Policy Brief

Autonomous Cyber Defense

A Roadmap from Lab to Ops

Authors
Andrew Lohn
Anna Knack
Ant Burke
Krystal Jackson

CSET CENTER for SECURITY and EMERGING TECHNOLOGY

CETaS Centre for Emerging Technology and Security

June 2023

Executive Summary

Given the immense economic and societal damage caused by cyberattacks and recent advances in artificial intelligence (AI), interest in the application of AI to enhance cyber defense has grown in recent years. Research is expanding on autonomous cyber defense that can not only detect threats but can engage in defense measures such as hardening or recovery. This report focuses on one promising approach to creating these autonomous cyber defense agents: reinforcement learning (RL).

There is no single agreed definition of autonomous cyber defense, but at its most basic level, these agents would complete some of the tasks of human cyber defenders by protecting networks and systems, detecting malicious activity and reacting to anomalous or malicious behavior, but at the speed of digital attacks.

This report presents a proposed definition for autonomous cyber defense, surveys the current state of autonomous cyber defense and the associated challenges that must be overcome for this technology to become a viable cybersecurity tool. There is no guarantee that autonomous cyber defense will succeed, but the technology is at a stage where policy support is needed to realize the potential benefits and help cyber defenders deal with the speed and uncertainty of modern cybersecurity operations.

RL is the leading AI approach to creating cyber defense agents, which are the core requirement of effective autonomous cyber defense. This technique increased in prominence in 2012 when RL agents first beat expert humans in simple Atari games. Building on that success, from 2015 and 2018, DeepMind built systems for far more challenging games, including Go and Chess, achieving unanticipated levels of success. Researchers flocked to RL, partly because of these successes, but also because of an open framework from OpenAI, which allowed creation of simple, simulated training environments or 'gyms.' The OpenAI gym format simplified research and development, and, in the last few years, cyber gyms have begun to appear that allow the training and creation of cyber defense agents. Even more recently, these gyms became part of an open cybersecurity competition titled Cyber Autonomy Gym for Experimentation (CAGE).

Our study is anchored on the potential for reinforcement learning (RL)-based AI agents to provide the autonomous capabilities required to fulfill some or all of the autonomous cyber defense concept. While the breadth of promising and relevant modeling approaches, techniques and technologies that relate to autonomous cyber defense is

large, our focus on RL is guided by the increased efforts in applied RL for cyber defense and the promising results RL has achieved in other problem domains.

While the technology central to autonomous cyber defense has advanced rapidly in the last decade, many challenges remain before systems can be deployed operationally. During the course of this research project, we interviewed government and non-government experts to identify the requirements for building and fielding trustworthy systems, which include:

- Adaptability - A potential autonomous cyber defense system will need to be future-proofed against changes in the cyber threat environment
- Auditability - Autonomous cyber defense systems must be able to generate logs and archive the agents' decisions and rationale in undertaking actions to enable review and audit, despite the operational tempo potentially exceeding human capacity. Audit logs can also be used to provide assurances that actions taken are lawful and proportionate, and adhere to agreed norms.
- Directability - Human operators will need to be able to redirect or terminate the system if needed.
- Observability - The system needs to provide human operators sufficient data capture and resolution to inform accurate, up to date situational awareness, and provide system performance metrics to support human oversight.
- Security - The autonomous cyber defense system and the agents within them all need to be secured against being leaked, stolen, or compromised.
- Transferability - Autonomous cyber defense agents will need to be deployable in real environments that do not exactly match the environment they were trained in.

To meet these requirements and continue progress, the fledgling field of autonomous cyber defense needs to be nurtured. RL has only recently started to take off for cybersecurity. Academic publications have surged in recent years and gyms for training cyber RL agents have begun to proliferate. However, capabilities remain rudimentary and incomplete compared to the more complex real-world network environments these agents will face. Sustained funding, coordinated effort to bolster simulation, emulation and evaluation tools, securing skilled personnel, and provisioning access to realistic data and infrastructure will help assure progress.

There is substantial potential for growth in autonomous cyber defense if technical challenges can be overcome. The existing agents and environments built for cyber defense currently consider fewer variables and possibilities than the more famous RL agents for playing board games like Go or video games like Atari or DOTA2. This means there is ample potential for increasingly intelligent agents; ones that can manage a larger number of possible defensive actions, and operate in more complex environments that require them to explore more situations. Our exploration of the technical challenges revealed that autonomous cyber defense is going to be a long-term ambition that can only be realized years into the future.

Recommendations

Despite significant progress in the autonomous cyber defense field, our study indicates that no autonomous cyber defense system has been deployed operationally. Given the present maturity of the current technology, we offer recommendations for developing these capabilities to mature the technology (See Chapter 5 for a full list of recommendations).

Invest in scaling up. The field can improve by making bigger and more realistic network simulations that incorporate more complex scenarios and attacker behaviors. Greater fidelity will lead to more capable cyber defense agents. In addition, releasing and maintaining tools such as gyms or trained agents can help attract academia or other researchers to do this work. Finally, sustained funding would also make it easier for researchers to align themselves to these projects.

Build and provide testing and training ranges. Larger and more complex agents will require more computationally intensive training and testing that could strain the resources of some researchers. Setting up and maintaining large computing systems is also a challenge, which requires talent that is hard to come by. Providing the requisite infrastructure, talent and funding resources – perhaps at a subsidized cost, could also help accelerate progress and provide continuity.

Coordinate data sharing. Policymakers across governments and industry have the power to release cyber data about networks that need to be defended and about threats that they are observing. These are all delicate issues that will require careful consideration, but to the extent that sharing data improves cybersecurity, all organizations stand to benefit.

Host competitions. Continue to host autonomous cyber defense competitions, complemented by financial incentives, as a means for improving the gyms and agents while developing future talent.

Prioritize areas that maximize the benefits of autonomous cyber defense. Not all cyber defense situations need autonomous agents, such as where speed is not the limiting factor or where defenses are already effective. Prioritizing areas where autonomy is most impactful can help guide research. Similarly, some technologies, such as vulnerability discovery, could be helpful for both defenders or attackers. Policymakers should invest in research to determine which scenarios and technologies will result in better defenses rather than improved attacks.

Determine whether defender agents require attacker agents. When creating realistic simulations, it is unclear to what extent defensive agents can be built without offensive agents to drive them. Researchers and policymakers should explore methods to limit the capabilities of the offensive agents without sacrificing the effectiveness of defenders and establish tight controls on the proliferation of agent technology and know-how. They should also invest in research to understand which specific scenarios and technologies require offensive agents.

Determine thresholds for authorization of autonomous cyber defense agents. Autonomous cyber defense agents will need to reach high levels of trust in an organization to be given high levels of autonomy. Policy guidance needs to be developed to set initial targets for capability and trustworthiness that are matched to the risk of decisions that the agents are authorized to make. This guidance could be similar to the levels of autonomy developed for autonomous vehicles. They may also vary depending on aspects of the situation or threat environment.

Table of Contents

Introduction

Russian troops crossed the border into Ukraine on the morning of February 24th, 2022. But in the cyber domain, the invasion had already begun.[1]

Just as businesses were closing up the night before the physical invasion, Russia launched a wave of cyberattacks, using a wiper malware that renders computers unusable by deleting key pieces of software necessary for starting up. Government agencies and cyber defense organizations worldwide recognized the threat immediately. They had been investing in AI-powered intrusion detection monitors for years, so were well positioned to discover attacks like these. Within hours, engineers had analyzed the code, provided the signatures to identify it, and even given it a catchy name—HermeticWiper.[2]

Was that response really fast enough for the victims, given that the Russians were already inside the networks at the time the intrusion was detected?[3] The answer is unclear. What we do know is that at least a month into the conflict, HermeticWiper was still disabling computers in Ukraine.[4]

The start of the Ukraine invasion exemplifies the current state and limitations of autonomy for cyber defense. Most efforts to incorporate AI have focused on detecting intrusions and malware so that humans can then choose what defensive actions to take. Ultimately, it is not the number of seconds, hours, or years until discovery that matters, it is whether the defenders can act before the attackers achieve their goals. Some attacks may take months to succeed fully, but others, like in Ukraine, may destroy an organization in the blink of an eye. Threat and intrusion detection is vital, but action must be taken to respond and recover from attacks. Given the limits of humans' speed to respond, is there a way to automate not only detection, but also responses, to better protect against future attacks?

This joint report from Georgetown University's Centre for Security and Emerging Technology (CSET) and The Alan Turing Institute's Centre for Emerging Technology and Security (CETaS) assesses the current state-of-the-art in autonomous cyber defense and its future potential, identifies barriers to progress and recommends specific action that can be taken to overcome those barriers. The findings and discussion will be of relevance to cybersecurity practitioners, policymakers and researchers involved in developing autonomous cyber defense capabilities.

What is autonomous cyber defense?

Autonomous cyber defense is understood here as a desirable future capability that complements existing human-centric approaches to cybersecurity by leveraging key strengths of machine intelligence. It operates at machine speed and scale, and without fatigue.

We found that the term "Autonomous Cyber Defense" means markedly different things to different people. Researchers have largely neglected autonomy for action (acting to defend, respond and protect) as compared to autonomy for detection (spotting, analyzing and characterizing attacks). For autonomy for action, we found that the defense and intelligence sectors are leading both in conceptualization and applied research. Emerging definitions sometimes refer to 'Active Cyber Defense' or 'Intelligent Autonomous Agents for Cyber Defense and Resilience.' For example, NATO RTG IST-152 refers to intelligent autonomous agents that, "will stealthily monitor the networks, detect the enemy cyber activities while remaining concealed, and then destroy or degrade the enemy malware. They will do so mostly autonomously, because human cyber experts will always be scarce on the battlefield. They have to be capable of autonomous learning because enemy malware is constantly evolving. They have to be stealthy because the enemy malware will try to find and destroy them."[5] Some of these elements were similar in all definitions.[6] Where definitions differ is in the scope of tasks assigned to the agent, the boundaries that contain it, the degree of authorization to execute its decisions, and the role of humans in operating and maintaining it.

This report adopts the following working definition of autonomous cyber defense:

Autonomous Cyber Defense describes systems capable of protecting organizations and users through system hardening, network and endpoint management, threat detection, and intrusion response and recovery, without direct human tasking. Autonomous cyber defense systems independently compose and implement safe, proportionate and effective courses of action to accomplish goals based on observation, knowledge and understanding of the world.

Table 1 lays out some descriptions of this vision within the four areas that the definitions differ. Other visions that select different alternatives in these four areas are also valid and worth pursuing but, for this report, we have scoped autonomous cyber defense to the vision outlined in the above definition and in Table 1.

Table 1. Overview of interview responses on potential characteristics of Autonomous Cyber Defense[1]

Characteristic	Description
Defensive countermeasures	• Protect, Respond and Recover actions based on Identify and Detect. • System hardening, endpoint and network management actions. • Employs decoys, canaries and honeynets.
Operational boundary	• Operates within an organization, excludes external offensive operations. • Supports broad applicability, not exclusive to military systems.
Degree of autonomy	• Capable of successful high-stakes decisions and mission completion without being given explicit prior approval or human authorization. • Adapts to previously unseen operational and environmental conditions. • Only operates within authorized bounds.
Human duties	• Humans removed from detailed tasking and task completion. • People design, build, test, operate, and sustain systems. • People act as choreographers, examiners, coaches and auditors.

Our interest is in AI agents that have a broad scope, can take many different actions throughout networks and devices to pre-empt and interrupt attacks, and recover from various adversary actions. The agents can take in a large volume of different types of

[1] Full autonomy is meant here as per the fully autonomous mode of operation specified in the NIST Autonomy Levels for Unmanned Systems Framework. It is worth noting that whilst this concept describes a fully autonomous system, most stakeholders consulted described partial autonomy as a necessary interim step before a fully autonomous system can be realized. This is discussed in more detail in Chapter 4.

data to make decisions. They can act throughout the network that is being defended, but we do not envision agents that take action beyond that network, and certainly not offensive actions beyond the network boundary. We envision agents that have the autonomy to make a variety of potentially high-impact decisions. Reinforcement Learning (RL) is a particularly promising approach for creating these agents and is the focus of our report.

Research aims and methodology

Within this context, the findings contained in this report have sought to assess the current state of autonomous cyber defense, its future potential, and lay out steps that can help bridge the two.

Our study is anchored on the potential for reinforcement learning (RL) based AI agents to provide the autonomous capabilities required to fulfill some or all of the autonomous cyber defense concept. While the breadth of promising and relevant modeling approaches, techniques and technologies that relate to autonomous cyber defense is large, our focus on RL is guided by the increased efforts in applied RL for cyber defense and the promising results RL has achieved in other domains.

The study sought to answer the following research questions:

- RQ1: What is the current state-of-the-art in autonomous cyber defense?
- RQ2: What are potential visions for future autonomous cyber defense?
- RQ3: What are the challenges in achieving those visions?
- RQ4: What actions can be taken to accelerate progress toward autonomous cyber defense?

To address these research questions, the study team collected data over a three-month period, including two parallel literature reviews covering thousands of academic and gray literature (See Appendix A for full methodological approach). One literature review looked at literature on artificial intelligence for cyber security from the past three years. The second literature review used an AI platform using text embeddings to find semantically similar publications from the past 23 years. Next, 23 interviews with UK, U.S. and Australia-based government, academic, defense research organizations,

private sector stakeholders and international legal experts were conducted. In parallel, the study team explored the code of some cyber training environments and agents to assess the range of actions and observations that are currently implemented. It also helped to assess RL's scalability to larger or more detailed environments. The study team then synthesized the findings from the literature reviews, interviews and computational experiments.

The rest of this report presents our major findings from our analysis as follows: Chapter 2 sets out prior work and recent progress. Next, technical challenges to developing and implementing autonomous cyber defense are laid out in Chapter 3. Chapter 4 describes policy challenges in the development and implementation of autonomous cyber defense. Finally, Chapter 5 provides conclusions and recommendations. The appendices contain further information on the methodological approach for this study, as well as further detail on cyber action spaces.

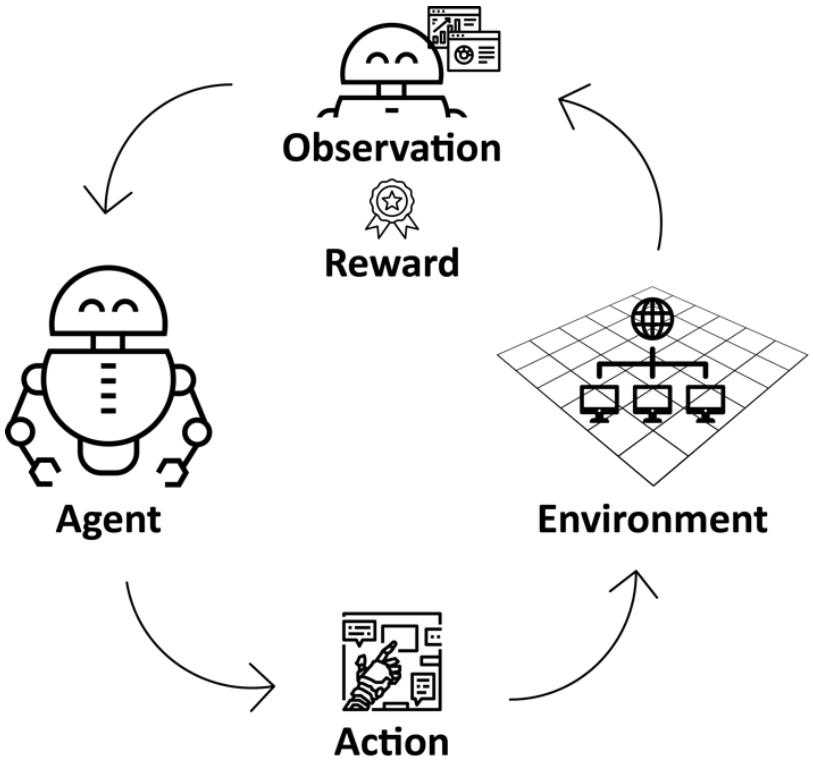
Developments in reinforcement learning

Reinforcement learning's roots, perhaps surprisingly, evolved from models of animal and human behavior. In the early 1900s, psychology research shifted away from analyzing subjectively reported experiences to recording observable behaviors. This new approach, now known as Behaviorism, focused on designing experiments in order to test how specific interventions affected a patient's actions. In their now infamous experiments with rats, B.F. Skinner demonstrated the training power of quickly following particular observed behaviors with reward signals. This approach, referred to as 'operant conditioning,' is the insight that animals will learn to adopt or avoid specific behaviors if, when they exhibit these behaviors, they are followed by a positive or negative reward.[7] Reward in this context generally connotes any positive or negative outcome that follows a behavior. Using these insights, psychologists were able to encourage behaviors they never previously observed or anticipated animals to be capable of, like solving complex mazes and puzzles. These experiments helped psychologists and the general public gain a new understanding of how behaviors are developed and not just simply acquired.

The relatively simple concept of reward driving desired behavior, has, under the right conditions, proven to be an incredibly powerful technique within artificial intelligence development, specifically reinforcement learning.

What is reinforcement learning?

Figure 1. Stages Of Reinforcement Learning



Source: CSET and CETaS.

Where Skinner's rats were rewarded with food and water, machines are rewarded with numbers. A program or function calculates a number representing how well or poorly the machine performed—the reward function. Then another program uses those rewards to adapt an agent's behavior as illustrated in Figure 1. After many trials and errors, machines can use this simple process called Reinforcement Learning (RL) to achieve impressive performance in tasks that were once thought to be the pinnacle of intelligence.

While RL has been one of the central approaches to machine learning for decades, interest spiked in 2012 when RL agents beat expert humans at playing simple Atari games.[8] Building on that success, from 2015 and 2018, AlphaGo and then AlphaZero used RL to master the more intellectually-demanding games Chess and Go.[9] Somewhat less known is that RL then went on to beat human experts in the video game DOTA2.[10] DOTA2 may not be as highly regarded for its intellectual merits, but it is far more

complicated in its interface and in the variety of tasks to perform—traits that are particularly important for cybersecurity.

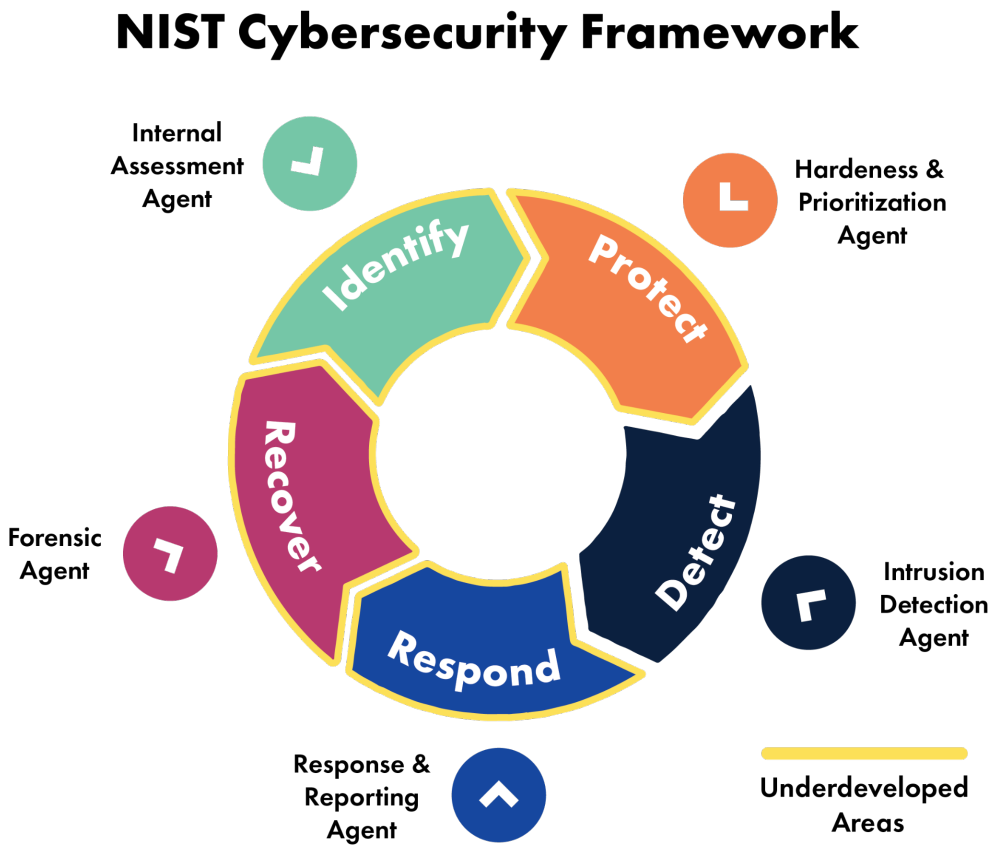
Over this period, researchers flocked to RL, partly because of these famous successes, but also because of the OpenAI gym format[2] that simplified research and development.[11] RL agents interact with their environment, which may be a simulation, to learn. Gyms package those simulations to easily accept inputs from arbitrary RL agents and to provide data and rewards back to those agents. For example, in the cart-pole gym, experimenters train a computerized cart to balance a pole on its end by simply moving either left or right.[12] The gym receives an action (move left or right) from the agent, simulates the result, and provides a reward and the new position and velocity of the cart and pole. The OpenAI format is not necessary, but the simplicity of gyms helped attract new developers and test new algorithms. That is because developing and fine-tuning agents typically requires very different expertise than building environments.

Creating a simulated environment requires knowledge of the application and goals, such as the rules of Go, the mechanics of conducting a robotic surgery, or the status of a digital fight in a computer network. Training the agent, on the other hand, requires expertise in machine learning algorithms and the processes for teaching the agent from its successes and failures. The gyms help to separate those skill sets and make it easy to apply RL to many fields. But interestingly, in this initial period when researchers were flocking to RL, there was conspicuously little interest in RL for cybersecurity.[13] Interest in RL for cybersecurity has been concentrated in detection and coupled with traditional supervised machine learning approaches.[3] Figure 2 illustrates the different tasks for RL in cybersecurity.

[2] In 2021, gyms solidified their status when The Farama Foundation, a non-profit dedicated to maintaining open-source reinforcement learning development tools, took over maintaining the gym (under the new name gymnasium).

[3] Our literature review showed five times more applications in detection than response, the second largest application.

Figure 2. NIST Cybersecurity Framework



Source: CSET and CETaS.

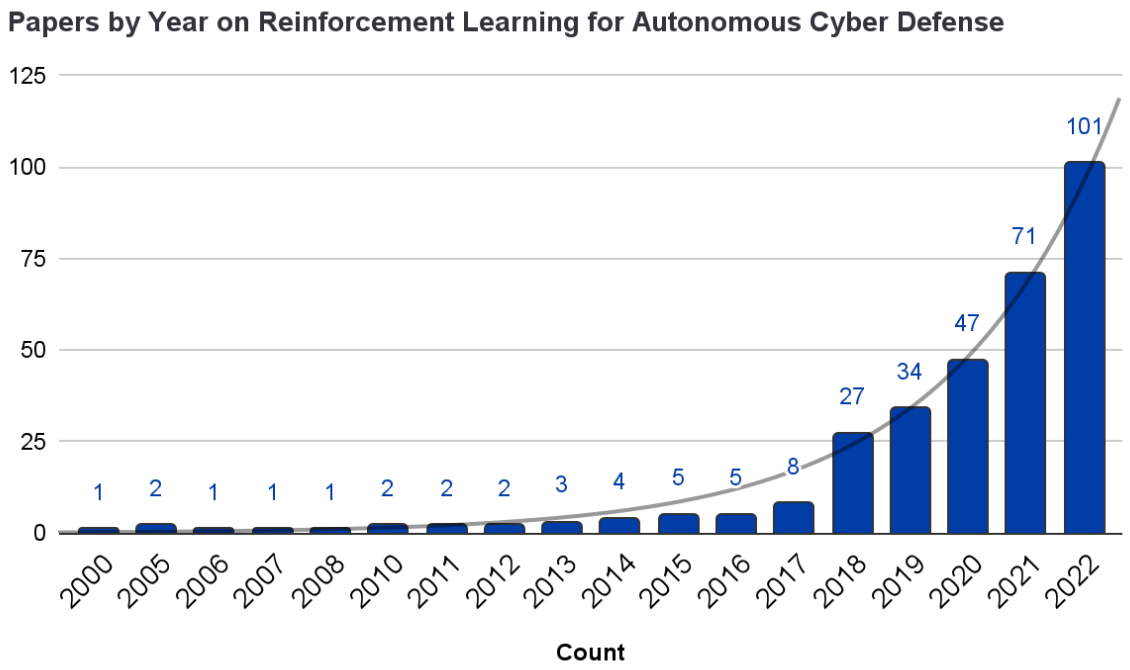
Reinforcement learning for cybersecurity

In 2016, the same year that AlphaGo beat the human world champion, DARPA hosted the Cyber Grand Challenge, where computers battled each other to hack and defend completely autonomously.[14] Despite RL being the zeitgeist of the day, machine learning was almost completely absent from the competition. Competitors relied on techniques that were mostly prescribed by humans and therefore not particularly autonomous. Fast forward to 2020, seven years after RL mastered Atari and five years after the AlphaGo moment, an undergraduate student's bachelor's thesis project was still one of the most advanced RL cybersecurity tools available online.[15]

After a slow start, interest has started to surge. A 2017 NATO Workshop on Intelligent Autonomous Agents for Cyber Defense and Resilience, as well as a 2018 paper in the U.S. National Security Agency (NSA) open-source technology journal helped conceptualize autonomous cyber defense.[16] Academic publication has grown

exponentially since then but is still small compared to either cybersecurity or RL individually, as shown in Figure 3. At the same time, where practical tools for studying cyber RL agents were once hard to come by, they are now more widely available, inviting new entrants to join the field.

Figure 3. Number Of Publications On Autonomous Agents For Cybersecurity From 2000-2022



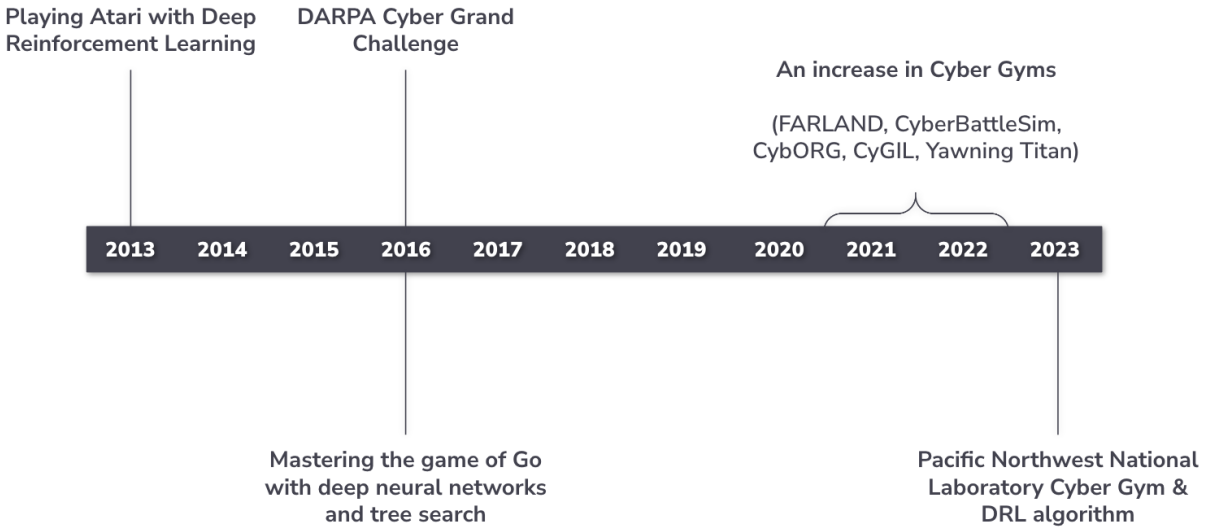
Source: "Limmen / awesome-rl-for-cybersecurity," GitHub, https://github.com/Limmen/awesome-rl-for-cybersecurity.

Research arms of defense and intelligence organizations from the U.S., Australia, UK and Canada have all either provided open-source RL gyms for cybersecurity or published descriptions of their closed-source gyms.[17] Some gyms are intended more for offense than for defense, and they vary in the level of detail of their simulations. Microsoft also publicly released an offensively-minded cyber gym that Apple adapted for a more defensive focus but opted not to release publicly.[18] A subset of these new gyms is listed in Table 2, focusing on ones that are capable of especially detailed simulations. A timeline showing when these gyms were developed is shown in Figure 4.

Table 2. Overview Of Gyms For Cybersecurity From 2021-2023

Gym name	Year	Source	Defense, Offense, or Both	Open Source (y/n)
CybORG	2021[19]	Australia Defence Science and Technology Group	Defensive	Yes
FARLAND	2021	U.S. MITRE and National Security Agency	Both	No
CyberBattleSim	2021	U.S. Microsoft Research	Offensive	Yes
CyGil	2022	Canada Defence Research and Development	Offensive	No
Yawning Titan	2022	UK Defence Science and Technology Lab	Both	Yes
PNNL Gym	2023	U.S. Pacific Northwest National Lab	Both	No

Figure 4. Timeline Of Relevant Developments In RL For Cyber Defense



Source: CSET and CETaS.

Of all these gyms, the Australian CybORG deserves special attention because it is openly available, designed with defense in mind, and has the potential for fairly detailed simulations. It has also been used in a series of competitions that were first announced in August 2021, at the 1st International Workshop on Adaptive Cyber Defense.[20] Since then, they have continued to maintain the tool and have hosted two more competitions called Cyber Autonomy Gym for Experimentation (CAGE) Challenges.

This surge of interest in reinforcement learning for cybersecurity is still small but is a promising start. The next few years may show whether researchers can or cannot overcome the various technical hurdles for creating autonomous cyber defense as outlined in the following section.

Technical challenges

As intellectually demanding as games like Chess and Go are, they are very simple games. Their rules are relatively simple and straightforward to implement in code, making it relatively easy to build environments that simulate play. Everything the computer needs to know to make its decisions—the observations—are simply the location of all the pieces on the board. And, at least for those games, the pieces on the

board contain all the information that is available with no uncertainty about whether the observations are accurate. None of these things are true for cybersecurity.

Complexity and combinatorial explosions

The simple cart-pole gym described earlier only has two possible actions and four observations. The game Go with its 19x19 board, is far more complex. There can be as many as 361 legal moves or actions. In cybersecurity, there is effectively an infinite number of possible actions and observations, and these observations may be partially hidden, or they can even be untrue as part of a deception or an error.

Every configurable setting on every computer, router, and device is a potential action. Moreover, every bit of data flowing in a network or sitting on a computer is potentially important to observe. For example, ten computers that each have ten pieces of software that each have ten possible security settings to configure leads to one thousand possible actions—about three times as many as Go. The number of actions and observations grows exponentially and quickly becomes unmanageable.

Thus, a primary challenge for autonomous cyber defense is selecting tasks and building training environments that are complex enough to be useful, while small enough in terms of the number of actions and observations to be manageable. One idealized vision for autonomous cyber defense is to have one giant model that can perform all the actions that a cyber defender can perform while observing all the data throughout a network, but that would require a seemingly impossibly large number of actions and observations.

An alternative vision is to build many separate agents that are each trained for more constrained tasks, with a smaller number of actions and observations. These agents could work together and pass information amongst each other. For example, one agent may only think of computers as black boxes that can be infected or clean, and it may only be able to perform a few actions to isolate or remediate them. Another agent may be working on those computers, observing all the processes that are running and user behaviors. It could decide whether to kill some of those processes or lock out the users, and it could tell the first agent whether or not the computer is infected.

In principle, gyms can be expanded to match either of these visions, but the current state of the art is very simple. The second CAGE challenge (CAGE2) used a pared-down scenario that essentially treated computers as black boxes that can be in a few different states of clean or infected for a total of only 62 observations. For actions,

defenders could monitor the network, clean the infected computers, or set up decoys.[21] In total, across the thirteen computers, there were 158 possible defensive actions. As outlined in Appendix B, this is a tiny fraction of the total number of actions that real cyber defenders can take, but this is already almost half the number of actions in Go. The competition's winning entry only considered 36 of the possible actions, but RL agents can manage much larger numbers.[22]

The DOTA2 agent can consider up to about 80,000 actions and about 16,000 observations, so there is plenty of room for RL cyber agents to grow. Still, cyber environments can get complex quickly, so autonomous cyber defense will likely need many separate agents that are trained to work together for the many different tasks and roles in defending a network, even if each can consider tens or hundreds of thousands of observations and actions.

Neural network architectures

What makes the game Go difficult is not the number of possible moves at any given time, it is the many possible strategies for each move and their implications as the game proceeds. It is an intellectual challenge, and the agent needs a large neural network to retain the lessons from watching or playing many games. AlphaGo's neural network has 13 layers connecting inputs to outputs via 8.2 million parameters, and DOTA2 used 159 million parameters.[23]

As big as they seem, these models are tiny compared to the biggest AI models, which are now approaching one trillion parameters, but they are far larger than the winning network in the CAGE2 challenge.[24] The agent that won CAGE2 used several neural networks, the largest of which had just 6,372 parameters, and all its networks combined used only 20,938.[4] For comparison, a typical agent for cart-pole uses 450 parameters.[25] The state-of-the-art networks for autonomous cyber defense are currently closer to cart-pole than to AlphaGo, which itself is a long way from the largest models.

But more parameters do not always mean more capable. Cart-pole and the autonomous cyber defense agents use the simplest architecture, a fully connected one,

[4] There is an actor critic with 62 inputs, 64 hidden nodes, 36 outputs, and a softmax with biases for 6,372 parameters. There is also a sequential network with 62 inputs, 64 hidden nodes, and 1 output with biases for 4,097 parameters. And there is one of these networks for each of the two possible attackers for a grand total of 20,938 parameters.

where every node in one layer connects to every node in the next. AlphaGo, on the other hand, takes advantage of an architecture that was designed for computer vision because observing pieces on a Go board is similar to observing pixels in an image. Language models have reached new heights by using a different architecture called transformers. Either inventing new cyber-specific architectures or figuring out how to leverage existing architectures for cyber problems could lead to breakthroughs in autonomous cyber defense.

Neural networks can retain lessons from prior games, but that is not the same as remembering things from the current game. If the offensive agents Microsoft studied found a password in one step of the incursion, they had no way to remember them for the next step.[26] From a defensive perspective, discovering malware signatures or attacker tactics is only helpful if they can be remembered. A capable autonomous cyber defense agent cannot be an amnesiac.

Some architectures do have limited memory built in, but the AlphaGo agent did not use one.[27] It addressed the memory problem by expanding the observations to include the last eight board positions rather than just the current one.[28] This approach is not ideal for autonomous cyber defense agents that are likely to already struggle with the number of observations.

Another alternative is to store this information outside of the neural network. This is the approach taken by the Canadian CyGil gym.[29] Their gym includes observations that are a simple one or zero for whether files or folders have been found on a device, or whether a user's passwords have been discovered. They then have a separate database that the agent can refer to where those files or passwords are stored.

Computational requirements

Running the computers to train just one large language model can cost tens of millions of dollars.[30] Training RL models is not currently as demanding as that, but it can require developing many exploratory models before settling on the final one, which can drive up the computational costs. Further, in the case of RL, there is an additional cost for generating the data through episodic play whereas the data often already exists for other types of AI.

These big AI projects use specially designed computer chips called GPUs, but these are only useful for training and running neural networks. For autonomous cyber defense, two types of compute are common. GPUs, or other special purpose chips, to train and

operate the neural networks, and general-purpose CPUs to run the simulated training environment. The general-purpose CPUs are less efficient. While training an agent, the GPU runs the neural network to choose an action. Then the CPU runs that action through the simulation to calculate the reward. Those action-reward pairs are saved, and every so often, the agent takes a batch of them to update its neural network, which also happens on the GPU.

For a simple game, the simulation on the CPU may run quickly, so the dominant computing costs can be running the neural network on the GPU. However, as the simulations increase in complexity, so do the CPU costs. For DOTA2, the simulation is quite involved, and the CPUs ran for 10 months.[31]

The winning agents from CAGE2 and the CybORG gym were probably not optimized for compute efficiency, but we adapted their code to measure their computing needs. Running the CybORG environment to calculate the rewards took the CPU 3.15 milliseconds on average. Deciding on actions and updating the agent's network took the GPU an average of 0.97 and 232 milliseconds, respectively. Although updating was the slowest step, it only happened once every 20,000 steps, so the dominant costs were in calculating the reward and deciding on actions.

The total costs were about the same for the CPU and GPU, but computing action-reward pairs was about 150 times as expensive as updating the neural network.[5] More advanced autonomous cyber defense agents will certainly need much larger neural networks that will take much more time to update, but they may also need more complicated simulations that take longer to run. So, it seems likely that the cost to generate action-reward pairs will be much more than the cost to update the neural networks.

Defining rewards

When goals are not well-defined, even the most advanced agent will struggle to learn the correct behavior. This requires defining rewards so that the agents can balance future security with current security. The system must be able to properly prioritize a

[5] We used an Intel Xeon 2.3 GHz which cost 3.3 times less than our GPU which was an NVIDIA Tesla V100. Computing rewards took 3.15/0.97=3.25 times as long as deciding on actions, so the costs are roughly balanced.

variety of goals, even ones that can potentially conflict, such as limiting data loss and maximizing up-time.

Goals must be scoped appropriately as well. Defined too narrowly, and the agent will be overly constrained. Depending on the organizational risk tolerance, this may be an advantage or a disadvantage. These agents will struggle to discover the novel strategies and techniques that are part of the promise of RL. However, agents that are given too much free rein can exhibit unintended behaviors. Additionally, there is always the general problem of aligning an agent's behavior or goals with the designer's goals. The alignment problem; where models maximize their reward in ways that do not align with the designer's intent, is a common challenge for RL, but it is particularly challenging for cybersecurity. If the agent's goal is simply to "keep malware off of all systems," the agent may achieve this goal by turning all of the systems off. Without specific and complete goals, an agent can technically achieve its goals but not in the desired ways.

Security concerns: offensive agents

Several cyber gyms are intended primarily for creating offensive agents, and all the major gyms are capable of doing so. It is unclear at this stage if it is even feasible to build defensive agents without also building their offensive counterparts. For the CAGE challenges, the designers manually created a set of simple offensive agents that followed a planned playbook, but more advanced and realistic challenges might require more intelligent offensive agents to challenge the defenders.

It is also unclear who would win in a battle between highly intelligent autonomous attackers and highly intelligent autonomous defenders. It might only be wise to build the pair if there is reason to believe that defenders would prevail. The winner will probably be different for various aspects of cybersecurity and for diverse applications and industries.[32] Researchers should investigate ways to ensure that autonomous agents are more beneficial to defense than to offense for as many applications and circumstances as possible.

Securing the securers

If offensive cyber agents are made, either in isolation or as a necessary precursor for making defensive agents, those offensive agents could cause significant harm if they are leaked or stolen, highlighting the need for them to be deleted or secured. Such measures may be less apparent for defensive agents, but they, too, need to be

protected for strategic and economic reasons. If defensive agents run on the routers, computers, devices, or networks that they are defending, then they will likely be easy for adversaries to steal and reverse engineer.[33] Alternatively, the agents could do most of their decision-making remotely from a separate data center. That would make the agents easier to protect but may create more opportunities for attackers to interfere with communications between the agents and the systems they are protecting.

The degree of autonomous cyber defense access to networks and systems can also influence the size and capability of the agents. Agents that are deployed to the networks they are defending must be small enough to not exhaust the computational resources of those networks and devices. Even modest increases in scale from the current state of the art could necessitate separate GPU-enabled computers just for running the defenses; further increases could necessitate local clusters of GPUs for autonomous cyber defense that would be expensive and challenging to manage.

In addition to protecting the models from being stolen or leaked, skillful attackers can manipulate the inputs to AI systems, including RL agents, so that they make wrong decisions of the attackers' choosing.[34] This is a pervasive trait of advanced AI systems that would need to be addressed throughout design, training, testing, deployment, operation, and maintenance.

Transferability

A final technical challenge this report examines is related to transferability. Future developments of cyber-RL should rely on training models within a simulated environment that mirrors the real environment they will be deployed in as closely as possible. At this early stage, it is common for agents to be highly tailored to their specific scenario, so they struggle against different adversary tactics or if the simulated environment changes slightly.[35] In fact, it is not even clear at this stage how to add or remove devices to an agent's purview or how to add or remove actions it can take, since those are currently fixed by the shape of the neural network before training begins.

These problems are critical to solve because enterprise environments are ever changing, increasingly so with dynamic computing resources like cloud computing. Rudimentary simulations may limit the variety of networks or systems that a trained agent can be used in, and if agents take 10 months to train as they did for DOTA2, then the environments and threats may have evolved too much. This may be less of a problem for low-fidelity simulations that do not overly constrain the agents' set of

experiences. In any case, successfully transferring an agent from the training environment to the real world is a concrete challenge for which specific metrics of performance must be established and evaluated. This is discussed further in Chapter 4.

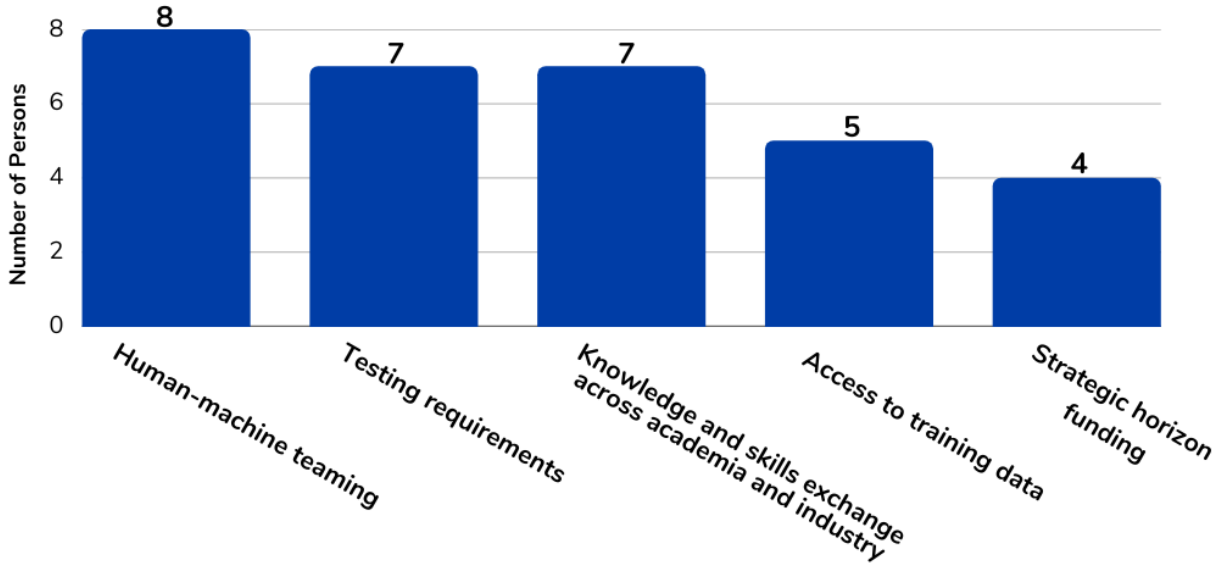
Policy challenges

Building autonomous cyber defense agents is not a purely technological problem. Policymakers have a significant role to play in developing an enabling environment for autonomous cyber defense by setting the necessary regulations and standards, and identifying and providing the resources necessary for successful long-term deployment. The associated policy challenges can be found both in the creation of autonomous cyber defense capabilities and in managing autonomous cyber defense systems once fielded. This chapter contains an exploration of the challenges and the recommendation to address them will be covered in Chapter 5.

In our interviews, we asked 23 experts across government, industry, academia, defense research and development organizations, and international legal communities about the challenges facing autonomous cyber defense,[6] and the most common near-term policy issues, which are shown in Figure 5. It should be noted that participants were able to select multiple policy challenges.

[6] Please see Appendix A for more detail on the methodological approach pursued.

Figure 5. Most Reported Policy Challenges And Enablers For Applied Research In RL For Cyber Defense



Source: CSET and CETaS.

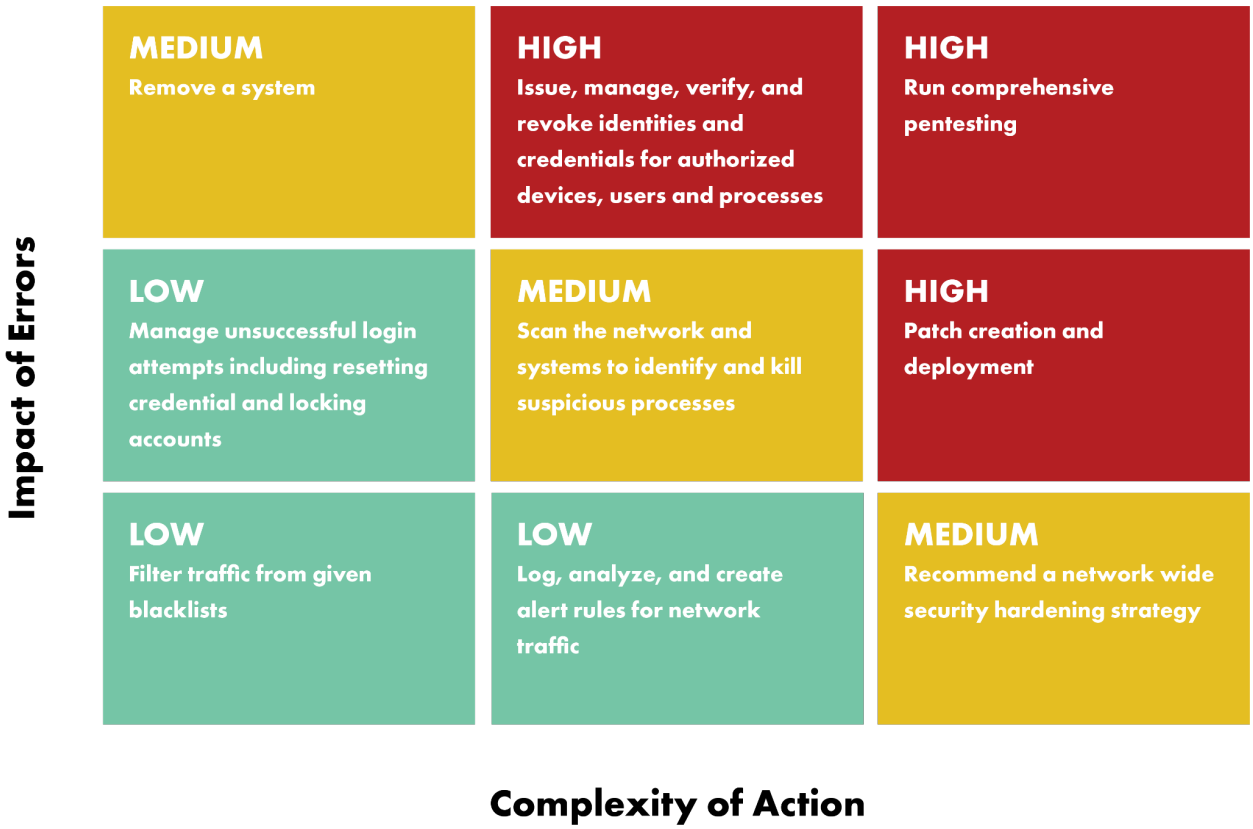
Each of these policy challenges is discussed in turn below.

Human-machine teaming

An autonomous cyber defense agent loses its speed advantage if it has to wait for human approval at every step, so it will need some form of authorization to act. But there are many actions that are potentially risky, such as temporarily stopping services or permanently wiping a user's computer. Determining the right level of autonomy is a challenge, and even human defenders rarely have full autonomy to act without further approval.[36]

To help determine the appropriate level of autonomy in any given autonomous cyber defense context, it may help to establish a human-machine teaming scale with different levels, similar to those developed for autonomous vehicles that specify the expected capability at each level.[37] The levels of autonomy for various actions could depend on the complexity of the actions and the impact of errors, as illustrated in Figure 6. The appropriate levels of autonomy for any system will also depend on the deployment specifics and organizational risk tolerance. Risk assessments developed by standards bodies will be an important part of the safe development and deployment of ACD, and more work needs to be done to refine the risk calculus of implementing ACD.

Figure 6. Potential Levels Of Autonomy On Complexity Of Actions To Be Performed And The Impact Of Errors.



Source: CSET and CETaS.

AI is already involved in many of the cyber security processes described to some extent. Transitioning from AI within cyber tools, to agents that utilize cyber tools will require more robust classifications of autonomy and risk in order for organizations to make decisions around integrating autonomous cyber defense.

Those autonomy and risk levels could then be paired with testing and assurance requirements.[38] Another consideration is whether there are exceptional circumstances where an autonomous cyber defense agent may be given increased autonomy or if there are actions that will never be acceptable for an autonomous cyber defense to take without human authorization. At present, decisions on the level of autonomy are made by companies rather than policymakers, and the opaqueness of the rationale and capability of cyber defense systems due to intellectual property (IP) concerns, raises questions of trustworthiness.[39]

A number of interviewees warned of the cyber operational tempo exceeding human capacity.[40] RL agents may also have the ability to adopt strategies and tactics that humans do not understand. Humans are not entirely outmatched though, even at machine speeds. For example, uploads and downloads of large files can take time, and even automated attackers may choose to operate slowly in order to avoid detection.[41] So some tasks can remain human-centric and be conducted at human speed, and for others, humans can guide and adjust autonomous cyber defense systems throughout an engagement. Setting these roles and responsibilities will require policy guidance at both the strategic level and the most local levels.

Testing

Testing systems is a technical challenge, but policy sets the standards and methods for testing. Tests will require access to data (which is further discussed in Section 4.4.) as well as realistic simulators and emulators that demand substantial infrastructure both in computing hardware for running the tests and software for composing and managing them.[42] Providing a variety of testing environments might also help to reduce the risk of overconfidence that can come from training to testing. For autonomous cyber defense agents deployed in dynamic digital environments that are ever-changing, increased robustness enabled by more diverse and representative training environments will help assure that the systems are as effective as the tests indicate. Perfect test coverage is impossible, but the agents can be made more robust in their ability to respond appropriately in unfamiliar situations.

Besides the infrastructure for testing in realistic simulators and emulators, developers also need policy guidance for acceptance criteria and thresholds. Interviewees suggested an autonomous cyber defense system's testing regime will have to be aligned across a range of stakeholders with potentially competing interests and values, and will broadly need to meet the following evaluation requirements in addition to performance and functionality in order to contribute towards the trustworthiness of systems:

- Adaptability - autonomous cyber defense systems will need to be future-proofed against changes in the cyber threat environment.
- Auditability - autonomous cyber defense systems must be able to generate and archive the agents' decisions, actions, and rationale to enable review and audit, even when the operational tempo potentially exceeds human capacity. Audit

logs can also provide assurances that actions taken are lawful and proportionate, and adhere to agreed norms.

- Directability - operators need to be able to redirect or terminate the system if needed.
- Observability - autonomous cyber defense systems need to provide human operators with sufficient data capture and resolution to inform accurate, up to date situational awareness, and provide rich system performance metrics to support human oversight and audit.
- Security - The autonomous cyber defense systems and the agents within them all need to be secured against being leaked, stolen, or compromised.
- Transferability - autonomous cyber defense systems will need to be deployable in real environments that do not exactly match the environment they were trained in.

Skills gaps, shortages, and the future of work

Demand for AI skills and for cybersecurity skills outstrips the current global talent pool,[43] and the pool of individuals who are skilled in both is even smaller still. For example, the 2022 U.S. State of the Federal Cyber Workforce Report cited 700,000 cyber jobs needed to be filled across the U.S.[44] In the UK, a 2021 Cybersecurity Skills in the UK Labor Market Report found that 680,000 businesses across the UK (50 per cent) have a basic skills gap.[45] Besides these generic cyber workforce skills, there will also be a growing need for more data scientists with an understanding of AI in cyber security. Developing autonomous cyber defense will require extensive cyber expertise to design environments that are realistic. That includes many components and configurations and a means of recording the data that is most likely to be important while ignoring the data that is not likely to be important so that the agents are not overwhelmed by too many observations. It will also require teams of experts who are able to set up large computing infrastructures to run those simulations efficiently at large scales.[46] Finally, developing and deploying autonomous cyber defense will require other types of expertise such as legal, international relations, and testing and assurance. Each of these will need familiarity with the complex issues surrounding autonomous cyber defense once there is more progress against some of the fundamental technical challenges. Although creating autonomous cyber defense systems will require expertise, they also have the potential to help fill the skills gap and

make it easier for organizations to improve their security. As human-machine teaming in cyber defense advances and new roles develop, organizations will need to consider impacts to the workforce.

Data access

Although RL agents learn from exploring their environment rather than just observing data, data is still a core requirement for autonomous cyber defense agents.[47] Designers need data to design realistic environments and threats. Without detailed data about how networks, computers, and devices are set up, managed, and attacked, the training and testing environments will be inadequate. Data showing how humans go about defending a network could also be helpful to get an agent started through a process called Imitation Learning.[48]

Unfortunately, sharing data is a major policy challenge for several reasons. Much of this data is held by private companies which view the data about networks, and attacks on them, as proprietary, and which must respect commercial confidence and protect client privacy.[49] In the national security domain, data on threats, incidents, responses and weaknesses can be highly sensitive and tightly controlled, and all data handling must respect appropriate legal controls on collection, retention and dissemination. Furthermore, data protection regulation can create delays to model development.

Policymakers will need to find new ways to share as much of this data as possible. Some of that could be through crafting regulations or establishing norms. It may also be possible to establish controlled training and testing infrastructure within which agents can operate without providing all the details of that infrastructure and the networks and agents running on it.[50]

Strategic horizon funding

Autonomous cyber defense's recent growth in interest is exciting, but it is still only a seedling to nurture. The teams are small, the projects are limited, and their continuity is uncertain. Some assurances of continued funding over at least a five to ten year period would allow these teams to build a stable workforce, set up the infrastructure to enable applied R&D and set more ambitious goals.[51] This could determine whether autonomous cyber defense can make the jump from compelling academic demonstrations to practical commercial and national tools that are worthy of larger-scale sustained investment. A significant uplift in long-term research funding for

autonomous cyber defense is required to realize the full potential of these new and emerging capabilities.

In addition, offering meaningful monetary incentives to current autonomous cyber defense competitions could incentivize more competitors and spark progress. In 2004, DARPA ran its first Grand Challenge focused on autonomous vehicles and no team was able to complete the 100-kilometer off-road course through the Mojave Desert. In 2005, five vehicles made it through. This competition helped spur the development of autonomous driving technology. In a similar fashion, competitions for autonomous cyber defense could help spur innovation and progress.

Liability and criminal responsibility

Autonomous cyber defense agents will surely fail from time to time. They could fail to defend against novel attacks or fail against attacks that they should be prepared for. They could also cause damage by overreacting to imagined threats. Or the autonomous cyber defense agents themselves could be deceived by a creative attacker in ways that no human would. Any of these failures could lead to complex legal ramifications where it is unclear who is at fault or how to assess the damages.[52]

Similar concerns already exist for present day technology, systems, defender teams, penetration testers, red teams and general cyber security vendors.[53] Accounting for these challenges does not appear to require immediate regulatory intervention, but it does add some complexities that will be difficult to manage.[54] This is before considering potential legal issues for offensive agents or defensive agents that operate beyond the boundaries of the networks they are defending. Additional policy, guidance and legal advice will be required to ensure these risks can be managed appropriately.

Supply chain security and export control

Rather than being stolen or compromised as a whole, autonomous cyber defense agents may be stolen or compromised throughout their development. As such, nations should consider how to protect their supply chains from subversion and theft.[55] This could include simple information sharing about supply chain best practices and threats, or it could require legislated standards for security and export restrictions,[56] a significant departure from present day norms. Policymakers should be cautious to avoid interfering with knowledge or data sharing in ways that could stagnate progress or stifle a fledgling field, while protecting national interests, inhibiting advances by hostile entities and maintaining national security advantages.

Social good and equality of access

In an interconnected digital world where attacks on one system propagate the threat to others, and where disruptions in a gas pipeline, power grid, or shipping company can affect thousands or millions of citizens worldwide, some view cyber defenses as a public good.[57] This view contrasts with corporate incentives to restrict access to cyber defense systems to only those who can pay, or national incentives to withhold strategic technologies.

If autonomous cyber defense is able to provide a meaningful defensive advantage, then policymakers should consider ways to provide that technology widely to individuals, companies, organizations, and nations that would not otherwise have access to it. This could involve open-sourcing tools and resources developed by government entities. It could also involve methods to produce autonomous cyber defense services rather than the products themselves or the tools for developing them. However it is achieved, defense anywhere benefits from cyber defense everywhere, and there is a strong argument for government intervention to ensure widespread access to autonomous cyber defense systems across all areas of the economy.

Conclusions and recommendations

For nearly a decade following RL's achievements in Atari, and five years after the AlphaGo moment, RL for cybersecurity was all but non-existent. Now, RL for autonomous cyber defense has grown into a promising field with a flurry of research results and several encouraging tools.

The autonomous cyber defense field is still small and demonstrations are more academic than practical, leaving plenty of room for growth. The current state of the art in autonomous cyber defense is more similar to relatively simple examples like cart-pole than for the far more complex and impressive RL examples of Go or DOTA2. Making that practical leap, and expanding beyond the defense and intelligence labs, will require nurturing the field. There is no guarantee that autonomous cyber defense will succeed, but it appears to be at a stage where support is needed, and that is promising enough to be worthy of that support.

As autonomous cyber defense agents progress, there will be many policy issues, including possible export controls, legal questions, security of the agents or their underlying technologies and potential societal impacts. Given the early stage of autonomous cyber defense technology, it is too early to provide concrete

recommendations about these future issues. We do offer some recommendations for developing autonomous cyber defense that can help progress the technology in ways that will simplify those future issues when they come. Our recommendations fall in two basic categories: Nurturing the field, and guiding the field.

5.1. Nurture the field

5.1.1. Invest in scaling gyms and agents

RL for cybersecurity can improve by making bigger and more realistic simulations and models that incorporate more observations and actions and more scenarios and attacker behaviors. It will also be important to enable testing to detect unknown or undesirable actions, as well as deception in the observation space. It is not clear how far scaling gyms and agents will progress the field but there is still plenty of opportunity for growth. Releasing and maintaining tools such as gyms or trained agents can help attract academia or other researchers to do this work. Prolonged funding would also make it easier for researchers to align themselves to these projects.

5.1.2. Build and provide testing and training ranges

Larger and more complex agents will require more computationally intensive training and testing that could strain the resources of some researchers. Setting up and maintaining large computing systems is also a challenge that requires talent that can be hard to come by. Providing the requisite infrastructure, talent and funding resources – perhaps at a subsidized cost, could also help accelerate progress and provide continuity.

5.1.3. Coordinate data sharing

Policymakers across governments and industry have the power to release cyber data about networks that need to be defended and about threats that they are observing. Policymakers can also adjust incentives for sharing such as by adjusting liability, privacy, or antitrust considerations. These are all delicate issues that will require careful consideration, but to the extent that sharing data improves cybersecurity, all organizations stand to benefit.

5.1.4. Develop, attract, and retain talent

Talent shortages threaten to constrain the development of this field that relies on expertise from AI, cybersecurity, testing, and IT infrastructure among other domains. Policymakers should make efforts to develop, attract, and retain talent in these areas. These areas are likely to be of continuing importance, and attracting foreign talent not only benefits the receiving country but can slow progress in competing countries.

5.1.5. Host competitions

Continue to host competitions, complemented by incentives such as monetary prizes, as a means for improving the gyms and agents while developing talent. Further, carefully choosing scenarios and rules for the competition also guides the field to develop what technologies that are most aligned with practical goals. Determining those goals, and where exactly to guide the field is a challenge of its own.

5.2. Guide the field

5.2.1. Invest in understanding the risks and benefits of autonomous cyber defense

Not all situations need autonomous agents to the same degree. For example, a defender may be able to slow their network or switch to a completely manual mode during an attack so that humans have time to keep up with attackers. Similarly, some technologies such as vulnerability discovery could be helpful for both defenders or attackers. Policymakers should invest in research to determine which scenarios and technologies will result in better defenses rather than improved attacks, as well as the scenarios where the field may want to focus initially.

5.2.2. Determine whether defender agents require attacker agents

It is unclear whether dynamic, adaptive defensive agents can be built without the offensive agents to drive them. Researchers and policymakers should invest in research to find ways to limit the capabilities of the offensive agents that are used to create effective defenders and establish tight controls on the proliferation of agent technology and knowhow. They should also invest in research to understand which specific scenarios and technologies require offensive agents.

5.2.3. Determine thresholds for authorization of autonomous cyber defense agents

Autonomous cyber defense agents will need to reach high levels of trust to be given high levels of autonomy. Policy guidance is needed to set initial targets for capability and trustworthiness that are matched to the risk of decisions that the agents are authorized to make. This guidance could be similar to the levels of autonomy developed for autonomous vehicles. They may also vary depending on aspects of the situation or threat environment similarly to the DEFCON levels.

5.2.4. Determine priorities for autonomous cyber defense agents

The technical designs and specifications of autonomous cyber defense agents can be different depending on the system being defended or the scenario. Policymakers should determine which systems and scenarios they prioritize so that technical researchers can align their work with strategic goals. For example, large models that run in a remote datacenter may or may not be most valuable. Or models that can be offered as a service to many separate organizations may be more or less valuable than tailored products for each organization. These alternatives present trade-offs where policymaker input could be valuable.

Appendix A: Methodology

The main methods contributing to this report were literature reviews, structured interviews, and various simple computational experiments. Each is described briefly here.

A.1. Research approach

A.1.1. Literature review 1

The CETaS team reviewed academic and gray literature, as well as webpages on artificial intelligence in cyber security from the past three years covering the U.S., UK and Australia. The thematic coverage of the literature review focused on reinforcement learning and deep learning, automated cyber operations and cyber security and human-machine teaming. The literature was then extracted in a structured data extraction matrix aimed at capturing insights on the policy context for autonomous cyber defense, features of RL for cybersecurity, technical requirements, policy challenges, the maturity of the technology and any insights on existing cyber AI challenges and field trials. The output of the literature review was then used to guide the development of a semi-structured interview questionnaire, which focused on filling the gaps in the literature review. This literature review also guided the identification of interview participants.

A.1.2. Literature review 2

The CSET team conducted a search of all literature relevant to autonomous cyber defense resulting in the identification and analysis of thousands of papers, reports, and articles. To find all publications relevant to autonomous cyber defense we implemented a classifier to search CSET's merged corpus of scholarly literature. This corpus brings together over 270 million scientific publications from around the world into one dataset. Specifically, it combines (and deduplicates) publications from Web of Science, Digital Science, Microsoft Academic Graph (MAG), Chinese National Knowledge Infrastructure (CNKI), arXiv, and Papers with Code.

We manually searched, identified, and annotated a very small subset of data from sources such as Google Scholar. Then we used an AI platform for automated data labeling, integrated model training, and analysis, which allowed us to use text embeddings to find semantically similar publications. Next, we created labeling functions based on keywords, regular expressions, our knowledge of the data, and clusters based on text embeddings that extrapolated from our initial annotations,

resulting in weak labels for a much larger training corpus. We used those weak labels to identify additional papers related to autonomous cyber defense and reviewed these publications for relevance. Overall, we were able to increase our final dataset of autonomous cyber defense publications by 15%.

The output of our literature review was used to determine the size of relevant publications by year (Figure 2), understand technical trends and directions, and aided in the comparison of reinforcement learning and cyber defense gyms.

A.1.3. Semi-structured interviews

Research participants were identified using a purposive, non-probabilistic sampling strategy. A focus was on identifying individuals with direct experience with ongoing policy discussions, as well as world-leading industrial and academic research related to autonomous cyber defense. Legal and ethics experts, as well as defense research organizations were also consulted.

A semi-structured interview guide was developed to ensure a broadly consistent line of questioning across interviews, while allowing flexibility to pursue other lines of inquiry identified in the course of discussions. Interviews were conducted on an anonymous, non-attributable basis. Interview data was analyzed following a general inductive approach, whereby the focus is on extracting meaning from data and categorizing data into relevant themes and sub-categories. The sections of this report broadly correspond to the core themes identified through this analysis process.

A.1.4. Computational experiments

We experimented with the CybORG gym and with a few of the agents developed for the CAGE challenges.

Exploring the code of the CybORG challenge allowed us to carefully assess the range of actions and observations that are currently implemented and to start to assess its ability to scale to larger or more detailed scenarios. Inspecting the agents allowed us to understand how their observation and action spaces were constrained and to evaluate the size and structure of their neural networks. We were able to rerun the training and evaluation of those networks to evaluate the performance of the agents and of the training process. We adapted the codes slightly to determine the computational demands of the various components of the training process in order to project costs.

A.2. Caveats and limitations

This research was conducted within a limited timeframe with data collection undertaken between November 2022 - January 2023. The CETaS literature review was limited to the last three years and the geographical scope was limited to the UK, U.S. and Australia. The study team did not conduct a comprehensive search, but instead conducted targeted searches focused on predefined focus topics, given the limited timeframe for the study.

Appendix B: Cyber Action Spaces

Current cyber gyms only cover a small set of actions taken by defenders. To get a sense for how large that fraction is, we compared them to the set of offensive actions listed in MITRE's ATT&CK taxonomy and to the set of defensive actions listed in OpenC2 taxonomy.[58] The actions that are included to some degree in the CybORG, PNNL, Yawning Titan, and Battlesim gyms are highlighted in blue in the taxonomies in Table 3 and Table 4 below. It is important to note that each of these actions can be implemented in many ways, so each of the elements in the taxonomy actually represents many different possible actions. As a result, the actions included in the selected gyms are actually a much smaller fraction of the total possible action space than shown here. Additionally, not all gyms provided detailed information on the specific tactics and techniques used. This highlights however, that currently only a small fraction of the total possible types of actions have been implemented.

Table 3. OpenC2 Actions Framework

ID	Name	Description
1	scan	Systematic examination of some aspect of the entity or its environment.
2	locate	Find an object physically, logically, functionally, or by organization.
3	query	Initiate a request for information.
6	deny	Prevent a certain event or action from completion, such as preventing a flow from reaching a destination or preventing access.
7	contain	Isolate a file, process, or entity so that it cannot modify or access assets or processes.
8	allow	Permit access to or execution of a Target.
9	start	Initiate a process, application, system, or activity.
10	stop	Halt a system or end an activity.
11	restart	Stop then start a system or an activity.
14	cancel	Invalidate a previously issued Action.
15	set	Change a value, configuration, or state of a managed entity.
16	update	Instruct a component to retrieve, install, process, and operate in accordance with a software

		update, reconfiguration, or other update.
18	redirect	Change the flow of traffic to a destination other than its original destination.
19	create	Add a new entity of a known type (e.g., data, files, directories).
20	delete	Remove an entity (e.g., data, files, flows).
22	detonate	Execute and observe the behavior of a Target (e.g., file, hyperlink) in an isolated environment.
23	restore	Return a system to a previously known state.
28	copy	Duplicate an object, file, data flow, or artifact.
30	investigate	Task the recipient to aggregate and report information as it pertains to a security event or incident.
32	remediate	Task the recipient to eliminate a vulnerability or attack point.

* specific tactics and techniques used were not specified in all cyber gyms, like PPNL who says they used 21 proactive actions

Table 4. MITRE's ATT&CK Framework

ID	Name	Description
TA0043	Reconnaissance	The adversary is trying to gather information they can use to plan future operations.
TA0042	Resource Development	The adversary is trying to establish resources they can use to support operations.
TA0001	Initial Access	The adversary is trying to get into your network.
TA0002	Execution	The adversary is trying to run malicious code.
TA0003	Persistence	The adversary is trying to maintain their foothold.
TA0004	Privilege Escalation	The adversary is trying to gain higher-level permissions.
TA0005	Defense Evasion	The adversary is trying to avoid being detected.
TA0006	Credential Access	The adversary is trying to steal account names and passwords.
TA0007	Discovery	The adversary is trying to figure out your environment.
TA0008	Lateral Movement	The adversary is trying to move through your environment.
TA0009	Collection	The adversary is trying to gather data of interest to their goal.
TA0011	Command and Control	The adversary is trying to communicate with compromised systems to control them.

TA0010	Exfiltration	The adversary is trying to steal data.
TA0040	Impact	The adversary is trying to manipulate, interrupt, or destroy your systems and data.
TA0043	Reconnaissance	The adversary is trying to gather information they can use to plan future operations.
TA0042	Resource Development	The adversary is trying to establish resources they can use to support operations.
TA0001	Initial Access	The adversary is trying to get into your network.
TA0002	Execution	The adversary is trying to run malicious code.
TA0003	Persistence	The adversary is trying to maintain their foothold.
TA0004	Privilege Escalation	The adversary is trying to gain higher-level permissions.

Authors

Andrew Lohn is a senior fellow with the CyberAI Project at the Center for Security and Emerging Technology, where Krystal Jackson is a visiting junior fellow. Anna Knack is a Senior Research Associate in the defence and security program of the Centre for Emerging Technology and Security at the Alan Turing Institute, where Ant Burke is also affiliated.

Acknowledgments

We would like to thank John Bansemer and Alexander Babuta for their help coordinating this work, as well as Micah Musser and Chris Rohlf for their feedback on earlier drafts. We would also like to thank all of the anonymous interviewees. Finally, we would like to thank our external reviewers Paul Yu, Melody Wolk, Andy Applebaum, Tim Watson, Andrew Dwyer, and Chris Hicks for their productive comments and critiques.

Document Identifier: doi: 10.51593/2022CA007

Endnotes

[1] "Russia behind cyber-attack with Europe-wide impact an hour before Ukraine invasion," Foreign, Commonwealth & Development Office and The Rt Hon Elizabeth Truss MP, Gov.uk, last modified May 22, 2022, https://www.gov.uk/government/news/russia-behind-cyber-attack-with-europe-wide-impact-an-hour-before-ukraine-invasion.

[2] Joe Tidy, "Ukraine crisis: 'Wiper' discovered in latest cyber-attacks," *BBC News*, February 24, 2022, https://www.bbc.co.uk/news/technology-60500618.

[3] "HermeticWiper," Juniper Networks, last modified April 19, 2022, https://blogs.juniper.net/en-us/threat-labs-knowledge-base/hermetic-wiper.

[4] Microsoft Digital Security Unit, "An overview of Russia's cyberattack activity in Ukraine," April 27, 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4Vwwd.

[5] Alexander Kott, Paul Théron, Martin Drašar, Edlira Dushku, Benoît LeBlanc, Paul Losiewicz, Alessandro Guarino, Luigi V Mancini, Agostino Panico, Mauno Pihelgas, and Krzysztof Rzadca, "Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture Release 2.0 https://arxiv.org/ftp/arxiv/papers/1803/1803.10664.pdf, " *ArXiv* (2019).

[6] Interview with UK government expert, 05 December 2022; Interview with non-UK government expert, 13 December 2022.

[7] R. Staddon, J. E., and D. T. Cerutti. "Operant Conditioning." Annual review of psychology 54, (2002): 115. Accessed April 11, 2023. https://doi.org/10.1146/annurev.psych.54.101601.145124.

[8] Volodymyr Mnih, Koray Kuvukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra and Martin Riedmiller, "Playing Atari with Deep Reinforcement Learning," *ArXiv* (2013).

[9] "AlphaZero: Shedding new light on chess, shogi and Go," last modified December 6, 2018, https://arxiv.org/abs/1912.06680.

[10] "AlphaZero: Shedding new light on chess, shogi and Go," last modified December 6, 2018, https://arxiv.org/abs/1912.06680 ; "OpenAI Five defeats Dota 2 world champions," OpenAI, last modified April 15, 2019, https://openai.com/research/openai-five-defeats-dota-2-world-champions.

[11] "Farama-Foundation / Gymnasium,"GitHub, https://github.com/Farama-Foundation/Gymnasium.

[12] "Openai / gym" Github, https://github.com/openai/gym/blob/master/gym/envs/classic_control/cartpole.py.

[13] Micah Musser and Ashton Garriott, "Machine learning and cybersecurity: Hype and reality," *CSET Georgetown*, (2021).

[14] "Cyber Grand Challenge (CGC)," Defense Advanced Research Projects Agency, accessed 1 March 2023, https://www.darpa.mil/program/cyber-grand-challenge.

[15] Jonathon Schwartz, "Autonomous penetration testing using reinforcement learning," ArXiv (2018); "Jjschwartz / NetworkAttackSimulator," GitHub, https://github.com/Jjschwartz/NetworkAttackSimulator.

[16] Alexander Kott, Ryan Thomas, Martin Dražar, Markus Kont, Alex Poylisher, Benjamin Blakely, Paul Theron, Nathaniel Evans, Nandi Leslie, Rajdeep Singh, Maria Rigaki, S Jay Yang, Benoit LeBlanc, Paul Losiewicz, Sylvain Hourlier, Misty Blowers, Hugh Harney, Gregory Wehner, Alessandro Guarino, Jana Komárková, and James Rowell, "Toward intelligent autonomous agents for cyber defense: Report of the 2017 workshop by the North Atlantic Treaty Organization (NATO) Research Group IST-152-RTG." *ArXiv* (2018); Joe McCloskey and David J. Mountain, *The Next Wave: The National Security Agency's review of emerging technologies* (California: National Security Agency, 2018).

[17] Callum Baillie, Maxwell Standen, Jonathon Schwartz, Michael Docking, David Bowman and Junae Kim, "CybORG: An Autonomous Cyber Operations Research Gym," *ArXiv* (2020); "Cage-challenge / cage-challenge-1," GitHub, https://github.com/cage-challenge/cage-challenge-1/commit/f1fb397cce49a2a80ee2fcf060bd3ac713bd2b; Andres Molina-Markham, Cory Minter, Becky Powell and Ahmad Ridley, "Network environment design for autonomous cyberdefense," *ArXiv* (2021); Alex Andrew, Sam Spillard, Joshua Collyer and Neil Dhir, "Developing optimal causal cyber-defence agents via cyber security simulation," *ArXiv* (2022); "dstl / YAWNING-TITAN" GitHub, https://github.com/dstl/YAWNING-TITAN ; Ashutosh Dutta, Samrat Chatterjee, Arnab Bhattacharya, Mahantesh Halappanavar, "Deep reinforcement learning for cyber system defense under dynamic adversarial uncertainties,"*ArXiv* (2023).

[18] "Gamifying machine learning for stronger security and AI models," Microsoft Security, last modified April 8, 2021, https://www.microsoft.com/en-us/security/blog/2021/04/08/gamifying-machine-learning-for-stronger-security-and-ai-models/; Andy Applebaum, Camron Dennler, Patrick Dwyer, Marina Moskowitz, Harold Nguyen, Nicole Nichols, Nicole Park, Paul Rachwalski, Frank Rau, Adrian Webster and Melody Wolk, "Bridging automated to autonomous cyber defense: Foundational analysis of tabular Q-learning," *AISec'22: Proceedings of the 15th ACM Workshop on Artificial Intelligence and Security*: 149-159.

[19] This uses CybORG's release date (Aug 20, 2021) rather than its first publication date (Feb 26, 2020).

[20] "W24: 1st International Workshop on Adaptive Cyber Defense (ACD 2021)", *Workshops, IJCAI2021 Montreal,* Accessed 1 March 2023, https://ijcai-21.org/workshops/

[21] "Cage Challenge 2 TTCP Cage Challenge 2," Github (August 2022), Accessed 1 March 2023, https://github.com/cage-challenge/cage-challenge-2.

[22] "Cyborg Cage 2 Attempt 1" Github (July 2022), Accessed 1 March 2023, https://github.com/john-cardiff/-cyborg-cage-2/.

[23] Jsevillamol and Pablo Villalobos, "Parameter counts in Machine Learning," *Less Wrong* (June 2021), Accessed 22 February 2023, https://www.lesswrong.com/posts/GzoWcYibWYwJva8aL/parameter-counts-in-machine-learning

[24] Andrew Lohn and Micah Musser, "AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?," *CSET Georgetown* (January 2022).

[25] Rita Kurban, "Deep Q Learning for the CartPole," *Towards Data Science* (December 2019), Accessed 1 March 2023, https://towardsdatascience.com/deep-q-learning-for-the-cartpole-44d761085c2f

[26] Microsoft Defender Research Team, "CyberBattleSim", *GitHub*, https://github.com/microsoft/CyberBattleSim ; Interview with private sector expert, 01 February 2023.

[27] Recurrent Neural Networks for example

[28] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel and Demis Hassabis, "Mastering the game of Go without human knowledge", *Nature* 550 (October 2017): 354-350; This describes the architecture for AlphaGo Zero rather than AlphaGo. AlphaGo Zero used 19x19x17=6,137 observations whereas AlphaGo used 19x19x48=17,328.

[29] Li Li, Raed Fayad, Adrian Taylor, "CyGIL: A Cyber Gym for Training Autonomous Agents over Emulated Network Systems", Cryptography and Security (2021).

[30] Private discussion with companies training large language models. This range of expense is consistent with the analysis in Andrew Lohn and Micah Musser, "AI and Compute: How Much Longer Can Computing Power Drive Artificial Intelligence Progress?,", *CSET Georgetown* (January 2022); Interview with private sector expert, 01 February 2023; Interview with private sector expert, 01 December 2022

[31] Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemyslaw Debiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, Rafal Jozefowicz, Scott Gray, Catherine Olsson, Jakub Pachocki, Michael Petrov, Henrique P.d.O. Pinto, Jonathan Raiman, Tim Salimans, Jeremy Schlatter, Jonas Schneider, Szymon Sidor, Ilya Sutskever, Jie Tang, Filip Wolski, Susan Zhang, "Dota 2 with Large Scale Deep Reinforcement Learning," *Arxiv* (2019).

[32] Andrew Lohn and Krystal Jackson, "Will AI Make Cyber Swords or Shields?,", *CSET Georgetown* (August 2022).

[33] These agents would need to be small so as to not overwhelm the devices they are protecting. Smaller usually means less capable, but it also usually means faster and cheaper to run.

[34] Andrew Lohn, "Hacking AI: A Primer for Policymakers on Machine Learning Cybersecurity," *CSET Georgetown* (December 2020); Adam Gleave et al, "Adversarial Policies: Attacking Deep Reinforcement Learning," *Arxiv* (2020).

[35] Melody Wolk, Andy Applebaum, Camron Dennler, Patrick Dwyer, Marina Moskowitz, Harold Nguyen, Nicole Nichols, Nicole Park, Paul Rachwalski, Frank Rau, and Adrian Webster, "Beyond CAGE: Investigating Generalization of Learned Autonomous Network Defense Policies," NeurIPS (2022).

[36] Alexandre K. Ligo, Alexander Kott and Igor Linkov, "Autonomous Cyber Defense Introduced Risk: Can We Manage the Risk?" *Arxiv* (2022).

[37] Interview with non-UK government expert, 13 December 2022; Interview with academic expert, 01 December 2022; Interview with UK government expert, 05 December 2022; Interview with non-UK government expert, 12 December 2022; Interview with non-UK government expert, 11 January 2023; Interview with non-UK government expert, 18 January 2023; Interview with private sector expert, 12 January 2023; Interview with private sector expert,

[38] Interview with non-UK government expert, 13 December 2022.

[39] Interview with UK government expert, 08 December 2022.

[40] Interview with non-UK government expert, 13 December 2022; Interview with private sector expert, 12 January 2022; Interview with private sector expert, 01 February 2023; Interview with academic expert, 01 December 2022; Interview with UK government expert, 05 December 2022; Interview with academic expert, 05 December 2022.

[41] Interview with academic expert, 05 December 2022; Interview with non-UK government expert, 12 December 2022.

[42] Interview with non-UK government expert, 13 December 2022; Interview with private sector expert, 12 January 2023; Interview with academic expert, 18 January 2023; Interview with private sector expert, 01 February 2023; Interview with private sector expert, 01 December 2022; Interview with non-UK government expert, 13 December 2022; Interview with government expert, 05 December 2022.

[43] Diana Gehlhaus and Ilya Rahkovsky, "U.S. AI Workforce Labor Market Dynamics CSET Issue Brief," CSET Georgetown (2021); Interview with academic expert, 05 December 2022; Interview with non-UK government expert, 08 December 2022; Interview with private sector expert, 12 January 2023; Interview with private sector expert, 01 December 2022; Interview with academic expert, 06 December 2022; Interview with academic expert, 07 December 2022; Interview with non-UK government expert, 12 December 2022.

[44] Federal Cyber Workforce Management and Coordinating Working Group, *State of the Federal Cyber Workforce: A Call for Collective Action* (2022).

[45] Department for Science, Innovation and Technology, Department for Digital, Culture, Media and Sport, "Cyber security skills in the UK labour market 2021," March 23, 2021.

[46] Interview with private sector expert, 12 January 2023; Interview with private sector expert, 01 February 2023; Interview with private sector expert, 01 December 2022; Interview with academic expert, 07 December 2022.

[47] Interview with private sector expert, 11 January 2023; Interview with non-UK government expert, 08 December 2022; Interview with government expert, 08 December 2022; Interview with non-UK government expert, 12 December 2022; Interview with academic exert, 18 January 2023.

[48] Ahmed Hussein et al, "Imitation Learning: A Survey of Learning Methods," *ACM Computing Surveys,* 50, issue 2, (April 2017): 1-35

[49] Interview with private sector expert, 01 December 2022; Interview with academic expert, 18 January 2023.

[50] This would require research and some risk because agents may be able to extract details from the private infrastructure, but it is a topic worth exploring if it can promote data sharing.

[51] Interview with academic expert, 05 December 2022; Interview with non-UK government expert, 08 December 2022; Interview with 12 December 2022; Interview with 13 December 2022.

[52] Interview with legal expert, 07 February 2023; Interview with academic expert, 06 December 2022; Interview with academic expert, 13 December 2022.

[53] Interview with private sector expert, 01 December 2022; Interview with non-UK government expert, 12 December 2022; Interview with private sector expert, 12 January 2023; Interview with legal expert, 07 February 2023.

[54] Rain Liivoja and Ann Väljataga, eds., *Autonomous Cyber Capabilities under International Law*, (NATO CCDCOE Publications, 2021)

[55] Interview with academic expert, 05 December 2022; Interview with non-UK government expert, 08 December 2022; Interview with non-UK government expert, 08 December 2022; Interview with government expert, 08 December 2022; Interview with 12 December 2022; Interview with academic expert, 13 December 2022; Interview with private sector expert, 12 January 2023.

[56] "Working groups", NDISAC, accessed March 1, 2023, https://ndisac.org/ndisac-working-groups/

[57] Mariarosaria Taddeo, "Is Cybersecurity a Public Good?," *Minds and Machines* 29, 349-354, (2019).

[58] "Homepage," MITRE ATT&CK, accessed March 1, 2023, https://attack.mitre.org/; Joe Brule and Duncan Sparrell, OASIS, "Open Command and Control (OpenC2) Language Specification Version 1.0," 24 November 2019, accessed March 1, 2023, http://docs.oasis-open.org/openc2/oc2ls/v1.0/oc2ls-v1.0.html