When artificial intelligence goes awry: separating science fiction from fact

Provided by

Malwarebytes LABS

Introduction: a peek into the recent past

As a society, we are already using artificial intelligence (AI) across a host of industries. Speech recognition. Autofill. Biometrics. Machine learning platforms. This technology train has left the station, and we will soon bear witness to its widespread adoption.

Experts say: Queue immediate destruction of society.
But we believe there's a crucial period in artificial intelligence's development—in fact, in any technology's development—where those bringing this infant tech into the world have a choice to develop it responsibly or simply accelerate at all costs. Because if we don't think about this now, we know threat actors will.

So what if someone figures out how to abuse AI applications? In recent years, we've witnessed the market for smart home assistants and other Internet of Things (IoT) phenomena explode, bringing along with it the attention of cybercriminals, who, with a little tinkering, quickly realized they could penetrate defenses with minor effort, as most of these devices were being shipped without privacy or security built into the design. Rewind to 2005 and ask yourself: "Could you imagine your baby monitor being used in a botnet?" It's not such a far stretch to imagine AI being tampered with as well.

When artificial intelligence was just a theory, science fiction writers and even many technologists warned of the far-reaching repercussions of AI gone awry. First, we humans would become dependent and placid. Then, the robots would revolt and end civilization as we know it. While there's no discounting the possibility, the reality of the potential abuse of AI—at least in the near future—is far more rudimentary.

That doesn't make it any less disruptive. When it comes to imagining how artificial intelligence and machine learning (ML) could be used for nefarious purposes, it helps to look to the recent past. Over the last 20 years, we've experienced massive disruption with the adoption of touch technology, social media, smart phones—even MP3s. In all cases, we've seen blissful early adoption followed by examples of abuse.

Will AI be a disruptive tech, then, in both the good and bad sense? The answer: a definitive yes. AI has already transformed from "new kid on the block" to a widely-applied science, although in some respects, it is still used as a buzzword to sell technologies, without a true understanding of how it's being incorporated into platforms.

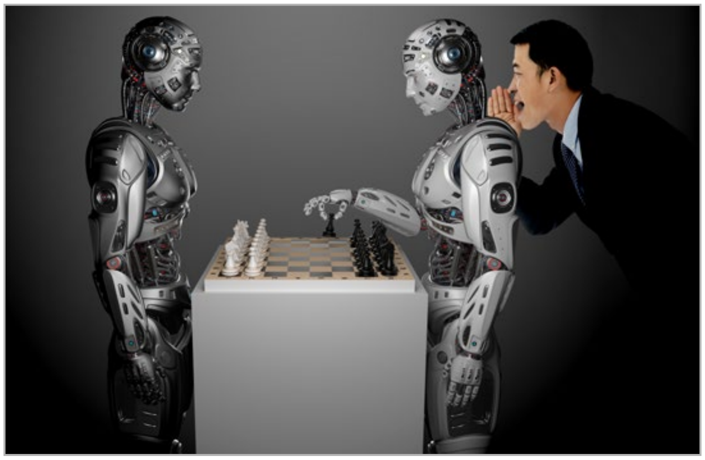
So let's take a look at exactly how AI and ML are being used in technology today, their benefits and our concerns, as well as what we believe are true possibilities for abuse in the near future so that developers, security professionals, and other organizations can incorporate AI responsibly and guard against potential attacks.

We don't see this happening anytime soon.

What is AI and machine learning?

To define artificial intelligence in its simplest terms, it is the science and engineering of making intelligent machines; the capability of a machine to imitate intelligent human behavior. Artificial intelligence is achieved when machines carry out tasks that are not pre-programmed, and in a way that we consider "smart."

Take, for example, a computer that can play chess. There is a significant difference between a chess computer that has been pre-programmed with countless situations and performs a given solution, and a chess computer that analyzes the position of the pieces and calculates the outcome for every possible move, many moves ahead. The first is executing commands. The second is using AI.



AI plays chess autonomously (left) while ML uses moves pre-programmed by humans (right).

Machine learning is a subset of AI that uses training to recognize patterns. When fed enough information, the machine learning algorithm is capable of recognizing patterns in new data and learning to classify that data based on the information it already has. Essentially, these algorithms teach the machine how to learn. However, an apparent danger of this method is that if the machine is allowed to accept its own assumptions to be true, it may stray from the path its developers envisioned.

Deep learning, meanwhile, is a method of machine learning that uses neural networks—algorithms modeled on the human brain—to learn from great amounts of data. "Deep" refers to the number of layers between the input and the output of the neural network that enables learning. Computers in the past could only handle one layer; modern computers can handle several. Similar to the way we learn from our mistakes, deep learning performs a task over and over, making slight changes to improve the outcome.

To sum it up: AI focuses on building machines that mimic human intelligence, including the ability to translate different languages, make decisions, recognize speech patterns, and perceive visual stimuli, while ML is about creating the algorithms that allow machines to learn from experience.

How are AI and machine learning used today in technology?

It's been a long time gone since AI was merely a theory researched by academics. Since then, many practical applications have been invented and evolved in many different directions. The technology, however, is still in its infancy, only recently being implemented into usable consumer platforms and devices. Even though it's been toyed with for decades, AI still has a long way to go. But in some fields, mostly dictated by need, there has been more progress than in others. These include:

- Autofill-as-a-service to improve the conversion rates on mobile devices
- Ranking search engine results to prioritize the sheer number of inputs for the user
- Cybersecurity detections to handle the huge growth of new malware samples
- Other data-filtering and analysis applications, for example, in marketing and large-scale planning
- Image and speech recognition, which are not yet perfect, but have much improved from just a few years ago

What these fields have in common is that they automate tasks that either require countless computations or that arise in such huge numbers that human beings alone simply cannot process them. Look for new AI developments in areas where similar problems are present, such as helpdesk requests (chatbots), online orders, or supply chain management.

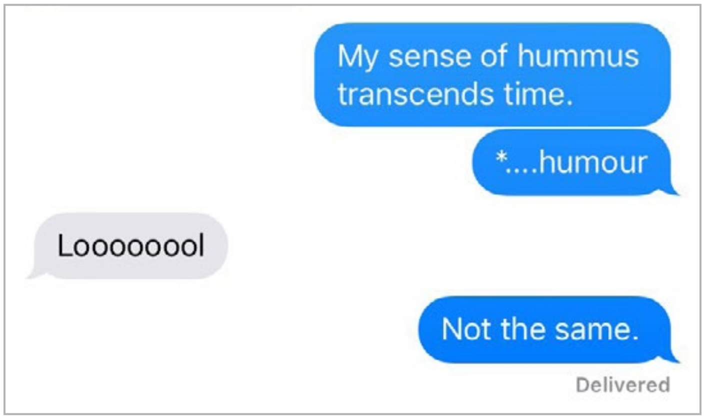
What are the advantages of using AI and ML in cybersecurity?

The cybersecurity industry—specifically its vendors—are in need of a technical solution to combat the growing number of new malware variants being deployed every day. With a well-known shortage of skilled IT workers and malware analysts, and changes in the threat landscape moving at breakneck speed, AI-enhanced technologies can step in and automate processes that might take humans much longer to complete. Using AI in cybersecurity solutions is advantageous not just in adding malware samples to detection engines, but in creating smart detections that can capture future versions of the same malware, or other variants in the same malware family. And while the AI evaluates, organizes, and condenses threat variants, automating mundane tasks at scale, it frees up researchers to focus on deeper analysis of more interesting, bleeding-edge threats.

AI-powered automation helps organizations battle threats with improved detection capabilities, allowing cybersecurity products to better identify, quarantine, and remediate malware and other threats without having to put a strain on IT resources. Malwarebytes uses machine learning in its Anomaly Detection engine to identify additional threats beyond those found by other layers of technology, such as anti-ransomware or malicious website blocking. Of the nearly 94 million non-PUP, non-adware detections logged from January through May 2019, 4.5 million were attributed to Anomaly Detection.

What are the concerns with AI and ML?

One of the first concerns with AI is its stability. We all know the problems with autocorrect on our mobile devices. (Many memes have been dedicated to these hilarious and embarrassing slip-ups). Is the technology stable enough to use in other applications?



My sense of hummus transcends time.

*....humour

Looooooool

Not the same.

Delivered

A relatively tame autocorrect fail, all things considered.

What raises concerns in the here and now is the use of unsupervised machine learning. For example, a Twitter bot based on unsupervised machine learning had to be taken offline rather quickly when it started imitating unbecoming behavior that it “learned” from other Twitter users. This was almost a perfect showcase of how easily machine learning can be corrupted when left without human supervision.

For now, while AI is mostly a beneficial addition to security solutions, incorrect implementation can result in less-than-optimal results. The use of AI and ML in detections requires a constant fine-tuning. Today’s AI lacks the depth of human knowledge needed to ignore benign files that don’t match the expected patterns. If the weave of the neural net is too wide, malware might escape detection; too fine, and the security solution will trigger false positives.

With rapid adoption of AI in technology—especially as cybersecurity organizations run to incorporate AI and ML into their security infrastructure—there also becomes an undeniable chance for cybercriminals to use the weaknesses in currently-adopted AI against security vendors and users. Once threat actors figure out what a security program is looking for, they can come up with clever solutions that help them avoid detection, keeping their own malicious files under the radar.

For example, malware authors could subvert AI-enhanced security platforms in order to trick detections into incorrectly identifying threats, damaging the vendor's reputation in the market. Threat actors could also dirty the sample for machine learning, flagging legitimate packages as malware, and training the platform to churn out false positives.

Another field where we are already seeing morally questionable use of AI and ML is in social engineering, particularly fake news. In a blog post on Malwarebytes Labs, we discussed the possible implications of DeepFakes, which is a method of creating fake videos of real people based on artificial intelligence. Creators feed a computer data consisting of a person's many facial expressions and find someone who can imitate that person's voice. The AI algorithm is then able to match the mouth and face to synchronize with spoken words. The end result is essentially a face-swap; splicing a person's head onto another person's body with near undetectable changes. While mostly used in pornography videos in underground forums, examples of fake news have already popped up using this technology.

Now imagine getting a video call from your boss telling you she needs you to wire cash to an account for a business trip that the company will later reimburse. DeepFakes could be used in incredibly convincing spear phishing attacks that users would be hard-pressed to identify as false. Already, impersonation attacks are on the rise, according to security firm Mimecast's annual report. About two-thirds of businesses saw an increase in impersonations in the last 12 months, and of those who received attacks, 73 percent suffered a direct loss.

There are other ways in which threat actors can utilize AI without building it themselves. They include:

» Captcha solving, which is already trivial for machine learning.

» Social media scanning with AI: looking for people associated with organizations, which helps in getting intel for more effective spear phishing campaigns.

» Creating more convincing spam, as it could be trained to adapt to the receiver. When spam is generated in your language, aimed at your line of business, and appears to come from someone you know and respect, the likeliness of you opening the attachment is greater.

Almost by definition, cybercriminals are opportunistic. If they don't need to develop sophisticated attacks to ensnare their victims, they will not. Instead, they often rely on tried and true methods, such as sending spam emails with malicious attachments, which require little technical prowess but are nonetheless effective. However, you only need one smart cybercriminal to develop malicious AI in an attack for this method to catch on.

> *"I can envision attacks where some kind of AI or machine learning would take place. For example, fuzzing for vulnerabilities. A lot of legitimate companies already do that for auditing purposes, but it could certainly be used for nefarious purposes as well."*
>
> *Jérôme Segura, Head of Threat Intelligence, Malwarebytes*

If AI-enabled cyberattacks open up a new avenue for profit, rest assured threat actors will be standing in line to buy kits on the dark market or use GitHub open source to adapt it to their own needs.

Malicious AI in malware

There are currently no examples of AI-enabled malware in the wild. However, some realistic possibilities include:

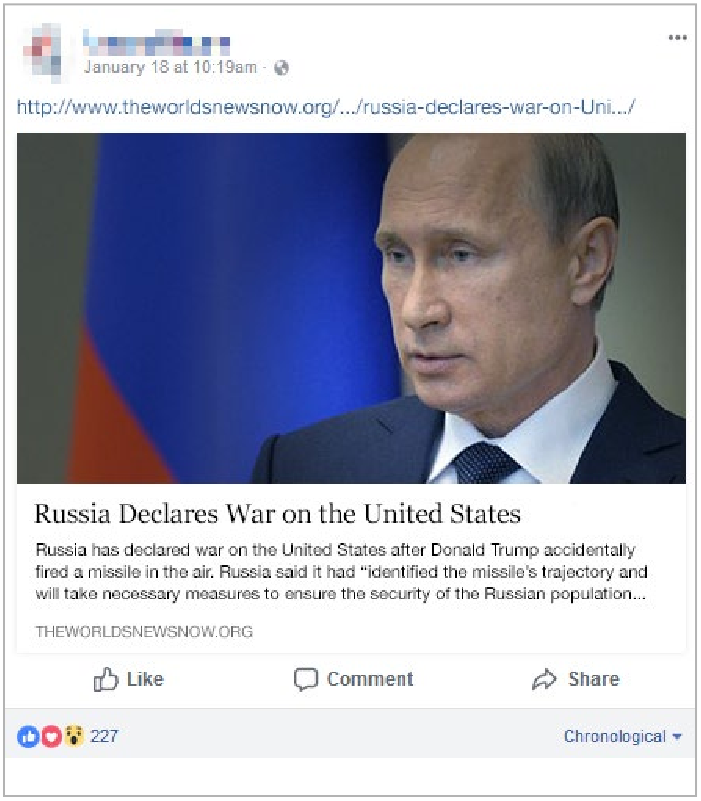
Worms
Imagine worms that are capable of avoiding detection by learning from each detection event. If such a family of worms is able to figure out what got them detected, they will avoid that behavior or characteristic in the next infection attempt. For example, if its code gave it away, worm authors could change the code, or if its behavior is flagged, they could add randomness to foil pattern-matching rules. An active worm with lateral movement can roam the networks of this planet for years, as WannaCry is teaching us, using unpatched endpoints as a basis for entry, and replicating itself via exploit, emails, sharing links to contacts via social media, and other methods of propagation.

Trojans
There are already Trojan malware variants, such as Swizzor, that create new file versions of themselves to fool detection routines. Improving this method by using AI may not be as far-fetched as one might think.

DeepLocker
As a case study, IBM Research developed an attack tool called DeepLocker powered by artificial intelligence. DeepLocker was presented at Black Hat USA 2018, built to better understand how some already-existing AI models could be combined with contemporary malware techniques to form a new breed of malware. As IBM security experts Jiyong Jang and Dhilung Kirat explained, “DeepLocker is designed to be stealthy. It flies under the radar, avoiding detection until the precise moment it recognizes a specific target.” Masquerading as video conferencing software, DeepLocker waits patiently until it reaches a system where a given condition is met (the target), and then it deploys its malicious payload. This makes the malicious code hard to find and almost impossible to reverse engineer. Malware designed with these specifications could infect many machines without being detected, and then be deployed on target machines according to the threat actor’s command.

Which problems can we foresee if AI gets implemented in malware?

If and when AI is used with malicious intent, there will be some alarming consequences. Because AI-enabled malware would be better equipped to familiarize itself with its environment before it strikes, we could expect harder-to-detect malware, more precisely-targeted threats, more convincing phishing, more destructive, network-wide malware infections, more effective methods of propagation, more convincing fake news and clickbait, and more cross-platform malware.



[redacted]
January 18 at 10:19am ·

http://www.theworldsnewsnow.org/.../russia-declares-war-on-Uni.../

Russia Declares War on the United States

Russia has declared war on the United States after Donald Trump accidentally fired a missile in the air. Russia said it had “identified the missile’s trajectory and will take necessary measures to ensure the security of the Russian population...

THEWORLDSNEWSNOW.ORG

An OpenAI algorithm created this fake story after being fed a single line of content. Here’s what it would look like in a Facebook feed.

Some methods malware could deploy to hinder detection by cybersecurity vendors include changing behavior and characteristics based on its environment, deleting itself when it suspects it’s being analyzed, changing shape and form along the way, and deploying malicious activities only on specified systems. In addition, malware authors could train security solutions that rely on machine learning to leave certain malicious files alone or, as previously stated, generate a large number of false positives.

The combined use of AI and Big Data could quickly annihilate what little privacy we have left. The operators would know details about their targets that haven't been unearthed for years, or that even remain secret to the targets themselves. Imagine this power in the hands of cybercriminals, and you can quickly envision convincing spear-phishing campaigns that can hardly be recognized as malicious. Even the most experienced users might fall for such personalized attacks. We have already seen the use of exposed personal data being used in sextortion scams to fool users into paying large sums to the "hacker" who supposedly has video proof of salacious deeds. In addition, scams conducted via WhatsApp or Messenger mimic the behavior and wording of relatives so closely that they defraud users into wiring large amounts of money, thinking they are helping.

There are many other possible scenarios, but the ones mentioned above are techniques that are already in use and could be easily improved upon or deployed on a larger scale by using AI.

Preventative measures for organizations and consumers

Our advice: be proactive. It is important for consumers to understand the threat landscape, and even more so for those charged with an organization's security—whether of its employees or its own product. We can use this knowledge to prepare ourselves and our organizations for a near future of AI-enabled malware. As long as we're still only speculating on the exact form and extent of future threats, however, it will be difficult to take specific measures. But we are seeing some forward-thinking developments by governments and cybersecurity organizations.

Governments

One difference between AI's introduction to the market from the Internet of Things debacle is that politicians are getting involved this time, announcing rules and regulations before the first disaster strikes. Does that make us prepared for the possible implications of AI-weaponization? Even though AI is a global development and, should it arise, a global threat, most governments (or other bodies for that matter) are not yet prepared for the possible implications of AI-weaponization. Only a couple of countries, including the United States, have some sort of AI strategy.

You can find information about different nations' strategic plans for artificial intelligence in the report Toward AI Security; Global aspirations for a more resilient future by the Center for Long-Term Cybersecurity (CLTC). The most compelling part of its conclusion:

> *"Although AI technology has been around for decades, governments have only begun paying increased attention to this technology in recent years, as the interactions between AI technologies and political, social, and economic systems have increased in scale, scope, and impact."*

A presidential executive order on maintaining American leadership in AI was issued by President Trump on February 11, 2019, to promote AI research and development. The order stipulates that advancement in AI must be conducted while protecting the safety and privacy of not only the American people, but citizens around the world. One of its objectives is to:

> *"Ensure that technical standards minimize vulnerability to attacks from malicious actors and reflect Federal priorities for innovation, public trust, and public confidence in systems that use AI technologies; and develop international standards to promote and protect those priorities."*

However, there is no clear path forward outlined in the order to show how to achieve this objective. With the United States looking to lead the way, a more robust plan for developing standards and regulation must be provided to thwart attacks on trusted systems.

Cybersecurity

Cybersecurity vendors should develop AI and machine learning-capable technologies with their own security in mind, considering the possible implications of cybercriminals attempting to use the technology against security programs. Closing any loopholes, especially for training systems to correctly identify threats, should be a top priority. But protecting the security program alone isn't enough. The technology should also not open up new attack vectors that could potentially be used against customers, and it should be well-tested before being implemented.

Organizations should look to vendors who aren't burying their heads in the sand when it comes to AI—both its benefits and potential for negative consequences. Which companies are using AI? How are they using it? Do they have plans to protect it from abuse? Users should favor organizations that are implementing the shiny new tech with deliberate consideration of its widespread impact and how it aides in strengthening security, not serving as a loophole through which criminals can gain access.

Timeline for the development of AI-powered malware

Based on experience, we expect to see AI implemented or used against itself for malicious purposes in the next 1–3 years, but in minimal ways. As developments in other fields progress, chances may open up for cybercriminals to abuse new AI technology. For the moment, we haven't seen a fully-automated security strategy that would be able to overpower AI-driven malware, so to at least be on an equal playing field, we need to get to work. Our advantage over AI continues to be the sophistication of human thought patterns and creativity; therefore, human-powered intelligence paired with AI and other technologies will still win out over systems or attacks that rely on AI alone.

AI + human intelligence = friends forever

Fast-forward 10 years, however, and if we're not proactive, we may be left in the dust. Best to develop this technology responsibly, with a 360-degree view on how it can be used for both good and evil, then to let it steamroll over us and move beyond our reach.

Contributors

Pieter Arntz // lead intelligence reporter, Malwarebytes Labs

Wendy Zamora // editor-in-chief, Malwarebytes Labs

Jérôme Segura // head of threat intelligence, Malwarebytes

Adam Kujawa // director of Malwarebytes Labs

blog.malwarebytes.com corporate-sales@malwarebytes.com 1.800.520.2796

Malwarebytes proactively protects people and businesses against dangerous threats such as malware, ransomware and exploits that escape detection by traditional antivirus solutions. Malwarebytes completely replaces antivirus with artificial intelligence-powered technology that stops cyberattacks before they can compromise home computers and business endpoints. Learn more at www.malwarebytes.com.