, and secure-by-design principles. This includes:
 - **Trust and Authorization**
 - **Identity and Access Management**
 - **Asset Management**
 - **Network Access Control / Quarantine Policies**
 - **Vulnerability Management**
 - **Continuous Monitoring**
- **Make information sharing easy and commonplace.** Companies should use existing standardized security information sharing structures, such as Structured Threat Information eXpression (STIX); Trusted Automated eXchange of Intelligence Information (TAXII), and the national vulnerability database for AI cybersecurity purposes. Where these are not well suited, companies should work together and with the government to develop protocols that facilitate quick and easy sharing.
- **Log by default.** Developers of AI models or tools should build logging into AI tools for cybersecurity, audit, and other legally authorized purposes.

APPENDIX

“GOOD PLACE” FUTURE SCENARIO

BACKGROUND

Modern AI tools could make the world safer from cybersecurity threats by helping organizations rapidly identify and respond to threats and improve the efficiency and effectiveness of their cybersecurity workforce. Corporate executives and IT security leaders could use AI to optimize decision-making, assess risks, and make financially-sound operational decisions. Security teams could use AI tools to analyze enormous volumes of data and detect anomalous activity or malicious users. This would enable defenders to also focus more on investigations that require human understanding.

How could this world exist and what would it look like? This future will be possible if defenders harness AI's advantages over attackers. Below are specific descriptions of how AI tools could provide those distinct advantages.

AI tools make [greater efficiency] possible by processing large quantities of data and identifying a myriad of anomalous behavior.

FUTURE SCENARIO: A.I. GREATLY IMPROVES CYBERSECURITY

IMPROVEMENTS IN EFFICIENCY FOR DEFENDERS

AI enhances the ability of organizations to run existing cybersecurity processes more efficiently, at greater speed, and with fewer resources. AI tools make this possible by processing large quantities of data and identifying anomalous behavior. These AI tools detect threats earlier and more accurately to enable analysts to take action more quickly. While attackers traditionally benefit

from the asymmetric advantage of attack execution, defenders are dependent on uncertain detection signals. AI tools shifted this balance. Specific capabilities include:

- **Vulnerability prioritization** — Defenders analyze vulnerabilities in terms of risk and cost to mitigate to optimize return on investment for cybersecurity spending.
- **Network flow data** — Defenders analyze network flow data to find unusual data transfers and unauthorized remote access by recognizing deviations from normal network behavior, enabling early detection of and response to malicious activity.
- **User behaviors** — Defenders analyze user behaviors such as login times, locations, and activities for deviations from typical behavior to detect insider threats.
- **Potential malware** — Defenders analyze files and system processes to find potential malware that is undetectable with traditional signature-based tools.
- **Endpoint data** — Security teams analyze endpoint data to find signs of unauthorized device access and processes or misuse of legitimate tools (i.e., ‘living off the land’ attacks).
- **Isolating infected endpoints or processes** — Security teams take proactive steps to isolate endpoints and mitigate suspected intrusions before they result in significant compromise or lateral movement across systems.
- **Phishing attempts** — Organizations analyze language in emails or other communications to detect and block phishing attempts.

Early and accurate detection reduces wasted efforts on false positives and helps flag true positives that may otherwise go undetected for further investigation. Assessment teams use AI tools for more effective decision-making through better metrics, visualizations, and decision-trees. Some examples include:

- **Metrics** — Organizations understand their cybersecurity effectiveness through automated analysis of mitigations, incidents, and responses.

- **Visualizations** — Defenders prioritize cybersecurity efforts by dynamically visualizing relevant data, such as network traffic, access patterns, and user behavior, to see anomalies in real-time.
- **Decision-trees** — Security leaders use customized contextual decision trees based on AI analysis of the impact and confidence levels of a particular incident.

Where malicious activity is suspected, AI tools deploy automated security measures to reduce the time between detection and mitigation:

- **Response time** — AI tools respond to deviations in patterns and deploy automated responses with minimal human intervention, thus minimizing the duration and impact of adverse events.
- **Iterative response time** — AI tools are trained with post-incident reports to refine the criteria by which they detect, assess, and respond to various scenarios. Every intrusion is a learning opportunity and improves the responsiveness of security tools in the future.
- **Detection quality** — AI tools are, on an ongoing basis, trained on event data to constantly refine their decision-making and detection capabilities.

These capabilities have several effects on the cybersecurity workforce:

- **Improved productivity** — AI tools increase efficiency and reduce the number of people required for cybersecurity tasks, thereby reducing the overall cyber workforce gap.
- **Increases workforce satisfaction** — AI tools perform mundane, tedious, or routine tasks, freeing up cybersecurity personnel to work more challenging problems that increase job satisfaction and reduce burnout.
- **Speeds onboarding** — AI tools enable new workers to integrate into the cybersecurity workforce more rapidly.

Finally, AI can improve the security and quality of existing and new code:

- **Code assessment** — AI tools can scan existing and new code for variants of vulnerable code patterns that would be missed by traditional static analysis.
- **Code recommendations** — AI tools are used to analyze secure coding practices and recommend improvements to legacy code, reducing the time spent on manual code analysis and rewriting.
- **Code monitoring** — AI tools examine code in real-time as it is developed and proactively identify vulnerabilities or deviations from secure coding practices.
- **Code forecasting** — AI tools are used to analyze how existing code as well as potential future modifications could lead to vulnerabilities.
- **Rewriting Code** — AI tools can rewrite legacy code using safer modern patterns, languages, and libraries.
- **Code automation** — AI tools can automatically generate code patches that mitigates the threat risks that it has identified.

BETTER RELATIONSHIPS IN THE DEFENDER SUPPLY-CHAIN

AI tools are also improving service quality and how users are treated in a cybersecurity process. Organizations are using AI tools to develop more effective customer service chatbots that adapt to a user's knowledge and capabilities. Some ways that AI tools accomplish this include:

- **Improved response time** — AI tools generate automated responses tailored for the users' role, environment, and the problem they are experiencing. These responses use log and event data to propose, or automatically apply, the most appropriate solutions for the problem leading to earlier and more comprehensive resolution.
- **Quality user engagement** — AI tools analyze which solutions and engagement models resulting in high user satisfaction to inform best-practices on user engagement.

- **Adapting to user needs** — AI service tools assess the knowledge and skill level of a user and provide instruction that is suited to an individual's needs, including elevating for human intervention when an end-user cannot remediate a situation.
- **Labor savings** — AI tools better utilize cybersecurity experts by addressing matters that don't require human analysis and provide cybersecurity experts with tailored background information and possible interventions for matters that require human analysis.

IMPROVEMENTS IN A.I. ACCOUNTABILITY AND QUALITY

AI tools are generating high-quality, accurate results and are accountable when they fail to do so. Organizations can distinguish between malevolent tampering and so-called "hallucinations" in Large Language Models, or LLMs, which provide free-text outputs. Humans are kept at the center and know when they're interacting with an AI system, any potential limitations and risks in the outputs, and in higher-risk scenarios can intervene or override the AI system. Organizations understand AI outputs and can distinguish between instances of malevolent tampering, potentially turning over such instances to the government. This is enabled by:

- **An AI social contract** — Defenders worked with governments to develop a commonly accepted and easily understood social contract that lists out key ethical responsibilities around the creation, use, and governance of AI models in cybersecurity.
- **Defender-lead innovation** — Defenders adopt a culture of innovation around AI models and develop processes to learn from experience and encourage controlled experimentation with regards to how models are created and deployed.
- **Greater control** — Defenders have access to the models, data, and previous output, including those found to be factually incorrect. This allows them to identify more quickly when outputs are the result of tampering vs. analytical errors. Attackers do not have access to this breadth of data or the tools and infrastructure needed to analyze the data that they do have.

- **Model monitoring** — Defenders monitor systems in real-time and use other AI tools to detect anomalous engagement with their models. This allows for early detection of tampering with inputs and detection of manipulated outputs, which can be blocked to prevent negative outcomes.
- **Model transparency** — Defenders understand the capabilities and limitations of models and the impact it has on their scenarios; they use risk frameworks, systemic measurements, and evaluation tools to ensure that their AI systems are safe, secure, and reliable.
- **Iterative improvements** — Defenders fix models even when tampering occurs, rendering such tampering attempts useless. This capability forces attackers to constantly innovate with how they tamper in the future, increasing costs for the attacker.
- **Generative AI countermeasures** — Defenders use content provenance and AI content detection to prevent deepfake content polluting their platforms.

Defenders have the advantage over most attackers in the ability to use AI more effectively as the best models require vast amounts of high-quality data that are available only to the largest organizations or most sophisticated nation states. Therefore, AI tools are more effectively wielded by governments and companies than by criminal enterprises. Defenders also benefit here because:

- **Data breadth** — Defenders have multiple readily available high-quality sources of data for training their model whereas attackers must rely more on “black-market” data. Black-market products in general are of poorer quality due to their limited size, the cost of acquisition, and the lack of quality monitoring.
- **Data depth** — Defending organizations can use larger quantities of data more easily due to greater capacity, labor specialization, and purchasing power.
- **Data iteration** — Defenders can get constructive feedback on the quality, limitations, and uses of their data from suppliers and customers alike. This lets them iterate on existing models and sources more effectively than attackers.

“BAD PLACE” FUTURE SCENARIO

BACKGROUND

Modern Artificial Intelligence tools could empower attackers and disadvantage defenders if criminals and rogue nations can harness them to improve their attacks, collaboration, and learning faster than defenders can adapt. Enterprising cybercriminals could use AI tools to write malware regardless of their coding knowledge. They could partner with a rogue nation, using the nation’s access to high-quality data to train AI models. Both the criminals and the rogue nations would then improve their attack techniques and ability to avoid detection, using them for theft, espionage, or destructive attacks. Because the attackers would be using models and datasets that defenders could not access or use, defenders could only react to each intrusion by which time the attackers can move on to a new AI-generated technique. The speed and power of these tools would minimize cost of modifying attacks, while the cost of defending against them increased at an uncontrollable pace.

How could this world exist and what would it look like? This future will be possible if AI gives attackers distinct advantages over defenders. Below are specific descriptions of how AI tools would function in this future.

FUTURE SCENARIO: A.I. GREATLY HARMS CYBERSECURITY IMPROVEMENTS IN EFFICIENCY FOR CRIMINALS

AI makes cyberattacks simpler and less expensive. The incentives for crime are higher because AI tools can improve the payoffs and success rate of several types of attacks, such as the following:

- **Concentrated assets in targets** — The use of generative AI to complement core business functions both creates and concentrates more sensitive data in specialized AI systems, raising the payoff for criminals to exploit them.
- **Bespoke malware** — AI tools generate on demand, bespoke malware that can be tailored to specific targets.

- **Weak deterrence** — Criminals care less about violating rules like intellectual property laws and ethical guidelines and have fewer constraints on their use of AI tools.
- **AI automation for cryptojacking** — Gangs use AI-based automation scripts to harness the computational power of victims' machines and improve the efficiency and payoff of cryptojacking.
- **Generative AI in spear-phishing** — Attackers use generative AI to create more personalized phishing emails and believable sender personas that increase the likelihood of a success.
- **Machine learning for OSINT** — Machine learning tools help attackers better understand their targets through improved analysis of publicly available data sources.
- **Machine learning for unauthorized access** — Machine-learning tools help attackers search the Internet to find vulnerable systems more easily, thereby increasing the likelihood of successful attacks.
- **Machine learning for superior malware** — Attackers incorporate machine learning into malware that allows it to learn from experience and modify its behavior dynamically to avoid detection.
- **Attack surface enumeration** — Machine learning tools help malware find valuable assets more effectively.
- **Vulnerability discovery** — Machine learning tools help attackers find vulnerabilities to exploit for access to enumerated asset management systems.
- **Generative AI for disinformation campaigns** — Attackers use generative AI to create fake audio and video content ("deepfakes"), improving the likelihood of deception for both macro-targeted disinformation campaigns and micro-targeted social engineering campaigns.
- **AI corrupts AI** — AI-enabled attacks detect and avoid AI-enabled defenses, rendering them useless and exploitable.

BETTER RELATIONSHIPS IN THE CRIMINAL SUPPLY-CHAIN

AI tools also affect the types of entities involved in cybercrime and how they interact with one another. Initially, AI models used massive high-quality datasets that were generally only available to legitimate organizations and responsible nations. Over time, criminal groups and smaller rogue nations responded by allying and integrating their efforts to create comparable tools for malicious use. This integration enhances the abilities of criminal actors:

- **Better attacker division of labor** — Criminals and nations have different strengths and their partnerships evolve to create an attacker economy of scale. Nations provide resources and target lists, while criminal organizations have specific expertise or a willingness to mount attacks that even rogue nations might not be willing to do. AI analysis of intended targets is used to identify potential collaborators and the attacks most likely to be successful.
- **Unequal global distribution of defenses** — Nations with fewer financial and computational resources are unable to deploy effective defenses against AI attacks. Conversely, improvements in the defenses of wealthier nations led attackers to focus on those with weaker defenses.
- **Improved ransomware deployment** — Criminal organizations skilled in using ransomware collaborate with rogue nations to disrupt target-state economies.
- **Improved influence operations** — Smaller, more nimble criminal groups collect personal information to tailor nation-state propaganda.
- **Malicious AI model development** — Rogue nations provide criminal organizations with data and infrastructure to be used as a training ground for malicious AI model development.
- **Integrated money laundering** — Criminal organizations aid nation states with money laundering by providing connections that bypass economic barriers and using AI to generate believable transaction trails.
- **Inconsistent compliance with legal boundaries** — While legitimate organizations comply with national laws and international agreements that limit AI, criminal organizations use AI across borders without limitation.

In addition, the democratization of AI tools reduced the barrier to entry for potential actors looking to get involved in cybercrime:

- **Cascading costs** — AI both reduces barriers to entry and increases economies of scale, both of which amplify the scale and size of attacks and enable more effective changes in tactics.
- **Simple hacking toolkits** — Criminals use automated hacking tools that require minimal knowledge but can penetrate sophisticated corporate or government defenses.
- **Improved phishing prompts** — Criminals use readily available generative AI prompts to generate personalized, culturally specific phishing content that they can distribute at scale.
- **Scam content** — Criminals can easily generate deepfakes for large-scale social media campaigns,
- **AI-based ransomware** — Criminals use ready-made AI-based ransomware that uses large volumes of data from previous victims' behavior to maximize the probability of payment.

Criminals have no limits creating attack tools because they do not follow the same ethical norms and rules that constrain legitimate developers

On the target-side, organizations using AI must rely on a small number of vendors that possess these large data sets. This creates concentrated points of vulnerability in the supply-chain that attackers can exploit to impose large-scale costs on their targets. Attackers can exploit these points by:

- **Poisoning the training data** — Criminals use AI-based application programming interfaces (APIs) to manipulate training data and models to further sabotage users or facilitate wide distribution of vulnerabilities.

- **Masked network traffic** — Criminals can create synthetic traffic that's harder to sort from human traffic.
- **Backdoor compromise** — Criminals place backdoors into AI models being used by customers; thereby gaining access into the customer systems as well.
- **Common vulnerabilities** — Criminals can exploit a particular vulnerability to hit multiple companies because all companies rely on the same underlying AI vendor.
- **Ransomware-as-a-Service (RaaS) attacks** — Criminal enterprises centered around RaaS can scale ransom payments more easily using AI-based attacks.
- **Spying across shared resources** — Criminals exploit shared AI infrastructure used by multiple companies to exfiltrate sensitive information.

DIFFICULTIES IN A.I. ACCOUNTABILITY AND QUALITY

AI tools regularly generate poor-quality, inaccurate results which are hard to distinguish from non-AI content. Moreover, models and developers are not held accountable for these errors. By reducing the cost of launching certain types of cyberattacks, adversaries can now outsource their attacks to smaller, more hidden entities thereby making it harder to hold such states accountable. The cost of defending against AI-enabled attacks far outpaces the cost of developing them and the capability gap is widening:

- **Global reach** — Even though some states adopted legal safeguards, AI platforms are developed and used for malice in legally permissible jurisdictions.
- **Struggle with novelty** — AI-enabled defenses struggle against attacks that are not part of their training sets.
- **Permissionless innovation among criminals** — Criminals have no limits creating attack tools because they do not follow the same ethical norms and rules that constrain legitimate developers.
- **No duty of care** — AI companies do not have clear legal obligations to protect their data and models and as a result many do not adequately invest in defense.

- **Advanced evidence tampering** — Criminals can use generative AI to cheaply create fake content that complicates evidence gathering processes.
- **Loss of public confidence and an erosion of social institutions** — Because AI tools have become more effective at facilitating malicious conduct, the public has lost confidence in new technologies and legitimate innovation lags while malicious actors continue to thrive. Loss of public confidence in technology then erodes the confidence in social institutions, democratic systems, and objective truth.
- **Polluted commons** — AI-based attacks such as deepfakes result in a worse digital commons, driving out innocent entities who lack the means to filter through such material. The result is that poorer entities must operate on worse platforms than those with the ability to pay.

COPYRIGHT © 2024 BY THE ASPEN INSTITUTE

This work is licensed under the Creative Commons Attribution Noncommercial 4.0 International License.

To view a copy of this license, visit:
<https://creativecommons.org/licenses/by-nc/4.0/>

Individuals are encouraged to cite this report and its contents.

In doing so, please include the following attribution:

“Envisioning Cyber Futures with AI.” Aspen Digital, a program of the Aspen Institute, Jan. 2024. CC BY-NC. <https://www.aspen-digital.org/report/cyber-futures-with-ai/>

**U.S. and Global
Cybersecurity Groups**

