CONTEMPORARY CYBER-SECURITY TRENDS AND RELATED ETHICAL DILEMMAS (PART 2) [1]

VESSELIN BONTCHEV
vesselin.bontchev@nlcv.bas.bg

DIMITRINA POLIMIROVA
dimitrina.polimirova@nlcv.bas.bg

National Laboratory of Computer Virology, BAS

Abstract
In Part 1 of this paper we considered the ethical dilemmas posed by ransomware and by the so-called "surveillance capitalism". In Part 2 we shall examine the ethical dilemmas related the offensive tactics in which people normally concerned with computer defense might engage in, to vulnerability disclosure, and to artificial intelligence.

Keywords: ethics, ethical dilemmas, cyber security.

2.3. Offensive defense

Normally, defends against cyber attacks involves hardening one's computer and network infrastructure by installing defensive programs, configuring them properly, regularly updating the systems, and monitoring them for early signs of attacks. However, occasionally the defenders have the opportunity to take more aggressive actions, in order not just to stop but also to hurt the attackers – actions, which are sometimes called "hacking back". Such opportunities often present various ethical dilemmas regarding the proper way to proceed [SolidState].

2.3.1. Accessing criminal infrastructure

Attackers almost never use their own computers to conduct an attack. They either use somebody else's machine that they have hacked into, or they set up cheap or free accounts at a cloud hosting provider. Since they have to access this machine, in order to conduct the attack from it, they need to set up remote access to it. Sometimes, either due to sloppiness or due to the incompetence of the attacker, this remote access is not properly secured. In such cases, when the victim detects an attack coming from that machine, they are able to access it remotely too.

This has clear benefits for the defender. Often it means stopping the attack by removing or otherwise disabling the attacking programs running on the remote machine. It can also allow the gathering of valuable data about the attacker – data, which can help with the investigation of the incident and can be provided to the law enforcement.

On the other hand, accessing somebody else's machine without authorization is at the very least legally questionable in most jurisdictions. Furthermore, any actions taken to stop the attacking software from running risk to damage something else on the machine – and in at least some cases, this machine belongs to an innocent victim that the attacker has hacked.

2.3.2. Taking down criminal infrastructure

A more extreme case of the above is when the defenders disable not just the attacking software running on a single machine that is attacking them. Attackers often harness thousands of hacked machine into so-called "botnets" that are used to attack victims indiscriminately. If this infrastructure, used by the attacker, is itself vulnerable to being hacked into, there exists the possibility to disable it entirely and thus put a stop not only on the attacks against one's own machines but also make them stop of all the victims of the criminal. Examples of such operations of taking down botnets include [Boscovich, 2011], [Fisher, 2017], and [Vojtěšek, 2019].

While the benefits from such actions are quite obvious, they are also much riskier, because the bots are running on a large multitude of different machines, the exact specifics of which are not known sufficiently well. The risk that the takedown of the attacking software would make at least some of them inoperable, thus causing harm to their legitimate owners, is much higher. This is why such activities are usually undertaken in coordination with law enforcement and only after careful weighting of the different arguments for and against and when no better alternative can be found.

2.3.3. State-sponsored cyber operations

The days when hackers were mostly smart kids working alone from their mother's basement are long gone. Nowadays the vast majority of cyber attacks are caused by professional criminals, often working for criminal organizations. However, there is increased percentage of attacks in cyberspace that are state-sponsored. In fact, there is an ongoing world-wide low-profile cyber conflict, involving many nations. In most cases their activities are related to cyber-espionage but, increasingly, other kinds of actions are undertaken as well. One example is the

Russian influence campaign during the US Presidential elections in 2016. Another is the activities of North Korea, which often targets financial institutions and steals financial assets, in order to finance its nuclear program [Wainer, 2019].

In some cases, however, these cyber attacks are much more direct and aim to cause harm to some physical infrastructure of the adversary. Probably the best known example is the Stuxnet virus [Shearer, 2010], which was developed by the USA and Israel (with some minor participation of the UK, France, and the Netherlands) and was used to infect the computers controlling the uranium enrichment centrifuges of Iran and to damage these centrifuges by operating them improperly, thus setting back Iran's uranium enrichment program for several years.

These kind of offensive cyber activities pose their own kind of ethical dilemma. On the one hand, the people conducting them usually serve in the military and it is their duty to conduct these attacks, in order to harm an adversary of the country they serve. On the other hand, such activities often have unintended consequences and do harm innocent victims. For instance, while indeed most Stuxnet infections were in Iran, as every virus, it spread uncontrollably and hit victims in Indonesia, India, Azerbaijan, Pakistan, and even in the USA itself. In another example, NotPetya, a ransomware worm that was developed by Russia and used against Ukrainian targets, managed to spread globally and cause more than $10 billion damage to completely unrelated companies around the world [Greenberg, 2018].

2.3.4. Unlocking smart phones

Contemporary smart phones (especially the iPhone) are very well-protected. The contents of their storage is encrypted and cannot be decrypted unless the user unlocks the phone. Given that these days people store information about their whole lives on their smart phone, such measures are very important, in order to protect the data if the phone is stolen, for instance.

However, by itself, technology is ethically neutral. It benefits all of its users – be they good or bad. The same technology that prevents the data of theft victims from being accessed by the thief also prevents the law enforcement from accessing the data stored on a criminal's phone. This is causing an endless amount of frustration among the law enforcement officials and they keep insisting that the technology companies should provide some way to access lawfully the encrypted contents of these devices.

A particularly well-known case was when the FBI wanted to decrypt the contents of the iPhone belonging to a domestic terrorist involved in a mass shooting who was already dead and could cooperate [James, 2016]. Attempting to guess the unlock code would have resulted in progressively increasing times until the next guess was allowed and after 10 incorrect guesses the device would have erased its contents anyway. The FBI demanded that Apple produced and new version of the operating system of the device that disabled this protection and pushed it to the phone in question. Apple refused, arguing that doing so would reduce the security of all users of their devices. The argument went to court but eventually the FBI dropped the case, after finding an undisclosed method to hack the device despite the protection.

However, such situations present a clear ethical dilemma: should the company-producer cooperate with law enforcement in order to help solve a crime (and potentially save human lives) – or should it rather refuse and help preserve the security of everyone, including the criminals?

2.3.5. Encryption backdoors

An issue similar to the one described in the previous section is posed by the so-called encrypted instant messengers. Those are mobile applications that allow their users to communicate in real-time via text messages, voice, and video. There are many of those – Signal, WhatsApp, Telegram, etc. When the encryption protocols used by them are implemented properly, the communication is end-to-end encrypted, meaning that only the two communicating parties can decrypt it at their respective ends. Even the company that has produced the application and is running the servers through which the communications are passing is unable to decrypt them.

On the one hand, this is a wonderful property that helps protect the privacy of the users. On the other hand, a relatively small percentage of the user population (drug dealers, terrorists) use it to communicate securely and avoid law enforcement interception.

Naturally, this does not sit well with the law enforcement officials and there are periodic calls that the companies producing these applications should built into them some way that allows the law enforcement to listen in, when empowered by an appropriate search warrant. Unfortunately, as the cryptography experts keep insisting, there is no way to achieve such selective access. If the backdoor is found, it can be used to violate the privacy of any user. Plus,

nothing prevents the criminals from using an application that does not contain such a backdoor. After all, the necessary mathematical algorithms are accessible to anyone.

In the USA, the most recent developments in this area was a speech by their Attorney General William Barr [Owen, 2019]. He repeated the calls for the communication companies to implement such access for the law enforcement and implied that if they don't do it on their own free will, the politicians will pass the necessary laws that force them to do so. In fact, such legislation has already been passed in Australia.

The ethical dilemma here is clear. Should the government be forcing an issue that would threaten to reduce the security of all citizens – in the name of solving (or even preventing) crimes that threaten the security of a few citizens? And should the technology companies cooperate, if they have the choice in the matter? These questions should be considered in the context that even without access to encrypted communications, the police is far from helpless to solve such crimes.

2.3.6. Government spyware

As mentioned in the previous section, the police has alternate means to monitor the conversation of criminals when end-to-end encryption is involved. One of them is infecting the criminal's phone the malware. The encryption is end-to-end, which means the communication cannot be read in-transit – but it is decrypted at the endpoint, meaning that malware running on the endpoint would be able to access its contents and, if necessary, forward it to the law enforcement.

This isn't as convenient as a backdoor built into the encryption protocol. The main problem is that such and approach does not scale well. You might be able to infect the phone of a single suspect, but doing it for thousands of people is problematic. Somebody (most likely the Italian police) has had the bright idea of solving this problem by Trojanizing popular mobile applications and uploading them to the official Android Play Store [Franceschi-Bicchierai, Coluccini, 2019].

The ethics of such an approach (even if we leave aside its legality) is clearly questionable. The malicious applications have endangered the privacy and security of all users – not just of selected suspects.

2.3.7. Whitelisting government malware

Given that the approach described in the previous section has become a reality and that it was discovered by security researchers, an interesting question arises. Should the producers of anti-virus programs "whitelist" (i.e., avoid the detection of) government-created malware – in order not to jeopardize a possibly ongoing investigation or an intelligence operation?

Fortunately, to the best of our knowledge, this question has never had to be answered in practice yet. Many anti-virus companies maintain that it is against their policy to do so [Westervelt, 2013]. However, it is questionable how much they would be able to resist if, say, served with a National Security Letter. When the Stuxnet virus was discovered and it became clear that it was US-made and was targeting Iranian installations, at least some American anti-virus researchers had doubts whether it would be proper to go public and disclose the operation. On the other hand, it is likely that this question will never have to be tested in practice, because it is relatively easy for a government intelligence agency to create malware that is not detected by any of the existing anti-virus programs (and to do so without any cooperation whatsoever from the producers of these programs) and by the time it is actually discovered, it has already done its work and the operation can be safely shut down anyway.

2.3.8. Anti-virus companies creating viruses

There is an even more interesting, from the ethics perspective, question than should the anti-virus companies avoid detecting government malware. This is the question should the anti-virus companies create new viruses themselves?

Anti-virus companies are often accused of doing that, based on the primitive reasoning that the more viruses there are, the more demand their will be for their products. Of course, this reasoning is completely unfounded – there are more than enough malicious programs floating around and new ones are created at the rate of more than a million per day. The anti-virus companies are overworked and barely capable of keeping up with this glut; they certainly have neither the time, nor the inclination of create even more malware. However, as explained in [Bontchev, 1998], there are cases when the question is not so clear-cut.

When Microsoft changed the macro programming language of Microsoft Word from WordBasic to Visual Basic for Applications (VBA) in 1997, they included a converter that would automatically translate the WordBasic macro of any opened document into VBA. This was done in order to ensure that existing macro packages that the users had would continue to

work – because starting from Word97, Microsoft Word can no longer execute WordBasic programs. Unfortunately, this means that the existing WordMacro viruses in any opened and infected document would be automatically converted into VBA too. WordBasic and VBA are two completely different languages, with different syntax and different internal representation. Such a conversion would effectively result in a new, previously unknown virus.

Some anti-virus researchers expressed the view that we should perform this action ourselves – take all existing WordBasic macro malware and produce the corresponding VBA malware from it (essentially creating new viruses in the process), then implement detection of the new malware in our products, so in the cases that it was created "naturally" in the wild, our products would already be capable of detecting it.

The grounds of this suggestion were clearly ethical and well-meaning – the idea was to do what is our duty and to provide the best protection to our users that we could. On the other hand, some of us argued that since we have vowed to fight computer viruses, it would be unethical for us to create more of them. A detailed discussion of the negative effects of such an act can be found in the paper cited above. Furthermore, we managed to come up with a method, that allowed us to implement detection of these possible future viruses without actually creating them [Bontchev, 2000].

2.3.9. Beneficial viruses

Traditionally, computer viruses have been used for various malicious purposes. However, given that they are just programs (that can replicate themselves), people are often questioning whether they could be used for something beneficial, too. Even the pioneer in the theory of computer viruses, Dr. Fred Cohen, has expressed such ideas.

A detailed discussion of the technicalities and the feasibility of this idea is beyond the scope of this paper. However, we would like to point the reader to a paper of ours [Bontchev, 1994], where this issue is discussed extensively. Despite being written quarter of a century ago, all the arguments presented there are still valid and actual. The general conclusion is that no, it is not possible to have beneficial computer viruses. Any beneficial program capable of self-replication would install itself on the user's computer only with the explicit permission of the user, would use cryptographically strong means for authentication (digital signatures), and wouldn't call itself a virus.

2.4. Vulnerability disclosure

Occasionally, security researchers discover vulnerabilities in the computer systems they work with. Silently exploiting them for their own benefit and to the detriment of the owners of these systems would be, of course, unethical. However, even revealing these vulnerabilities, so that they could be fixed, is fraught with ethical dilemmas. This section will attempt to cover these dilemmas.

2.4.1. Full vs. coordinated disclosure

There are two main schools of thought regarding how the information about a vulnerability should be disclosed [Trull, 2015]. The *first* school of thought, the so-called "full disclosure", maintains that the information should be made public, available to anyone who is interested.

The main advantage of this approach is that it has a wide reach and any user of the vulnerable product can be informed about the vulnerability. Unfortunately, the main disadvantage of this approach is that malicious actors would also be informed and are likely to start exploiting the vulnerability for criminal purposes and to the detriment of the legitimate owner of the vulnerable installation.

The *second* school of thought is that the vulnerability should be disclosed privately to the company making the vulnerable product and sufficient time should be allowed for the company to publish a fix before the vulnerability is made public. Originally, this was called "responsible disclosure" but due to the implication that any other disclosure method is considered "irresponsible", nowadays the term "coordinated disclosure" is used instead.

The advantage of this method is clear – only the company responsible for fixing the product is informed and the problem can be fixed before any bad actors get the chance to exploit the vulnerability. Unfortunately, companies often tend to ignore or downplay private reports of security problems in their product – while a public report tends to force them to fix it. In addition, the term "sufficient time to fix the problem" is rather ill-defined and subjective, although often a period of 90 days is used. In one extreme case the person reporting a security issue was arrested and convicted for hacking [Osborne, 2016].

Which disclosure approach to use is an ethical dilemma for the security researcher discovering the vulnerability. Probably the most ethical approach is to use a combination of them – attempt coordinated disclosure first and then go with full disclosure if the company does

not react or refuses to address the issue. Various resources for aiding coordinated disclosure exist. For instance, the US CERT has published guidelines on the subject [Householder, Wassermann, Manion, King, 2017]. There exists a whole platform [HackerOne], specialized in coordinating the reports of vulnerabilities from the security researchers with the companies affected by them.

2.4.2. How much to publish

Another ethical dilemma that a security researcher who has discovered a vulnerability often faces, especially if doing full disclosure, is how much information exactly to publish. A technical explanation of what the problem is, how to check for its presence, and what is its impact (and the possible mitigations) is, of course, a must. However, this information alone is often insufficient for the average non-technical user to determine whether they are affected. In order to solve this problem, researchers often publish a so-called "proof-of-concept" (PoC) – a small program that exploits the vulnerability in a non-destructive way, demonstrating its existence in some obvious way (e.g., by launching the calculator, or by displaying some information that would be otherwise private and unobtainable).

One the one hand, doing so helps the masses asses whether the vulnerability affects them and how serious it is. On the other hand, it allows even unskilled attackers to easily exploit the vulnerability. So, whether to include a PoC in a public vulnerability report or not poses an ethical dilemma.

2.4.3. Bug bounties

Sometimes some companies give monetary rewards to the security researchers who have reported a particularly bad security vulnerability in the company's product – the so-called "bug bounties". Platforms like [HackerOne] coordinate the bug bounty programs of those companies who have such programs and ensure that the researchers who report security vulnerabilities via the platform receive their reward.

While it is always nice and stimulating if one's efforts are rewarded, this has also given rise to the opposite extreme. Some researchers refuse to submit bug reports unless they are rewarded appropriately. Their logic is that they have done some work on behalf of the company (who really should have found the problem in its product itself) and they are entitled to a compensation – the so-called "no more free bugs" slogan.

It is an ethical dilemma that the security researcher who has stumbled upon a vulnerability faces. On the one hand, reporting it and having it fixed is the duty of the security researcher. On the other hand the logic of refusing to do the job of a large corporation for free also has its merit.

2.4.4. What to do if the company doesn't react

In some cases, when the security researcher who has discovered a vulnerability in some company's product, decides to take the coordinated disclosure route, and contacts the company in question, the latter does not respond at all, or refuses to acknowledge the seriousness of the problem and/or to fix it. Such cases pose another ethical dilemma.

One the one hand, the researcher could decide to disclose the vulnerability publicly, hoping that pressure from the public would force the company to fix the problem. But, as mentioned in the previous sections, this is likely to inform the bad actors too, leading to unintended damage to the society. On the other hand, the researcher could decide not to disclose the problem. But in such cases the vulnerability remains unfixed and there is always danger that the cyber criminals would discover it on their own – again resulting in damage to the society.

All these cases are usually not equivalent; in every particular case the researcher should consider the circumstances carefully and should take the correct decision in accordance to their ethical principles.

2.4.5. Sale of exploits

An extension of the case of bug bounties is when the security research simply sells knowledge about the vulnerability they have found (and how to exploit it) not to the company that is making the vulnerable product but to the highest bidder. Market demand for such goods definitely exists, as witnessed by the existence of several companies whose main line of business is brokering such deals. One of the best known companies of this kind is Zerodium; they often publish their "price list" of the kind of vulnerabilities they are willing to buy and the prices of some often exceed a $1-2 million [Greenberg, 2015].

Such companies often advertise their services as being able to help security researchers who have found a vulnerability to get a fair price for their hard work. However, anybody considering doing business of them should be aware that they are likely to sell the vulnerability to governments. In some cases these governments might use it perfectly legally – e.g., in order to obtain access to a criminal's computer system after obtaining a proper court order. However

not all governments are democracies and sometimes the buyers are likely to use it against political dissidents, suppressed minorities, and in other unsavory ways. Once the researcher has sold the vulnerability to a brokerage company, they have no control over who and how would use is further.

In addition, the marketplace of such vulnerabilities is often risky and not all participants are sufficiently trustworthy, as explained in [Miller, 2007].

2.4.6. Sale of data leaked in breaches

As the world becomes increasingly computerized, practically every company keeps computer databases containing all kinds of information. Very often this information is sensitive. It could contain business data, like customers, financial records, etc., or personal data – names, addresses, buying habits, and so on. Unfortunately, not every company employs sufficient security workforce to safeguard this information. As a result, the sensitive information is often leaked outside the company. This process seems to get progressively worse with time – for instance, 2019 was the worst year (with the most data breaches) so far [Turner, 2019]. One of them even occurred in our country, when personal data for practically the whole adult population of Bulgaria was made publicly available [Krasimirov, Tsolova, 2019].

The leaked data often finds its way to underground marketplaces where hackers offer it for sale. The data sold this way is often humongous (multiple gigabytes) and the prices are relatively low on a per-unit basis. There is absolutely no background check of the buyers or the purpose for which they are obtaining the data – in fact, it is often presumed that the buyers are other criminals who will use it for criminal purposes. It could be simple things like spam (when e-mail addresses are leaked), credit card fraud (when credit card numbers are leaked), or even something more convoluted, like identity theft, blackmail, finding prospective targets for a burglary, and so on.

This is probably the single ethical issue described in this paper which we do not think poses any ethical dilemmas. We firmly believe that the sale of illegally leaked confidential data is unethical. However, taking measures to prevent future data breaches could pose an ethical dilemma. For instance, the study [Choi, Johnson, Lehman, 2019] shows that taking such measures in a hospital environment made the systems harder to use by the medical personal and, as a consequence, increased the mortality rate of the patients.

2.4.7. Whistleblowing

A somewhat related issue to discovering a security vulnerability in a product is discovering proof of wrongdoing in an organization. Normally, the ethical decision would be to report this wrongdoing to the authorities – this is known as "whistleblowing". Things get complicated, however, if the organization in question *is* part of the "authorities" – i.e., is a government organization. Even further complications arise when the person who has discovered the problem works for this organization, when it is an extremely sensitive security or military agency, and when its employees are bound by law to maintain secrecy. While in most democratic countries there exist established means and processes to report such problems, they are often ineffective and the reporters often found themselves ignored or even persecuted in various bureaucratic ways.

A typical example of such a situation was the case of Edward Snowden – a contractor for the National Security Agency who discovered that the Agency was breaking the law by monitoring the phone calls of American citizens [Wired, 2014]. He found the illegal activities so egregious and the reporting process so inadequate, that he saw himself forced to flee the country, taking with himself a large archive with documental proof of this wrongdoing and providing it to various media outlets.

His actions are an excellent example of the ethical dilemma a person in his position is presented with. On the one hand, a government agency was doing something clearly illegal that was invading the privacy of millions of his compatriots. Clearly, this had to be reported, so that a stop is put to these activities. On the other hand, these activities were not a goal in itself; the Agency was doing it in order to fulfill its duty of guarding the country from terrorism and the monitoring of phone calls made by Americans was just a by-product of a global dragnet of phone call monitoring. Furthermore, by making public a large trove of secret materials, he endangered ongoing operations and possibly even the lives of active agents.

This is a very hard ethical dilemma to resolve and while it is obvious what his decision was, it is by far not obvious that it was the most ethical one.

2.5. Artificial intelligence

Artificial intelligence (AI) has seen strong development and adoption lately. It is a vast area, consisting of many different sub-areas like machine learning, computer vision, face

recognition, automated language translation, expert systems, and so on. They mostly involve developing algorithms that allow computers do perform activities normally associated with intelligent beings. However, it is important to always keep in mind that AI is just software. It is not really intelligent. It can contain various bugs and reflect the biases of the people who have created it. It cannot have ethics of its own. But it is important that the people developing it and using its output are aware of the many ethical dilemmas that its use can create.

2.5.1. Racial biases in machine learning

Machine learning (ML) is a sub-field of AI. It uses algorithms, which are "trained" to recognize patterns by providing them with labeled inputs and letting them choose themselves what properties to of the inputs to consider and what weights to assign to them, in order to differentiate the "good" from the "bad" inputs. It is important that the human does not had-code in the algorithm how the differentiation should be made. The algorithm decides it itself ("learns") by examining the data.

ML has many applications in areas where pattern detection and classification is important. These include from facial recognition, stock market prediction, spam e-mail filtering, fraud detection, and many others. However, it is important to note that how good the model developed via ML is depends very much on the data it has been given to train from. If the data is biased in some way, so will be the decisions of the ML model. The data could even contain hidden biases that the person training the mode is not aware of. Since ML is so good at detecting patterns, it is likely to "discover" this bias and similarly bias its results [DeBrusk, 2018].

For instance, in 2016 Microsoft released a Twitter bot that was using ML to learn to converse with other people via Twitter. It was learning from the tweets sent to it. This was quickly exploited by pranksters who kept tweeting racist phrases to it. As a result, the bot itself became "racist" – praising Hitler and making misogynistic remarks – and had to be taken off-line.

In another example, Google engineers built an ML tool to detect automatically offensive from inoffensive tweets. When it was tested in the real world, it was found to incorrectly flag as "offensive" 1.5 times the tweets from African Americans than those from Caucasian ones. The reason was because it was trained mostly in tweets from one the latter racial group, which often happens to use slightly different English wording than the other group [MIT, 2019].

In another case, an ML-based facial recognition software was able to identify with high precision the faces of Caucasian people – while it often confused with each other the faces of people of other races. This was caused by the fact that it had been trained with a database of photos of Caucasian faces. For more examples of racial bias in ML models, the reader is referred to [Kim, 2018].

The developers of such algorithms clearly have the ethical obligation to ensure that their algorithms are not biased. This includes not only refraining from intentionally training them on biased data, but also doing everything they can with reveal any hidden biases in the data and to eliminate them.

2.5.2. Judicial decisions

While the examples given in the previous section are mostly curiosities, things are starting to get very serious when the output of AI algorithms is used unthinkingly in decisions that can decide the fate of real people.

For instance, banks have started to use ML algorithms to decide whether an applicant should be granted a loan or not. The courts in the USA are sometimes using such algorithms to decide how much jail time a convict should get, whether an accused person is a flight risk and should be let go on bail (and of what size), or whether a convict should be released early. Unsurprisingly, these algorithms are often found to contain racial bias [O'Brian, Kang, 2018].

The apologists of this approach point out that it helps the overworked judges and that eventually it is always a human that makes the final decision. However, it is way too easy for the humans in question to get into the habit of approving the computer's decisions without questioning them, simply because they turn out being correct "most of the time".

In our opinion, ML algorithms should not be used at all when the fate of a human being is at stake. The main problem of such algorithms is that they work as a black box and produce a solution, which is correct most of the time – but sometimes can be wildly incorrect. They are incapable of explaining how exactly the solution was reached. In this aspect, it is preferable to use another AI approach – the so-called *expert systems*, which are capable of explaining the line of reasoning that has lead to their decision.

One should be especially wary of AI used in law enforcement and any police unit considering such use should carefully weigh the ethics of doing so [Dechesne, Dignum, Zardiashvili, Bieger, 2019].

2.5.3. Deep fakes

One of the application of ML is to learn from a set of existing images (e.g., of a face) and then replace the face in a picture with the one just learned, so that the picture remains looking realistic. The same approach can be applied not just to static images but also to vides and voice recordings. This allows the falsification of sound and video materials in a way that looks very convincing and difficult to detect. It can have various applications of the unsavory kind.

All the examples described below pose essentially the same ethical dilemma. On the one hand, the development of ML tools for image manipulation have all sorts of useful and beneficial applications and advance the human knowledge. On the other hand, they are easy to misuse. So, those who work in this field should undertake a careful consideration of what the outcome to the society would be of any tools they create and whether it is indeed worth creating them.

2.5.3.1. Fake nudity

One example is the creation of a mobile application that could take a photo of a woman and manipulate it in such a way, that the woman looks naked [Cole, 2019]. This could be used for all kinds of nefarious purposes – for blackmail, revenge porn, bullying, and so on.

The author of the application out of curiosity and wish to experiment with the technology. However, he later reconsidered and after seeing what it could be used for, he made the ethical decision and took the application off-line.

2.5.3.2. Fake news

Another area where falsified video material could be used to cause harm is the production of so-called "fake news" – realistically looking news reports about things that did not really happen or happened in ways different than depicted. Some speculate that this will be the next generation of propaganda war, psychological operations, and influence into other countries politics.

However, others maintain that this is unlikely. People tend to believe what they want to believe (the so-called "confirmation bias") and are likely to trust rumors that confirm their worldview even if the fake news are not accompanied with high-quality falsified video material. So, using deep fakes would be mostly a wasted effort, thus it is not likely to become widespread [Grugq, 2019].

2.5.3.3. Accounting scams

Another interesting case of a deep fake (this time of a voice) used for criminal purposes is described in [Damiani, 2019]. Reportedly, the scammer used deep fake software to modify his voice to become very similar to that of the CEO of a company and ordered the fraudulent transfer of $243,000.

While there are some doubts about the veracity of this story (the scam indeed happened but whether deep faked voice was used remains unproven), it definitely sounds plausible and we are likely to see this method more often used in phone scams in the future.

2.5.4. Autonomous weapons systems

Probably the most dangerous application of AI is in the military, in the so-called "autonomous weapons systems". These are military robots, usually drones, that are capable of autonomous action – i.e., without direct human supervision. They could be set to roam freely a specified area of operations and could use computer vision and machine learning to detect and identify targets and to engage them automatically, without waiting for an explicit order from a human operator.

This, of course, creates a serious ethical dilemma about the application of such machines. On the one hand, it can be argued that by eliminating the human factor one can eliminate a lot of human errors and emotion from the battlefield. This could result in lower fatalities, especially among the civil population, and other human mistakes (e.g., caused by fear) that now often result in unnecessary human fatalities. In addition, the use of such machines is likely to lower the casualties among one's own military force. No matter how expensive they are, from ethical grounds, it is preferable to lose a robot than a human.

On the other hand, as we saw in the previous sections, AI algorithms are often biased and their judgment is still by far not as good as that of a human. Such autonomous killing machines could easily make mistakes that result in the unnecessary loss of human life. Especially in modern guerilla warfare, it is difficult even for a human, let alone to an algorithm, to determine correctly who is a combatant and who is a civilian. According to some ethicists, the use of such machines violates a fundamental principle of international humanitarian law, *jus in bello*, which requires that some human person must always be held responsible for any civilian deaths.

For a more detailed consideration of the issues involved, the reader is referred to [Etzioni, 2017].

2.5.5. AI in control of nuclear weapons

Perhaps the most dangerous application of AI would be to let it control the nuclear arsenal. Yet this is precisely what some are proposing [Lowther, McGiffin, 2019].

The argument here is that with the development of hypersonic nuclear missiles, the warning time for a nuclear attack has shrunk to mere minutes, even over intercontinental distances. Therefore, the idea is to allow an AI to launch a return strike, if a nuclear first strike by the enemy manages to obliterate the human command structure. Any other options, like making the nuclear response infrastructure more robust (and likely to withstand a first strike), improving the early warning systems, or placing weapons for a return strike closer to the potential adversary are considered insufficiently adequate.

Of course, given that AI algorithms tend to fail in rare and unexpected ways, such a solution creates tremendous risk of starting a nuclear war and obliterating the humanity, instead of ensuring the security of the country deploying it from a nuclear first strike.

3. Conclusion

As we saw in this paper, the field of information technology is rife with problems that pose ethical dilemmas. Therefore, we consider it imperative that computer science students, especially those specializing in cyber security, are required to attend courses in ethics, so that they become more capable to come up with ethical solutions to the problems they face in their future profession.

NOTES

[1] The preparation of this paper is supported by the National Scientific Program "Information and Communication Technologies for a Single Digital Market in Science, Education and Security (ICT in SES)", financed by the Ministry of Education and Science of the Republic of Bulgaria.

REFERENCES

Bontchev, V. (1994). *Are 'Good' Computer Viruses Still a Bad Idea?*, Proc. EICAR'94 Conf., pp. 25 – 47, https://bontchev.nlcv.bas.bg/papers/goodvir.html

Bontchev, V. (1998). *The "Pros" and "Cons" of WordBasic Virus Upconversion*, Proc. 8th Int. Virus Bull. Conf., 1998, pp. 153-172, https://bontchev.nlcv.bas.bg/papers/upconv.html

Bontchev, V. (2000). *Solving the VBA Upconversion Problem*, Proc. 10th Int. Virus Bull. Conf., 2000, pp. 273-299, https://bontchev.nlcv.bas.bg/papers/upsolve.html

Boscovich, R. (2011). *Taking Down Botnets: Microsoft and the Rustock Botnet*, https://blogs.technet.microsoft.com/microsoft_on_the_issues/2011/03/17/taking-down-botnets-microsoft-and-the-rustock-botnet/

Choi, S., E. Johnson, C. Lehman. (2019). *Data breach remediation efforts and their implications for hospital quality,* Health Services Research, 54, pp. 971-980, https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-6773.13203

Cole, S. (2019). *This Horrifying App Undresses a Photo of Any Woman With a Single Click*, https://www.vice.com/en_us/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman

Damiani, J. (2019). *A Voice Deepfake Was Used To Scam A CEO Out Of $243,000*, https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

DeBrusk, C. (2018). *The Risk of Machine-Learning Bias (and How to Prevent It)*, https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-and-how-to-prevent-it/

Dechesne, F., V. Dignum, L. Zardiashvili, J. Bieger. (2019). *AI & Ethics at the Police: Towards Responsible use of Artificial Intelligence in the Dutch Police*, https://www.universiteitleiden.nl/binaries/content/assets/rechtsgeleerdheid/instituut-voor-metajuridica/artificiele-intelligentie-en-ethiek-bij-de-politie/ai-and-ethics-at-the-police-towards-responsible-use-of-artificial-intelligence-at-the-dutch-police-2019..pdf

Etzioni, A., O. Etzioni. (2017). *Pros and Cons of Autonomous Weapons Systems*, https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/May-June-2017/Pros-and-Cons-of-Autonomous-Weapons-Systems/

Fisher, D. (2017). *Taking down an android botnet*, https://digitalguardian.com/blog/taking-down-android-botnet

Franceschi-Bicchierai, L., R. Coluccini. (2019). *Researchers Find Google Play Store Apps Were Actually Government Malware*, https://www.vice.com/en_us/article/43z93g/hackers-hid-android-malware-in-google-play-store-exodus-esurv

Greenberg, A. (2015). *Here's a Spy Firm's Price List for Secret Hacker Techniques*, https://www.wired.com/2015/11/heres-a-spy-firms-price-list-for-secret-hacker-techniques/

Greenberg, A. (2018). *The Untold Story of NotPetya, the Most Devastating Cyberattack in History*, https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/

Grugq, T. (2019). *Cheap Fakes beat Deep Fakes*, https://medium.com/@thegrugq/cheap-fakes-beat-deep-fakes-b1ac91e44837

HackerOne, https://www.hackerone.com/

Householder, A., G. Wassermann, A. Manion, C. King. (2017). *The CERT® Guide to Coordinated Vulnerability Disclosure*, https://resources.sei.cmu.edu/asset_files/SpecialReport/2017_003_001_503340.pdf

James, R. (2016). *Apple Vs. FBI – The Summary Of Events*, https://www.beencrypted.com/apple-vs-fbi-events-summary/

Kim, S. (2018). *Racial Bias in Facial Recognition Software*, https://blog.algorithmia.com/racial-bias-in-facial-recognition-software/

Krasimirov, A., T. Tsolova. (2019). *In systemic breach, hackers steal millions of Bulgarians' financial data*, https://www.reuters.com/article/us-bulgaria-cybersecurity-idUSKCN1UB0MA

Lowther, A., C. McGiffin. (2019). America Needs a "Dead Hand", https://warontherocks.com/2019/08/america-needs-a-dead-hand/

Miller, C. (2007). *The Legitimate Vulnerability Market*, https://www.econinfosec.org/archive/weis2007/papers/29.pdf

MIT Technology Review (2019). *Google's algorithm for detecting hate speech is racially biased*, https://www.technologyreview.com/f/614144/googles-algorithm-for-detecting-hate-speech-looks-racially-biased/

O'Brian, M., D. Kang. (2018). *AI in the court: When algorithms rule on jail time*, https://www.apnews.com/20efb1d707c24bf2b169584cf75c8e6a

Osborne, C. (2016). *Hacker thrown in jail for reporting police system security flaws*, https://www.zdnet.com/article/hacker-thrown-in-jail-for-reporting-police-system-security-flaws/

Owen, M. (2019). *US Attorney General Barr doubles down on encryption backdoors call*, https://appleinsider.com/articles/19/07/23/us-attorney-general-barr-doubles-down-on-encryption-backdoors-call

Shearer, J. (2010). *W32.Stuxnet*, https://www.symantec.com/security-center/writeup/2010-071400-3123-99

SolidState, *Hack Back Pros and Cons: What You Need to Know Before Striking Back*, http://solidsystemsllc.com/hack-back/

Trull, J. (2015). *Responsible Disclosure: Cyber Security Ethics*, https://www.csoonline.com/article/2889357/responsible-disclosure-cyber-security-ethics.html

Turner, S. (2019). *2019 Data Breaches – The Worst So Far*, https://www.identityforce.com/blog/2019-data-breaches

Vojtěšek, J. (2019). *Putting an end to Retadup: A malicious worm that infected hundreds of thousands*, https://decoded.avast.io/janvojtesek/putting-an-end-to-retadup-a-malicious-worm-that-infected-hundreds-of-thousands/

Wainer, D. (2019). *North Korea Hacks Banks, Cryptocurrencies for Funds, UN Finds*, https://www.bloomberg.com/news/articles/2019-08-06/north-korea-hacks-banks-cryptocurrencies-for-funds-un-finds

Westervelt, R. (2013). *Antivirus Firms: Whitelisting Malware For Law Enforcement Against Policy*, https://www.crn.com/news/security/240159502/antivirus-firms-whitelisting-malware-for-law-enforcement-against-policy.htm

Wired (2014). *Edward Snowden*, https://www.wired.com/2014/08/edward-snowden/