# Legal Challenges of Next Generation AI

Peter R. Stephenson, PhD

*Keywords— cyberspace, cyber law, jurisdiction, international law, private law, cyber definitions, autonomous malware, Internet*

*Abstract*

*There are several areas of artificial intelligence (AI) that impact digital forensics and the law. The impacts are due to some recurring issues.*

*First, the legal community does not, overall, understand the foundations of how the Internet operates and, more specifically, how artificial intelligence works.*

*Second, the technical community is, on the whole, ignorant of how the law works in relation to events in cyberspace, more specifically how to determine jurisdiction.*

*Finally, the issue of jurisdiction in cyberspace is among the prickliest of all cyber-legal challenges. This issue is exacerbated when AI is involved.*

*This paper explores, briefly, the issues and one approach to addressing them. It is based upon some full-length peer-reviewed papers presented covering important aspects of the topics as well as unpublished scholarly research. The language is intended to be plain English, avoiding jargon, accessible to both the legal and forensic communities. The audience is, largely, the legal community but technical managers and entry-level technical individuals also may find it useful.*

## I. Introduction

We have entered an era where machines, albeit ostensibly under our control, are able to take control of many of life's daily functions. We are seeing the dawn of autonomous vehicles, widespread use of what marketers call, simply, AI, intending that their potential customers don't question the characterization too closely. They assume marketing planned and carried out by machine learning.

The other side of that coin is that the same technologies – presumably focused by their creators on simplifying our daily lives and performing computational tasks in seconds or minutes that humans take long times to perform – enable cyber criminals and cyber fraudsters.

One of the most recognizable of these nefarious tasks is cyber fraud carried out using "deepfakes", machine learning-generated images, videos and voices that either substitute for real people or even create bogus people who can act without any form of control once they have their mission encoded in their software.

While this certainly generates quite a collection of technical challenges – how do you locate the creator of a deepfake, for example – the legal challenges may even be greater.

The biggest of these challenges is a two-fold issue: first, how does one determine the proper venue to claim personal jurisdiction over a fraud perpetrated by a hidden identity using a deepfake image, and, second, how does the legal community attribute such a fraud when even the computer scientists are challenged?

## II. Bot nets, hive bots and swarms

Before we get into the depths of the problem, we need to define a few technical concepts for the legal community that will carry us forward. Terms associated with artificial intelligence (AI) can be confusing and other information security terms even more so.

For our purposes we will concentrate on some key AI-related terms and some key infosecurity-related terms. This is not the entire universe of relevant terminology but it will serve well for our purposes. The most important purpose of this briefing paper is to associate – in plain English rather than jargon – technical terms with legal constructs. The author once was accused by a law professor of "not sounding like a lawyer". That, of course, is the point. Technologists don't speak legalese and lawyers do not think in technical terms. There needs to be a common meeting point and that achievement is the purpose for this paper.

We begin with some core AI terms.

First, we need to understand that AI is a collection of technologies and techniques that have as their objective managing systems that do not require human intervention to function[1]. AI encompasses, for our purpose, machine learning, deep machine learning and neural networks. We focus upon machine learning.

Machine learning (ML) takes a body of data, uses it to learn about the data and its subject matter and then trains itself to apply those data to real interactions[2] [3]. The core issues are no human intervention and learning about data so that similar data may be addressed in whatever way the ML programmer wishes.

Machine learning may be supervised or unsupervised. If it is supervised, the ML program is provided with a training dataset and from that learns what is expected of it and how to interpret inputs. In that regard it is much like a student in a math class who, provided with the method for analysis of a certain type of math problem along with a set of representative data – multiplication tables, for example - that describes the task and expected outcome, does an assignment of problem solving using those resources.

If the ML is unsupervised, it gets no pre-determined training set and must derive its own from its observations. This is far more like the way most humans learn. Also, unsupervised training is far less prone to compromise.

ML can be compromised by stealing the training set – supervised learning – and introducing small perturbations in the data such that a disallowed operation looks as if it is permitted. This is far harder to do with unsupervised learning

Certain types of machine learning – e.g., neural networks – may be compromised using a black box attack where the attacker knows nothing about the training dataset. In this type of attack, the adversary "queries the oracle". In other words, it sends input to the ML system and records the response.

Over time the adversary may be able to deduce the training data and introduce small perturbations through repetitive bombardment with slightly altered data. Once those perturbations are established in the ML system's "mind", the system accepts them as legitimate and the attacker can avail him or herself of the corrupt training set. This is not always successful, and it certainly is far less successful – some say impossible - with unsupervised learning training.

Now, on to describing various forms of applicable malware.

"Bot" is another way of saying "robot". This implies an autonomous operation but it usually, today anyway, is not. Bots may be collections ("bot nets"), often large, of malwares that infect victim computers and then take their instructions from a command and control (C2) server. Those instructions may be exfiltration of victim's data, such as payment card information, or using the victim as a base of operations – a "zombie" – for attacks against other targets. When the bot takes its instructions from a C2 it is not autonomous. While there are many types of bots today, they are, largely, controlled by C2 servers.

The next generation of bots will be autonomous meaning that they will be programmed with a mission, given the core knowledge needed to accomplish the mission,

[1] See https://www.britannica.com/technology/artificial-intelligence : "Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings."

[2] See WestLaw, https://intl.westlaw.com/Document/I2eb1463e3a3411e89bf099c0ee06c731/View/FullText.html?navigationPath=Search%2Fv1%2Fresults%2Fnavigation%2Fi0ad6ad3c0000016ff81ce90135792d90%3FNav%3DANALYTICAL%26fragmentIdentifier%3DI2eb1463e3a3411e89bf099c0ee06c731%26parentRank%3D0%26startIndex%3D1%26contextData%3D%2528sc.Search%2529%26transitionType%3DSearchItem&listSource=Search&listPageSource=34ec7096432424be2ab03998e0988765&list=ALL&rank=1&sessionScopeId=a1c0eef0e252bcef324d9ff1c1c0bf9826383d382fa075193e7342a4c51053b4&originationContext=Search%20Result&transitionType=SearchItem&contextData=(sc.Search)&navId=BB4ABA8E2DD9B7AE719FADBFF924CE4C: "Machine learning is the computers' ability to learn without being explicitly programmed to do so"

[3] Expert System, *What is machine learning? A definition*, Expert System(2017), *available at* https://expertsystem.com/machine-learning-definition/.

and left on their own[4]. A collection of these autonomous bots is called a "hive". The hive has collective intelligence, much like the Borg in the old Star Trek TV series. The individual bots share data and information with other members of the hive. The result of this collective sharing is that the hive gains much more intelligence than could a single bot. It also means that there is no need for a C2 server. This is similar to many types of insects, particularly ants. When we start developing "hives of hives" we have what is called a "swarm". Theses swarms, and their swarm bots, are usually quite large and can launch attacks on their own initiative without C2 intervention.

They usually use some form of machine learning and, when that machine learning is used to disable a legitimate system also using machine learning, we call that "adversarial machine learning"[5].

While there are significant technical challenges in attribution of a hive or swarm, the legal challenges, arguably, are greater. At the top of these legal challenges is determining jurisdiction over an unknown actor in an unknown location.

# III. Addressing Legal Challenges

## 1. Jurisdiction

In his draft PhD dissertation[6], Stephenson lays out a three-pronged approach for determining jurisdiction in cyberspace:

1. *The Cyber Event Test, and…*
2. *The Modifier Test, and…*
3. *The Cyber Effects Test*

*In order to be classified as a Cyber Event, the case must adhere to both the definition of cyberspace[7] and the definition of cyber science[8] (the Cyber Event Test).*

*In order to be subject to jurisdiction in cyberspace, the case must demonstrate purposeful availment within the context of cyberspace (the Modifier Test).*

*In order to be subject to jurisdiction in cyberspace the case must be able to apply the standard effects test within the context of cyberspace (the Cyber Effects Test).*

Of the three prongs, the Modifier Test is the most potentially troublesome because demonstrating the bot herder's purposeful availment of the victim's jurisdiction is arguable based upon a lack of specific target knowledge by the bot herder.

Stephenson's research also confirms the position of the Tallinn Manual[9]: cyberspace is not a separate and unique domain. For every cyber event[10] there are at least two unique sides. From a legal perspective those may be thought of as as a plaintiff and a defendant.

While the three-prong test is useful in virtually every case using current adversarial technology, or in every instance where the event results in a dispute, its use may be akin the shooting a mosquito with a cannon. Where the

[4] Derek Manky, *The Evolving Threat Landscape - Swarmbots, Hivenets, Automation in Malware*, CSO Online(2018), *available at* https://www.csoonline.com/article/3301148/the-evolving-threat-landscape-swarmbots-hivenets-automation-in-malware.html.

[5] Nicholas Carlini, *A Complete list of all (arXiv) Adversarlial Example Papers*, Nicholas Carlini(2019), *available at* https://nicholas.carlini.com/writing/2019/all-adversarial-example-papers.html.

[6] Peter Stephenson, A Framework for Determining U.S. Jurisdiction in Cyberspace (2019) University of Leicester, School of Law).

[7] Cyberspace is a complex global information infrastructure that facilitates communication between technology such as computers, networks and other digital systems, both independently and on behalf of people using it. Cyberspace per se is distinct from physical space and the constraints imposed by it such as geographic boundaries (Stephenson).

[8] Cyber science is the study of phenomena caused or generated within the cyber space, which may or may not interact with phenomena caused or generated within the physical space (Stephenson).

[9] Michael N. Schmitt, Tallinn manual 2.0 on the international law applicable to cyber operations (Prepared by the International Groups of Experts at the Invitation of the NATO Cooperative Cyber Defence Centre of Excellence 2 ed. 2017).

[10] While the legal community prefers to characterize the "events" as "disputes", we prefer the more generic term. An event in cyberspace need not be a dispute, although a dispute certainly may arise at some point in an event.

three-prong test becomes truly useful is in identifying jurisdiction in what we choose to call “next generation events”.

These events involve events where there is a plaintiff in one jurisdiction, a defendant in another and the data path passes through one or more intermediate jurisdictions, at least one of which modifies the data in some way.

Another type of next generation event is one where identities and locations are masked using several technologies. In the next section w set up a hypothetical attack where identities and locations are masked using fake identification and encrypted connections.

## IV. Hypothetical

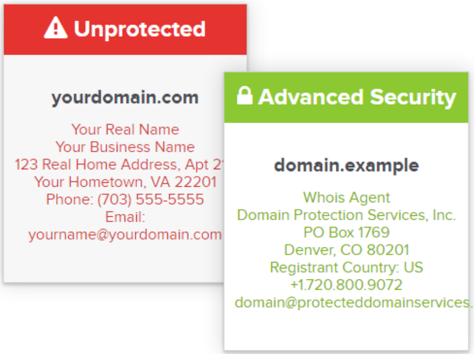
### 1. Setting up the hypothetical: a fake identity

For our hypothetical, we created an alias named Diana Schmidt with the following characteristics[11]:

- Nationality: German
- Address: Alter Wall XX, Lohr, Germany (with geo coordinates)
- Mother’s maiden name: Pfeiffer
- Phone: (actual, though not active, with country code)
- Birthday: 31 Aug 1974
- Age: 45
- Email address: DianaSchmidt@jourXXXXXX.[com]
- Username: (given but redacted here)
- Password: (given but redacted here – both username and password work)
- Mastercard: (real pattern but not usable to make purchases)
- Full employment details
- Height, Weight and Blood Type
- UPS, Western Union and MoneyGram tracking numbers
- Vehicle details

This tool is free to use, readily available on the Internet and users may select from dozens of nationalities and countries. Once an Internet alias is created, a “handle” or nickname comes next. Typically, the actor will not use the alias’ email but will establish an anonymous, encrypted email account using a free service such as ProtonMail[12]. For that account, the actor will use his or her handle. Using that account, the actor will register the IP address. That means that even if the actor is traced back through his or her handle to the alias – in this case, Diana Schmidt – the trail will end there. Service of subpoenas to the alias is likely to fail.

Additionally, these actors often use a service to mask their registration details. This is called “whois protection”. Whois is the open registration database for IP addresses on the Internet. By using whois privacy the only thing an investigator sees upon conducting a domain search is the information for the privacy provider. Figure 1 shows an example[13].



⚠ Unprotected

yourdomain.com

Your Real Name
Your Business Name
123 Real Home Address, Apt 2
Your Hometown, VA 22201
Phone: (703) 555-5555
Email:
yourname@yourdomain.com

🔒 Advanced Security

domain.example

Whois Agent
Domain Protection Services, Inc.
PO Box 1769
Denver, CO 80201
Registrant Country: US
+1.720.800.9072
domain@protecteddomainservices.

*Figure 1 - Whois privacy as provided by name.com*

Once the fake name has been created, all transactions in the attack will trace no further than the fake identity. Additionally, for realism – and misdirection of investigators – a deepfake image (or images) of the fake person may be created. These deepfake images – and, even videos – may be used to set up social network accounts on Facebook, Twitter, WhatsApp and Instagram. The fake accounts lead the investigator to rabbit holes that waste investigator time and end up nowhere except the fake ID.

[11] Fake Name Generator, *Your Randomly Generated Identity*. Fake Name generator online tool (2019), *available at* https://www.fakenamegenerator.com/gen-random-gr-gr.php. The tool is made by Coraban Works which appears to be in San Antonio, Texas, an apparently one-person shop owned by Jacob Allred which may or may not be the owner’s real name or loation.

[12] https://mail.protonmail.com/login

[13] name.com, *Get Advanced Security for Your Domains*, name.com(2019), *available at* https://www.name.com/whois-privacy.

The challenge from a technical perspective is clear: identifying the actual owner of a hive's IP. Analyzing a Microsoft case[14], we find that the same challenge exists when identifying the bot herder of a current generation botnet.

Microsoft addressed the problem simply by attempting to serve the 82 domain registrants at their individual domain name service providers associated with identified C2 servers. All attempts, of course, were unsuccessful but the court determined that they had been constructively served.

## 2. Exploring the Hypothetical

An easy way to explain the issues involved is the use of a hypothetical attack, ostensibly conducted by our fake individual, Ms. Schmidt. To our knowledge, no attack such as this has yet occurred. However, it is entirely plausible, and all the technologies – if not the products - employed here currently exist.

In this hypothetical attack, the victim is a bank located in Los Angeles, and has highly developed cyber defense systems including such capabilities as artificial intelligence-based monitoring. The monitoring uses unsupervised training algorithms for its machine learning functionality. It is capable of detecting, classifying and responding to imminent threats as soon as the threats hit its perimeter. It is constantly learning from the network traffic. From the outside (Internet), anyway, it appears impenetrable.

Over a weekend the on-duty help desk engineer notices that several accounts are being accessed simultaneously and their contents exfiltrated. This amounts to a mass withdrawal of money and payment card information from all accounts – saving, checking, etc. – held by the affected customers.

However, analysis of the readouts of the cyber defense systems show nothing wrong. There appears to be no attack. Clearly something is amiss however because over 500 accounts are being simultaneously emptied. There is no indication of where the attack is coming from and no indication of where the money is being exfiltrated to. The help desk engineer disconnects the bank's Internet connection and calls for help.

The forensic engineers respond immediately and begin to perform cyber forensic analysis on the affected servers to determine a cause. They find no evidence of malware. Whatever bot was in the system has erased its tracks and destroyed itself. Loss to the bank's 500 affected customers is over a million dollars. The attack appears to have taken less than 5 minutes to succeed and has left no trace, even within the sophisticated AI defenses.

The bank's AI defense systems are, as we pointed out, unsupervised learning systems. That means that it develops its ML training set "on the fly" and there is no preconceived training set. The benefit to that is that the unsupervised training set cannot be guessed as easily as a preconceived training set used in supervised learning. In other words, a supervised learning training set is just a database that can be stolen, modified and returned to reprogram the ML.

Conversely, an unsupervised training set notices everything that happens on the network and adds it to its database. Were it to see an effort to reprogram it, the AI would interdict the effort. The only way to do an attack is querying the oracle [15], recreating the training set empirically and attempting to replace portions of the existing training set with the modified one. Attacking the training set directly both would not work and any bot entering the network for that purpose would be intercepted.

Obviously, it develops that the only way this black box attack could have succeeded would have been querying the oracle, creating a nearly duplicate training set and imposing portions of the training set on the target. The easiest way to do that is to take the supposed training set created by querying the oracle, creating a small perturbation that would cause no damage and repeatedly hit the target with it using a large number of swarmbots. Since it causes no damage the ML decides it is OK – normal operation – and starts to ignore it.

Once the ML starts to ignore it, the attacker might introduce a bot that looks to the ML as harmless and that bot starts exfiltrating user credentials. The exfiltration might look to the ML like legitimate web site accesses. Once the

[14] Microsoft Corporation v. John Does 1 - 82, (United States District Court for the Western District of North Carolina Charlotte Division).

[15] See Section V, Black Box Attacks

bot is finished it cleans up all traces of its activity and destroys itself. There is no indication that it ever was in the bank's system and all the cyber forensics engineers in Silicon Valley cannot find its artifacts.

Now, on the weekend when the attacker is sure that the bank is on a skeleton staff it performs what security professionals refer to as a "smash-and-grab". At a predetermined time, using a preprogrammed script, swarmbots impersonating all bank customers whose credentials were stolen "log in" to their accounts. The system does not see this as an attack because proper credentials are being used and no withdrawal policies are being violated.

The accounts are emptied to remote servers that immediately ship the money to anonymous bit coin accounts and the servers destroy themselves and any evidence of their existence. The attack has succeeded completely thus far, cyber forensic engineers have nothing and the bank's legal team – even with the assistance of outside cyber lawyers – has no evidence to pursue.

We could make this more difficult still to track if we specify that all authentication and identification of and to the servers that receive the money payloads be created using the blockchain[16]. This creates the ultimate (today) anonymity. That means that the only place where a trail might be picked up – the exfiltration – is obfuscated by the blockchain. This constitutes an almost perfect next generation attack.

We now must ask what resources the adversary would need to accomplish this attack successfully. The use of a hivenet with thousands of swarmbots would accomplish the attack nicely. The original hive is programmed to accomplish the mission of emptying a million dollars from the bank's customer accounts. The hive then creates the hivenet without human intervention and destroys itself destroying any connection to its creator. The hivenet creates the swarmbots necessary to accomplish the mission, again, without human intervention. The attack could be coordinated to run simultaneously against several banks across the globe with disastrous results.

Today there are no laws that are designed to address this. If we could gather the evidence we could, perhaps, establish jurisdiction in Los Angeles. But jurisdiction over who? We don't even have enough evidence pointing to IP addresses and their supposed owners to take the Microsoft approach.

In United States law you cannot establish personal jurisdiction if you cannot serve the defendant. Microsoft assumed that the IP addresses were registered by their owners (even though the aliases made that unlikely) and served the pseudo owners.

Here we have no owners of IP addresses because the IP addresses were created using blockchain technology[17] and all communication is encrypted and via the blockchain. The attack succeeds and there is no forensic evidence for the lawyers to use. Were there such evidence the legal issues would simplify greatly.

This hypothetical illustrates the need for close collaboration between legal and cyber experts. It also illustrates the need for more and better-trained cyber lawyers and cyber experts.

# V. Legal Challenges

For adversarial machine learning to occur, the adversary must do several things. What those things are depends upon the adversary's chosen attack model. There are three[18]. They all need to retrain the ML classifier to recognize a malicious training element (a "perturbation") as non-malicious. It is the classifier's job to classify elements as malicious or not and it uses an algorithm for that purpose.

**While box attacks** – the attacker knows everything about the target's model and can reproduce it, making small perturbations in it to fool the classifier and retraining the victim. This is almost always likely to be an insider attack since access to the underlying algorithm and the original training set are necessary.

[16] A block chain is a database of "blocks" linked cryptographically. Only the cryptographic hash can access the contents of the block and there is no non-blockchain identification associated with it. It is used commonly by cryptocurrencies such as bit coin but has numerous other applications as well. – various sources

[17] See https://blockchain-dns.info/ with the understanding that the technology likely would be used but no service such as this one would offer the requisite anonymity.

[18] Mohammad Samragh Bita Darvish Rouani, Tara Javidi, Farinaz Koushanfar, *Safe Machine Learning and Defeating Adversarial Attacks*, 17 IEEE SECURITY & PRIVACY 31(2019).

**Gray box attacks** – The attacker knows the underlying algorithm, the model topology and nothing else. The training set either must be deduced or captured in some way before the classifier may be reprogrammed to believe that introduced perturbations in the training set are acceptable.

**Black box attacks** – The attacker knows nothing about the target and must derive the training set and the underlying algorithm for the classifier. The attacker does this by "querying the oracle". That consists of getting the target to respond to a large variety of sophisticated attacks and collecting the results. Having done that, the attacker may be able to reproduce the training set and the classifier algorithm allowing retraining of the target.

All adversarial machine learning (AML) attacks fall into one of these three categories. The ability of the law to establish personal jurisdiction depends upon what data about the hive master the technical investigative team can uncover from the residue of the attack.

The biggest legal challenge is availability of data. Without identifying data, the Microsoft approach discussed above is our best bet and, as we showed, not particularly reliable. What is needed is a closer partnership between lawyers and cyber technologists. Beyond that, a new area of cyber forensic science is necessary.

Almost no universities teach malware forensics. In a sample of the top 34 universities worldwide teaching computer science only 14 teach malware analysis, the starting point for the level of cyber forensics needed here[19]. The most difficult problem in identifying a cyber adversary is attribution. Lin emphasizes the difficulty of attribution, its many faces and approaches to solving the problem [20] by citing Rid and Buchanan [21] " … thinking about attribution is currently based on three assumptions …". Those assumptions are,

1. *"attribution is a largely intractable problem because of the technical characteristics and the geography of the Internet"*

2. *"attribution is either possible or not possible in any given case of interest"*

3. *"the main challenge in attribution is finding the evidence itself and not in interpreting or using it"*

All three assumptions, true in 2014, are equally true now. The only difference is that there now is a fourth assumption that we add here as a corollary to the third:

The main challenge in *finding the evidence* may be attributed to the increasing complexity of malware including botnets, swarmbots, hivenets and machine learning, including adversarial machine learning.

Thus, we have a circular problem. First, we have the technical characteristics of the Internet. Second, we have the challenge of finding the evidence. Third, we have the prickly issue of jurisdiction in cyberspace, in part due to the technical characteristics of the Internet.

Stephenson takes up the third issue – jurisdiction in cyberspace – and addresses it with the three-prong test[22]. But Rid and Buchanan's third assumption is, we submit, really the first and most difficult to overcome. It is difficult because it subsumes assumptions one and two.

Taken together, if one can solve the challenges leading to the three assumptions one can manage jurisdictional conflicts in any number of standard, well-established ways.

The legal community must, therefore, address its forensic needs with the technical community so that the technical community can provide *evidence*, not, simply, data. The evidence that the legal community needs is focused upon the requirements for attribution as laid out by Lin in his Aegis Series paper[23].

[19] Swetha Gorugantu, Malware Analysis Skills Taught in University Courses (2018) Wright State University, The Graduate School).

[20] Attribution of Malicious Cyber Incidents. No. Aegis Series Paper No. 1607(2016).

[21] Thomas Rid and Ben Buchanan, *Attributing Cyber Attacks*, 38 JOURNAL OF STRATEGIC STUDIES (2014).

[22] See Section III above

[23] Herbert Lin. 2016.

# VI. Conclusions and Possible Solutions

For as dire as the hypothetical sounds, there are solutions. However, the solutions are not simple, and they require global participation. The finding of the very small number of universities globally that teach the skills needed to start a career in malware forensics is quite troubling.

### 1. An Emerging role of Cyber-Legal Subject Matter Expert is needed

Further, a global recognition of what the field of cyber-legal practice requires on the technical side is lacking. The American Bar Association is typical of key legal organizations underplaying both the role of a cyber-legal practitioner and underplaying the educational requirements. It takes the position that[24], "…you do not need to major in computer science, information technology, or cybersecurity—though certainly that can help." We do not agree.

Consider the medico-legal field. Most medico-legal practitioners have an advanced degree in both the law and medicine – usually a JD and an MD. Our position is that cyber-legal practitioners have no less need for advanced education. Certainly, a JD (if one wishes to practice) or PhD/LLD (if one wishes only to advise) in law and a PhD in cyber science, cyber security or a related field are necessary.

This presumes that there is an actual field called cyber-legal practice. A brief survey of the Internet using Google and searching for "+cyber +legal +practice" revealed that the common definition is nothing close to the issues discussed in this paper. The focus is on cyber law and addresses the usual public and private legal issues albeit in an Internet context.

However, like the medico-legal field, the cyber-legal field is fraught with special problems, most arising from forensic evidence in the emerging autonomous world of the Internet. Consider, for example, a defamation case in which an individual is defamed in a public forum such as Facebook or Twitter.

If the defamer is an individual who can be identified the usual findings around jurisdiction and the elements of defamation are met. The suit goes to court.

However, suppose the defamer is an autonomous bot. Today we know that such bots routinely make political statements in online forums, it is no stretch at all that such bots – not even autonomous – could do likewise and defame.

Consider some of the technologies discussed in our hypothetical. Clearly such bots – now part of a swarm with instructions from a hivenet - could do the same thing in online forums. Today's "cyber lawyer" would be helpless without outside help.

Cyber scientists, likewise, would not have the legal insights to understand what evidence would need to be discovered without legal training. The role of the cyber-legal specialist is to know both sides of the equation and help bring the defamation to a satisfactory conclusion.

### 2. Cyber Forensic Training Including Advanced Malware Analysis and Evidence Identification and Gathering Needs to enter Education Programs

As we mentioned earlier in this paper and noted in a paper published over a year ago[25] in the "Legal Issues Journal", Digital – or "cyber" if one prefers – forensics education is in a sad state. In 2018 we agreed with the author's position that even the designation "digital forensic science" was arguable as to its accuracy. That may or may not still be the case but in either case it is of less concern here than the depth of the education required to make such a discipline useful today for the legal community.

The legal community depends for its success on evidence. One can testify ad nauseum but if evidence does not support the testimony the case will not go well. The problem, as we have abundantly illustrated, is significantly

[24] Leonard Wills, *How to Become a Cybersecurity Lawyer*, American Bar Association(2018), *available at* https://www.americanbar.org/groups/litigation/committees/minority-trial-lawyer/practice/2018/how-to-become-a-cybersecurity-lawyer/.

[25] P. R. Stephenson, *Digital Forensic Science: An Oxymoron?*, 6 LEGAL ISSUES JOURNAL 95(2018).

exacerbated when the stakes are high, and the adversary takes advantage of technologies such as artificial intelligence.

The solution is a difficult one both from the perspective of willingness of universities to add courses that are not profitable – no matter how much they benefit society – and from the more important perspective of qualified instructors. The solution to this challenge truly is a “bootstrap” operation.

## 3. Lawyers and Courts Need to Become Better Acquainted with Advanced Cyber Threats

While there may be ample income in cyber law as it is seen today (which we doubt given some interviews with cyber lawyers) there are very few law firms that view cyber law as a practice in itself. We contacted several self-designated cyber lawyers in the preparation of this paper and found that that put the designation on their web sites for marketing purposes, but they really have no practice in the field. What they do have generally is built around child pornography, defamation and infringement of intellectual property. None of these are inherently cyber crimes even though they may have cyber elements.

We have been privileged to work with one very large law firm that had conducted an intricate cyber investigation of a complicated (at the time) breach. We were called upon to validate the results of a private security firm’s investigation of the breach.

The lawyers we worked with had an excellent knowledge of the technology involved and we ended up coming to the same conclusion. Although the primary lawyer we worked with did not have a PhD in cyber science he was a prototype of the type of cyber lawyer that our society needs today.

## 4. Conclusions

There are some important conclusions to draw from this paper and the research behind it.

First, the stage on which malicious cyber event play out is changing dramatically. Although well over 80% of so-called cyber-related legal cases can be solved without recourse to cyber technology [26], there is an increasing number that are hard cases[27] from the cyber perspective.

For those cases that involve artificial intelligence in general and machine learning in particular, the challenges border on the extreme as our hypothetical illustrates. We are not, today, prepared legally or technically to address these new challenges. However, and much more important, lawyers and cyber subject matter experts (SMEs) are not yet prepared educationally to work together.

When the law takes on the creator of an autonomous malware system (hivenets and swarmbots, for example) it will have to depend upon expert witnesses and even then the interactions are very complicated. Our conclusion is that the better solution is to develop a cadre of cyber-legal expert practitioners who can aid the legal system in responding to the enhanced cyber threats that we see on the horizon.

Perhaps even more important, though, is the obvious prediction that nation-states and sub-state actors will be among the early adopters of these malicious technologies due to their resources and global objectives. That suggests that the cyber-legal aspects of artificial intelligence need to become part of every nation’s cyber governance.

We are not there yet and the clock is running out as more and more AI appears in virtually every aspect of our society, especially cyber security.

## Bibliography

Expert System, *What is machine learning? A definition*, Expert System(2017), *available at* https://expertsystem.com/machine-learning-definition/.

[26] Research for the author’s as yet unpublished book on cyber jurisdiction

[27] Martin Stone, *Formalism*, *in* OXFORD HANDBOOK OF JURISPRUDENCE AND PHILOSOPHY OF LAW (Jules Coleman; Scott Shapiro ed. 2004). The author distinguishes between hard cases and easy cases, easy cases being those for which a solution is obvious using well settled law

Derek Manky, *The Evolving Threat Landscape - Swarmbots, Hivenets, Automation in Malware*, CSO Online(2018), *available at* https://www.csoonline.com/article/3301148/the-evolving-threat-landscape-swarmbots-hivenets-automation-in-malware.html.

Nicholas Carlini, *A Complete list of all (arXiv) Adversarlial Example Papers*, Nicholas Carlini(2019), *available at* https://nicholas.carlini.com/writing/2019/all-adversarial-example-papers.html.

Peter Stephenson, A Framework for Determining U.S. Jurisdiction in Cyberspace (2019) University of Leicester, School of Law).

Michael N. Schmitt, Tallinn manual 2.0 on the international law applicable to cyber operations  (Prepared by the International Groups of Experts at the Invitation of the NATO Cooperative Cyber Defence Centre of Excellence 2 ed. 2017).

Fake Name Generator, *Your Randomly Generated Identity*. Fake Name generator online tool (2019), *available at* https://www.fakenamegenerator.com/gen-random-gr-gr.php.

name.com, *Get Advanced Security for Your Domains*, name.com(2019), *available at* https://www.name.com/whois-privacy.

Microsoft Corporation v. John Does 1 - 82, (United States District Court for the Western District of North Carolina Charlotte Division).

Mohammad Samragh Bita Darvish Rouani, Tara Javidi, Farinaz Koushanfar, *Safe Machine Learning and Defeating Adversarial Attacks*, 17 IEEE SECURITY & PRIVACY 31(2019).

Swetha Gorugantu, Malware Analysis Skills Taught in University Courses (2018) Wright State University, The Graduate School).

Attribution of Malicious Cyber Incidents. No. Aegis Series Paper No. 1607(2016).

Thomas Rid and Ben Buchanan, *Attributing Cyber Attacks*, 38 JOURNAL OF STRATEGIC STUDIES (2014).

Leonard Wills, *How to Become a Cybersecurity Lawyer*, American Bar Association(2018), *available at* https://www.americanbar.org/groups/litigation/committees/minority-trial-lawyer/practice/2018/how-to-become-a-cybersecurity-lawyer/.

P. R. Stephenson, *Digital Forensic Science: An Oxymoron?*, 6 LEGAL ISSUES JOURNAL 95(2018).

Martin Stone, *Formalism*, *in* OXFORD HANDBOOK OF JURISPRUDENCE AND PHILOSOPHY OF LAW (Jules Coleman; Scott Shapiro ed. 2004).