# Artificial Intelligence:
## Implications for IoT Security

by Lars Reger (SVP & CTO NXP Automotive) & Fariborz Assaderaghi (SVP & CTO NXP IoT and Security)

If there is one true indicator to measure the disruptiveness of a new technology, it's certainly the public outpouring of fear and suspicion. If we use societal angst as a measure, the current renaissance of artificial intelligence (AI) is a good candidate for groundbreaking technological disruption. As NXP we're looking beyond today's IoT, toward a future where smart connected devices not only talk with each other but where they use AI to interact with each other on our behalf.

By classic definition, artificial intelligence is a rather unspectacular affair. In his groundbreaking 1976 paper Artificial Intelligence: A Personal View, British neuroscientist and AI pioneer David Marr states: The goal of AI is to identify and to solve useful information processing problems and to give an abstract account of how to solve it, which is called a method.

It's true that AI computing systems are vaguely inspired by the biological neural networks that constitute brains. However, it is a popular myth that AI looks to re-engineer the function of the human brain to enable machines to solve problems the way humans would. The artificial neural network (ANN) rather is a framework for many different machine learning algorithms to work together and process complex data inputs. The main deviation from biology is that ANN are focusing to perform specific tasks rather than universal problem solving and planning capabilities.

As a subset of artificial intelligence, machine learning (ML) uses statistical techniques to give computers the ability to learn without being explicitly programmed. ML algorithms operate by building a model from an example training set of input in order to make data-driven predictions expressed as outputs. To make use of AI in this challenging environment, an agile application that can retain learning and apply it quickly to new data is necessary. This capability is called inference: taking smaller chunks of real-world data and processing it according to training the program has done. The combined capabilities of AI, ML, and inference create a multitude of business opportunities. However, the massive new potential for digital business models does not come without certain risks. Let's look at how AI and ML-based technology can have an impact.

Charter of Trust

**ML: A double-edged sword**

Every second, five new malware variants are discovered. Organizations across the globe are hit by one hundred previously unknown malware attacks every hour. And every day, one million new malicious files appear in the connected world[1]. With ever more devices and systems connected to the web, Cybercrime has become an increasing threat to our technological assets – and to the safety of our society as a whole.

It's only a matter of time when hackers will rely on AI to extract secrets and critical information from secure systems, as it only enhances their "learning" capabilities. We must think about and check our defense mechanisms against these coming approaches.

Progress in artificial intelligence is closely related to the development of cyber threats. Machine learning proves to be a double-edged sword: While ML enables industry-grade malware detection programs to work more effectively, it will soon be used by the bad actors to enhance the offensive capabilities of their attacks. As a matter of fact, a group of researchers from the University of Amsterdam recently demonstrated how this can work. In their side-channel attack that leaked information out of the CPU's translation lookaside buffers (TLBs), the white-hat hackers used novel machine learning techniques to train their attack algorithm and bring it to a new level of performance. They are confident that machine learning techniques will improve the quality of future side-channel attacks.

In order to prevent the emergence of new, effective AI and ML techniques from changing the balance of power, we must focus on how to leverage artificial intelligence to improve system security and data privacy.

**ML can add to system security**

A good example of ML-based security is anomaly detection, where the system "detects" anomalous behavior or patterns in the data stream. This process has been routinely applied to SPAM and malware detection in the past, but machine learning can be expanded to look for more subtle and complex anomalous behavior in a system.

While monitoring and protecting from external threats is crucial for an effective system defense, few organizations are aware of inside threats. In a survey from Accenture in 2016, they found that two thirds of the surveyed organizations fell victim to data theft from inside the organization. In these instances, 91% reported that they did not have effective detection methods

[1]Mc Kinsey & Company: T-30, AI Wars, Return of the Hardware, June 7-8, Pebble Beach, CA.

for identifying this type of threat. Machine learning can significantly aid in the development of effective, real-time profiling and anomaly detection capabilities, to detect and neutralize user-based threats from within the system.

**Privacy-preserving machine learning**

It's easy to identify applications in which the data providers for AI, either in the training phase or in the inference phase do not want to provide their data unprotected. With the new EU General Data Protection Regulation (GDPR) in effect since May 25, 2018, privacy protection is mandatory for any business dealing with the data of EU citizens, and non-compliance can result in heavy fines.

In medical and financial applications, for example, businesses are held accountable for the privacy of users contributing to the dataset. A typical use case is the training of a diagnosis model from a patient's medical records. A related threat comes when the machine learning model is made publicly available, e.g., when hospitals perform diagnoses in the previous use-case. A malicious user having access to the model might be able to analyze its parameters and to recover some of the data used to train the model.

There are also applications in industrial environments where data privacy is crucial to system providers. For example, in predictive maintenance machine data is used to determine the condition of in-service equipment to precisely predict when maintenance should be performed. This approach achieves substantial cost savings over routine or time-based preventive maintenance because tasks are performed only when required and hopefully in advance of system failure. Machine owners participating in the service have a clear intent to benefit from the generated data, however, they also have a strong interest in not sharing their data with competitors that use the same machines. This puts the maintenance service provider in a dilemma. The key question is: How can businesses continue to respect privacy concerns while still permitting the use of big data to drive business value?

This has led to the general research area of homomorphic encryption, which is referred to as a privacy enhancing technology that encrypts data into computable cipher text. Any data being used in the computation remains in the encrypted form, and only becomes visible to the intended user. The result of the computation – once decrypted – matches the result of the same computation applied to the plain text. In a machine learning context, companies looking to feed data into an externally provided, cloud-based machine learning model can use homomorphic encryption to avoid giving access to unencrypted data while still allowing complex computations to be applied to their own data.

Attribute-based cryptography is another privacy preserving technique that enables machine learning programs to run in compliance with strict data protection and privacy regulation. Attribute-based authentication is based on the Identity Mixer protocol[2] developed at IBM® Research and allows for strong authentication and privacy at the same time. It relies on a combination of flexible public keys (pseudonyms) and flexible credentials that allows a user to share only the information required for a certain transaction, without revealing any other attributes. In addition, this enables external parties to create a profile of the user based only on his pseudonym.

The advantages are obvious: "The Internet is like the lunar surface — it never forgets a footprint. With Identity Mixer, we can turn it into a sandy beach that regularly washes everything away," says Jan Camenisch, cryptographer and co-inventor of Identity Mixer.

**Attacking machine learning: attacks at training time**

What happens when an attacker goes after the security of machine learning itself? Before presenting potential threats in this arena, let's briefly recap how machine learning works. All machine learning starts with training data. The output is a set of parameters, essentially a model. In a second phase (inference), when given a new sample, the model infers the corresponding output. For instance, if the machine learning algorithm is an image classifier, one inputs a new image, the model returns its category (for example, that the image represents a cat). All the steps of this process, from training to inference, may be subject to attacks.

Attacks can occur even when the training data are collected and being fed into the ML model. While stealing data can be one objective of an attacker, changing the data or manipulating the outcome of the ML model may be another. For an AI model to make predictions that are in accordance with the physical reality, it is of utmost importance that the training data can be trusted. This property is sometimes difficult to achieve. A typical application is an anomaly detection tool trained from data sent by users. If a user "poisons" the training data by purposely sending incorrect inputs, this may resultin inferior performance or even failure of the machine learning model at inference time.

**Attacks at inference time: adversarial examples**

The user's privacy must also be guaranteed during the inference phase. This is especially relevant when inference is done on private or sensitive data. The user can also be the attacker. As a means of attack, the user may employ so called adversarial examples. An adversarial example is a valid input data that

[2] Identity Mixer (IDEMIX) is an anonymous, attribute-based credential system developed at IBM Research that enables strong authentication and privacy at the same time.

will cause the machine learning model to misinterpret it. This seemingly benign attack can create catastrophic consequences, for instance, if we think of road sign classification in safety critical situations.

By attaching a specially crafted sticker on a stop sign, researchers have shown that they can trick the image classifier into misinterpreting or not recognizing the sign at all. While the sign appears as regular stop signs to the human eye, the machine learning model is unable to see it as such. The concept of adversarial examples is not new. What is new is the severity of their consequences, like crashing an autonomous car in the stop sign adversarial example.

**IP Protection**

The value of machine learning models mostly resides in the associated data-sets. Training data can be very expensive to collect or difficult to obtain. When machine learning is offered as a service, the user has access only to the inputs and the outputs of the model. For example, in the case of an image classifier, a user submits an image and gets in return its category. It might be tempting for a user to make a copy of the model itself to avoid pay_ing for future usage. A possible attack is to query the service on chosen input data, obtain the corresponding output, and train the so-obtained data-set to get a functionally equivalent model.

The list of the presented attacks is certainly not exhaustive. Attacks can be combined to create even more damage. For example, once a model has been stolen, it can be used to try to recover training data or to craft adversarial examples. To ready ourselves against these evolving threats, the implications of AI must become an integral part of IoT system security. If training and inference of AI machine learning models are not to be become a wide-open gate for future adversaries, security by design and privacy by design principles must be considered from the beginning. Fortunately, it's not too late for the AI realm to apply the lessons already learned from IoT security.

**A Glimpse into the future: Artificial intelligence of things**

By designing things with smart properties and connecting them into the Internet of Things, we have created a global web of assets that have enhanced our lives and made them easier and more secure. The IoT gave us eye and ears, and even hands, to reach out from the edge of the network into the physical reality where we gather raw information, which we stream to the cloud, where it's processed into something of superior value: applicable knowledge. By adding high-performance processing, we've started to process and analyze information less often in data centers and the cloud, and more

now at the edge, where we see the magic occur. We witness that magic in smart traffic infrastructure, in smart supply chain factories, on mobile devices, in front-end stores, and in real time, where all the action takes place that makes our lives colorful.

The IoT in its present shape has equipped us with unprecedented opportunities to enrich our lives. Yet, it is only a stopover on the way to something even bigger and more impactful. We are talking about the artificial intelligence of things. Today's smart objects, even though they stream data, learn our preferences and can be controlled via apps, they are not AI devices. They 'talk' to each other, yet they don't play together. A smart container that monitors the cold chain of a supply of vaccines is not an AI system unless it does 'something', such as making a prediction about the temperature development in the container and automatically adjusts the cooling.

An autonomous car or a search-and-rescue drone that autonomously navigates off-shore is in fact an AI system. If it drives or flies on behalf of you, you can trust that some serious AI capabilities are involved. Reading, speaking or translating language, predicting the mass and speed of an object, buying stock on your behalf, recognizing faces or diagnosing breast cancer, are all artificially intelligent characteristics when done by an algorithm.

Now, imagine a world in which the entirety of AI things was connected. Expanding the edge of the IoT with cognitive functions such as learning, problem-solving and decision making would turn today's smart things from mere practical tools into true extensions of ourselves, multiplying our possibilities to interact with the physical world.

As an integral part of the IoT, artificial intelligence is the foundation for entirely new use cases and services. Siemens®, for example, is using AI to improve the operation of gas turbines. By learning from operating data, the system can significantly reduce the emission of toxic nitrogen oxides while increasing the performance and service life of the turbine. Siemens is also using AI systems to autonomously adjust the blade angle of downstream wind turbines to increase the plant's yield. In medical care, Thomas Jefferson University Hospital in Philadelphia seeks to improve patient experience with natural language processing that will enable patients to control room environment and request various information with voice commands. And Rolls-Royce® is developing an IoT-enabled airplane engine maintenance service that uses machine learning to help it spot patterns and identify operational insights that will be sold to airlines[3].

[3] See above, Schatsky et al.

On the consumer side, Google's Duplex offers an idea of what the future holds: A virtual assistant that can carry out "real world" tasks over the phone, performing functions like scheduling a dentist appointment or making a dinner reservation. These are tasks that typically require human interaction on both ends, but not anymore. While far from a natural understanding of the interaction, Duplex's AI voice sounds so natural that the person taking the call could be unaware that they're chatting with a machine.

Now, what does this mean for the future? The truth is, even with a broad range of nascent AIoT applications emerging, we can't even fathom what else is coming. One thing is for sure - Today's digital age society is undergoing a fundamental change. The paradigm shift that comes with the convergence of AI and the IoT, will be even greater than the one we have witnessed with the introduction of the personal computer or the mobile phone. Effective security, based on the guiding principles of security and privacy by design, will be crucial to mitigate against the risks that come with it. People, organizations and entire societies will support this transformation only if the security of their data and networked systems can be ensured. Therefore, it is essential that global players team up, join forces and work together on equal footing in industry, government and society to create an industrial cybersecurity network to instantly share new insights and information about attacks and incidents. And this is exactly where the Charter of Trust (CoT) comes in.

To keep pace with the continuous progress in the digital economy and the threats posed by criminal activities, Siemens works with partners from industry, government and society to sign a "Charter of Trust" to define and implement principles that can make both a difference and the digital world a safer place. Among them are AES, Airbus, Allianz, AtoS, Cisco, Daimler, Dell, Enel, IBM, NXP, SGS, Deutsche Telekom, Total and TÜV Süd.

The CoT focuses on three goals: protecting the data of individuals and companies; preventing harm to people, companies and infrastructures; and establishing a reliable foundation on which confidence in a networked digital world can take root and grow. As pioneers in digitalization, the Charter of Trust is taking a stand in favor of binding rules and standards that will create a new basis of trust and equality of competition. Only when we become active together, intensify the cooperation between companies and policymakers and create a common understanding of cyber threats, we will succeed in the long run and achieve our goals. The Artificial Intelligence of Things (AIoT) holds the power to transform our lives. It's upon us, to turn the black box of the future into a bright one.