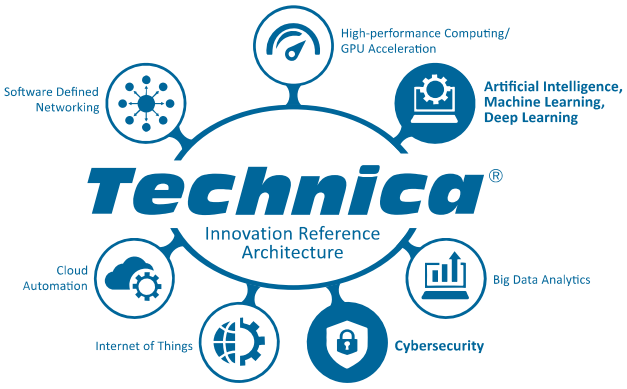
Technica®

This paper introduces deep learning and then looks to anticipated use cases that fit within a sweet-spot for Technica: cybersecurity. Baidu—the Google of China, with 92% of search users in China—claims that deep learning improved its antivirus filters. While promising, doing deep learning for an organization is still very complex. It is a bit like manually coding your own TCP/IP stack to do networking.

The Technica Innovation Platform White Paper Series presents advanced topics that will drive competitive advantage for next-generation IT over the next three-to-five years.



DEEP LEARNING FOR CYBERSECURITY USE CASES

INTRODUCTION

The form of machine learning known as “deep learning” is one of the hottest trends in technology. Google, Facebook, IBM, Twitter, Baidu, and Salesforce have all recently initiated deep learning projects.

On June 16, 2016 Google announced that it was opening a dedicated Machine Learning group in its Zurich engineering office, the largest collection of Google developers outside of the US. The group is tasked to lead research into three areas: machine intelligence, natural language processing, and machine perception. In other words, the Google developers will work on building systems that can think, listen, and see.

Google’s endeavor follows a Facebook announcement in 2015 that it was opening a “Deep Learning Lab” in Paris, France. The acquisition landscape for deep learning is equally hot. In February 2015, IBM acquired AlchemyAPI, Google acquired DeepMind, Twitter acquired Madbits, and Facebook acquired wit.ai. Salesforce has acquired three deep learning startups in 2016: MetaMind, PredictionIO, and TempoAI.

The DeepMind acquisition by Google was the most ambitious. The DeepMind founder listed the purpose of the company as, “Attempting to distil intelligence into an algorithmic construct may prove to be the best path to understanding some of the enduring mysteries of our minds.” Thus far, Google has used DeepMind to hone their offerings for visual recognition and response, in particular teaching a computer to play simple video games. Most recently, the DeepMind research arm of Google created a deep learning-based algorithm that

led to a victory over a world champion in the ancient Asian board game known as "Go." Go is orders of magnitude more complex than chess.

Where the archetypal startup of 2008 was "x, but on a phone," and the startup of 2014 was "Uber, but for x"; this year is the year of "doing x with machine learning."

This paper introduces deep learning and anticipated use cases that fit within a sweet-spot for Technica: cybersecurity. Baidu—the Google of China, with 92% of search users in China—claims that deep learning improved its antivirus filters. While promising, implementing deep learning within an organization is still very complex. Currently, it is a bit like manually coding your own TCP/IP stack to perform networking.

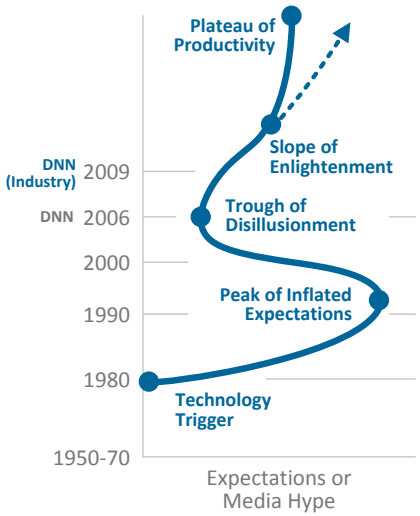


Figure 1 – Gartner Hype Chart for Neural Networks

Artificial Intelligence is a broad field that deals with:

- Reasoning
- Knowledge Representation
- Automated Planning and Scheduling
- Machine Learning
- Natural Language Processing (NLP)
- Computer Vision
- Robotics
- General Intelligence, or Strong AI (Computers equal the performance of humans)

DEEP LEARNING IN A NUTSHELL

Deep learning is part of a broader subset of technologies known as machine learning. Machine learning is a component of artificial intelligence (AI). None of these concepts—deep learning, machine learning, or AI—are new. They are as old as computers themselves.

Artificial Intelligence (AI)

AI was coined as a term in 1955, by John McCarthy who defined AI as the "science and engineering of making intelligent machines." Neural networks and machine learning algorithms were created in the late 1950s. Neural networks are the key concept within deep learning. As will be shown, neural networks are the 'deep' in deep learning.

Figure 1 provides a Gartner Hype cycle chart for neural networks. The chart is instructive because of the length of time it incorporates, with the hype reaching a peak in the 1990s. The upsurge in neural network activity is a clear indicator that deep learning is not a passing fad.

As a subset of machine learning, deep learning is affecting many of these AI fields, beyond image recognition and computer vision. For example, IBM's AlchemyAPI touched upon knowledge representation with its categorization of taxonomies; and NLP with its keyword extraction.

Deep Learning

With a complex topic such as deep learning, it may be helpful to review several definitions:

Definition 1: A class of machine learning techniques that exploit many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification.

Definition 2: Deep learning is a set of algorithms in machine learning that attempt to learn in multiple levels, corresponding to different levels of abstraction. It typically uses artificial neural networks. The levels in these learned statistical models correspond to distinct levels of concepts, where higher-level

concepts are defined from lower-level ones, and the same lower-level concepts can help to define many higher-level concepts.

Definition 3: Deep Learning is a new area of Machine Learning research, which has been introduced with the objective of moving Machine Learning closer to one of its original goals: Artificial Intelligence. Deep Learning is about learning multiple levels of representation and abstraction that help to make sense of data such as images, sound, and text.

The key concept within deep learning is the notion of "layers" or "levels," where each layer recognizes progressively more complex features and learns from the previous layer. **Figure 2** portrays a neural network composed of numerous layers to perform facial recognition.

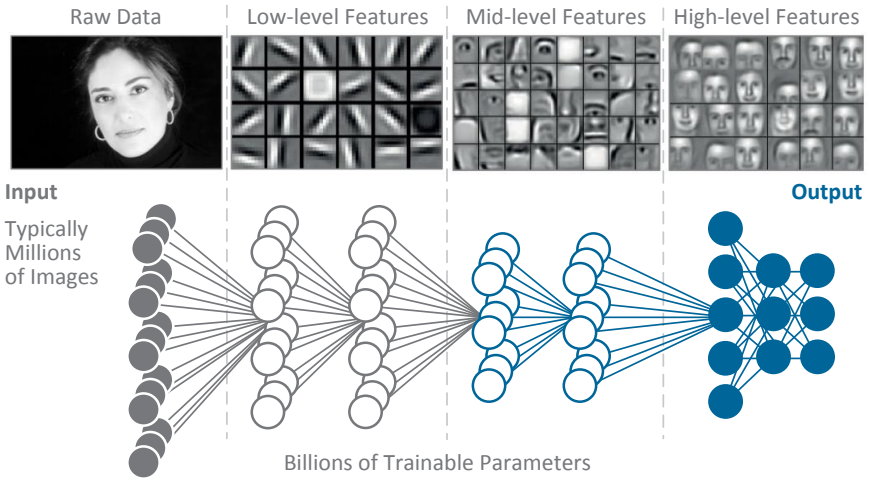


Figure 2 - Deep Neural Network (DNN)

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual competition to motivate the development of innovative algorithms for object detection and image classification. Each year the task is to take 1.2M training images with 1,000 object categories, i.e. tags, or classifications, and see how well the computer can recognize the images. In 2012, the first deep neural network solutions employing GPU acceleration entered the contest. The winner of that contest used a deep neural network, i.e. deep learning, and improved image recognition by a phenomenal 16%.

By 2014, every entrant used deep learning and a majority used GPU acceleration. Google was the winner of that contest and had reduced errors to under 5%. This is an important milestone because the 5% error rate is on-par with performance achieved by human classifiers. That is to say, Google is able to recognize images as fast as a human, i.e., Google is performing strong AI for image recognition utilizing GPU accelerated deep learning. This was performed with a neural network ten layers deep. The reason that GPUs and deep learning work so well together is discussed in a following section.

"A Person Riding a Motorcycle on a Dirt Road"

"A Yellow School Bus Parked in a Car Park"

Figure 3 – Google Image Recognition Software

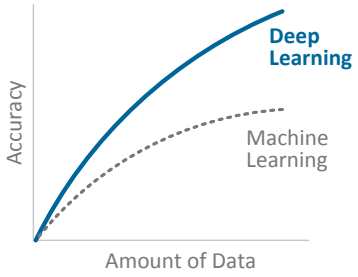
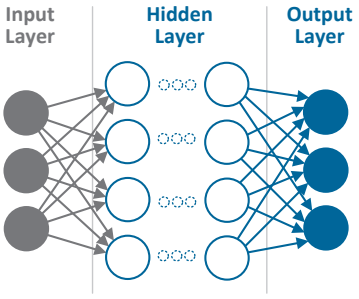
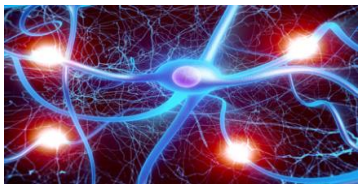


Figure 4 – Deep Learning Setting Records



Simple Neural Network



Biological Neuron

Figure 5 – Neural Network vs Biological Neuron

This impressive performance, where computers outperform human beings, led Bill Gates, Elon Musk, and Stephen Hawking to raise warning flags about AI. They all call for governance for AI before progressing much further.

Google is not only able to recognize images better than human beings, but also examines images and then automatically creates an English sentence describing the picture.

In the top picture of **Figure 3**, the computer recognized that a person was riding a motorcycle on a dirt road. The bottom picture of **Figure 3,** shows a mistake created by the deep learning algorithm in creating the textual image tagging. To fix this error, a typical solution would be to train the deep learning algorithm with more images of yellow school busses.

The amount of data available to the training set has a direct relationship to the quality of the algorithm, as portrayed in **Figure 4**.

This highlights one of the reasons that deep learning is hot right now. There is simply much more data, i.e. Big Data, that can be used to train the deep learning algorithms. The second factor is the increasing processing power of multi-core CPUs and GPUs. More specifically, GPU acceleration has drastically lowered the cost for deep learning. Finally, ideas have evolved as to how best train deep learning architecture, including those that have many layers relative to earlier efforts.

Andrew Ng is an associate professor in the Department of Computer Science and the Department of Electrical Engineering at Stanford University and Chief Scientist at Baidu Research in Silicon Valley. Ng recently stated, "I've worked all my life in machine learning, and I've never seen an algorithm knock over benchmarks like deep learning."

Often, deep learning is said to mimic the human brain. Along these lines, deep learning has the concept of neurons as the fundamental unit of the deep learning algorithm. It is important to note that these artificial neurons are very different from biological neurons. The human brain has 100 billion neurons and 100 trillion connections, called synapses that operates on 20 watts. Compare this to the 2012 Google brain project with 10 million neurons and a billion connections on 16,000 CPUs at about 3 million watts. **Figure 5** highlights this comparison.

The top of Figure 5 shows a traditional diagram used in describing deep learning. The circles represent individual neurons. The input layer is the data that will be processed by the neural network. The hidden layer, where the 'magic' occurs, is composed of three layers; therefore, the figure portrays a neural network that is three layers deep. The output layer is the result of the processing of the deep learning algorithm. In contrast, the bottom of Figure 5 is a picture of a single biological neuron.

TYPES OF NEURAL NETWORKS

The key to deep learning algorithms is the mathematics of the hidden layer. In other words, the design of the neural network impacts the types of problems the

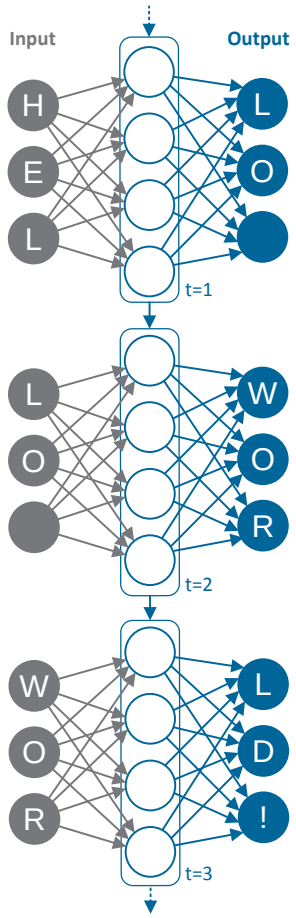


Figure 6 – Recurrent Neural Networks

neural network can address. There are a number of designs for neural networks. The two most successful designs for deep neural networks (DNNs) are convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

Convolutional Neural Networks (CNNs)

CNNs are inspired by the human visual cortex. CNNs are a feed-forward DNN in which individual neurons are tiled in a manner that they respond to overlapping regions in the visual field. CNNs are used to learn a hierarchy of visual features, and as such, are heavily used for image recognition. The "essence" of a visual object is learned through generalizations from successive exposure to similar images.

Recurrent Neural Networks (RNNs)

Neurons in an RNN form a directed cycle, a concept from graph theory. The neurons make numerous connections both forward and backward. This creates an internal state of the network that allows it to exhibit dynamic temporal behavior. See **Figure 6**.

Unlike CNNs, RNNs can use internal memory to process arbitrary sequences of inputs. RNNs have found success in NLP, handwriting recognition, and speech recognition. Recently, advances in RNNs have included incorporating "memory."

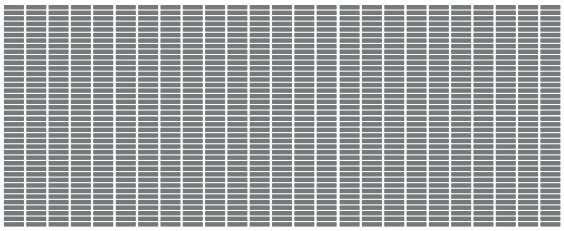
Long short-term memory (LSTM) RNNs are responsible for many of the advances in speech-recognition and NLP. Human thoughts have persistence; for example, each word you understand in the paper builds on an understanding of previous words. Similarly, LSTM allows the network to contain a configurable amount of persistence rather than continuously starting over from scratch.

DeepInsight

DeepInsight is Technica's deep graph analysis algorithm. It uses local information gathered from random walks on the graph to train a binary tree network. This network learns representations of all vertices in the graph. These representations contain the most important information about how the vertices relate to one and other in a condensed form. They can be used for a variety of tasks including community detection, label prediction and link prediction.

Deep Learning and GPU Acceleration

GPU acceleration is the term used to describe how an application is split between the CPU and GPU. **Figure 7** compares the configuration and cost to perform the same task using the Google Brain with 16,000 CPUs and no GPUs, and the Stanford AI Lab solution with GPU acceleration.



Google Datacenter
1000 CPU Servers
2000 CPUs, 16,000 Cores, 600 kWatts
$5,000,000

Stanford AI Lab
3 GPU–Accelerated Servers
12 GPUs, 18,432 Cores, 4 kWatts
$33,000

Figure 7 - GPUs Make Deep Learning Accessible

GPUs and deep learning go hand–in-hand. **Figure 8** comes from a slide prepared by NVIDIA. The essence of the slide is that the mathematics of neural networks are not particularly complex. The feed forward mechanics of the neural network are performed with matrix operations upon which GPUs excel. Finally, deep learning heavily uses floating point operations (FLOPS), another area in which GPUs excel.

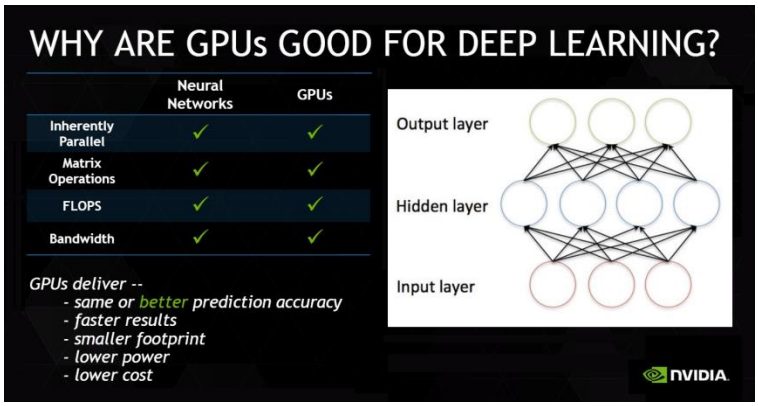


Figure 8 –Why are GPUs Good for Deep Learning? NVIDIA Slide

This means that GPUs meet or exceed the performance of CPU-only solutions at a fraction of the cost, power consumption, and datacenter space.

GENERAL DEEP LEARNING APPLICATIONS

The initial deep learning impact has been upon image recognition/object detection, speech recognition, and NLP.

However, *deep learning is a general purpose pattern recognition technique*.

As such, the applications for deep learning are boundless. Over the coming years, deep learning will impact multiple aspects of every industry. Any activity that has access to large amounts of data may derive benefits from deep learning. Currently deep learning shines with labeled data, i.e. in supervised learning. After enough exposure to labeled training set data, and fortified with the right algorithms, computers use annotated data to teach themselves to spot useful patterns, rules and categories within the data.

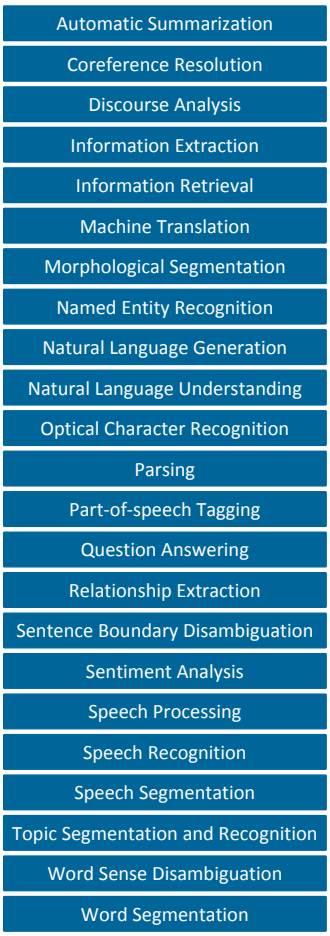


Figure 9 – Experienced Success

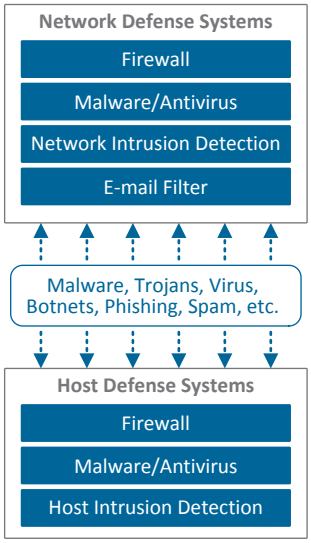


Figure 10 – Enterprise Cybersecurity System

Some progress is being made with unsupervised learning. For example, after watching millions of YouTube videos, the computer was able to categorize common objects including human faces.

Figure 9 lists a number of areas in which deep learning has experienced success.

A number of deep learning algorithms have been developed and trained within days and weeks that outperform specialty programs developed over years.

Most importantly, the deep learning algorithms were created by deep learning experts that did not have insight into the specific domain in question. For example, a deep learning algorithm designed to spot signatures of subatomic particles performed better than software written by physicists.

In a recent Kaggle data science competition, deep learning expert Geoffrey Hinton beat all drug discovery algorithms with deep learning algorithms his team designed in two weeks, yet no one on the team had a life sciences or medical background.

Traditional machine learning is generally composed of three steps: 1) collection of raw data, 2) architect feature extraction, 3) program classifier/ detectors. Domain specific experts work to improve feature extraction and classification, and life science/medical experts add insights to develop the best algorithm. In deep learning, however, the neural network combines these steps without human intervention. This has radical and profound implications for the future of computer programming.

CURRENT CYBER TRENDS

With an overview of deep learning behind us, let us know move to the topic of cybersecurity—more specifically current trends within the cyber landscape.

Figure 10 depicts the typical enterprise cybersecurity system.

Cyber-threats are met by network defense systems including firewall, malware software, and intrusion detection systems. Email protection typically occurs on network servers. Similar systems are deployed on the hosts within the system.

The table below presents the findings from a survey conducted by the Information Systems Audit and Control Association, Inc. (ISACA) and RSA Security, a division of EMC Corporation. Results of the survey were presented at the RSA Conference in the winter of 2016. The findings are instructive in detailing the frequency of specific categories of cyber-attacks.

	Daily %	Weekly %	Monthly %	Quarterly %
Phishing	**29.67**	**16.82**	**15.19**	**18.69**
Malicious Code	**16.36**	**12.38**	**12.85**	**26.40**
Physical Loss	**1.42**	**6.38**	**9.69**	**37.12**
Hacking	**11.06**	**7.29**	**9.18**	**25.18**
Online Identity Theft	4.08	4.56	5.52	20.62
Intellectual Property Loss	1.44	2.40	4.08	19.90
Intentional Damage	0.95	1.43	5.01	18.38

Denial of Service	4.05	5.48	9.76	27.38
Insider Damage	2.91	1.69	9.69	21.79
Don't Know	13.13	2.32	3.86	6.18
	8.40	6.30	8.70	22.80

This list gives a sense for the priorities of enterprises in thwarting cyber-attacks. According to the survey, the top four current threats are:

- Phishing
- Malicious Code (i.e. Malware)
- Physical Loss
- Hacking

Additional survey results identified items that RSA attendees wanted to discuss. The top four were:

- Internet of Things (IoT)
- Industrial Control Systems and the Industrial IoT
- Encryption
- AI and Machine Learning

IoT dominated the desired topics because the conventional cybersecurity system depicted in Figure 10 was not designed to deal with a multiplicity of hosts, especially hosts that reside outside the network perimeter.

Finally, survey respondents were asked about their expectations for AI in cybersecurity. **Figure 11** presents the results.

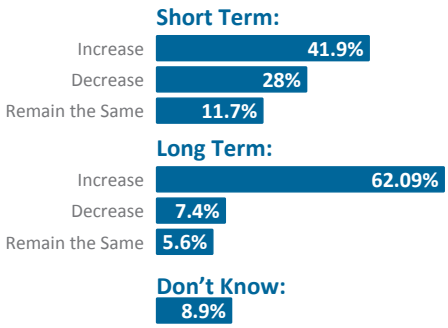


Figure 11 – AI and Cyber Risk

It is clear that the market expects that AI will massively impact cybersecurity in the coming years and is interested in AI's benefits in defense of the enterprise.

CYBERSECURITY USE CASES FOR MACHINE LEARNING/ DEEP LEARNING

Whereas cybersecurity has heretofore been about firewalls and antivirus software, coming advances will make cybersecurity smarter, more adaptive, and less human-intensive.

Figure 12 – Technica Innovation Platform

In fact, Technica believes that cybersecurity will be a key component driving innovation for next-generation IT, as highlighted in **Figure 12**. Working in coordination with other components, like AI, IoT, and GPU acceleration; Technica believes that cybersecurity will become a key differentiator of innovation for enterprise IT.

Given Technica's expertise with GPU acceleration, Big Data Analytics via FUNL and SQuadron, and knowledge of machine learning and deep learning, the company seeks to expand innovation within the enterprise by marrying cybersecurity with AI.

FUNL already implements a number of GPU accelerated machine learning algorithms including, Collaborative Filtering, Principal Component Analysis,

Support Vector Machine, and K-Means Clustering. A deep learning algorithm called DeepInsight is also currently supported.

The literature applying machine learning and deep learning to cybersecurity is sparse. A recent deep learning webinar for Defense—conducted by NVIDIA—mentioned cybersecurity briefly on a single slide, and all the use cases that were variations of object detection or speech recognition.

The current dearth of technical literature leaves areas within cybersecurity that are ripe for further investigation. Applications of DeepInsight to the relevant use case are articulated.

Phishing Detection

As previously discussed, phishing is the number one attack vector in terms of frequency faced by the enterprise today. Over 90% of all cyberattacks are initiated in this manner. Attackers carefully craft emails that are designed to trick recipients into providing access to their accounts. Typically, these emails mimic valid communications from a company or individual familiar to employees.

A successful deception can unravel an organization's entire cybersecurity framework. One compromised account can provide attackers with credentials to infiltrate a network, escalate privileges, gain access to core enterprise resource, and steal sensitive data.

Pattern recognition performed by machine/deep learning could mitigate many of these attacks. The use case would involve training on the corpus of existing enterprise emails. This analysis would yield relationships between internal and external domains, frequently contacted partners, and specific authentication patterns.

Trained by a large number of analyzed emails, real-time alerts could be rendered by the system to flag threats before they morph into data breaches.

DeepInsight can be used to analyze the structure of the web around links within emails and identify suspicious sites. In addition, DeepInsight's text analysis features can be used to identify patterns within emails to help distinguish between malicious and innocuous emails. The combination of the two provides a powerful tool for identifying phishing attempts.

Malware, Zero Day, and APT Detection

Current detection of malware and advanced persistent threats (APTs) rely on signatures, heuristics, sandboxing, and traditional machine learning. According to the AV-TEST Institute, 390,000 new strains of malware—zero day attacks—are produced every day. Symantec Corporation estimates that this number is closer to one million. These traditional approaches are not keeping up the volume and sophistication of the attacks.

In the same way that a deep learning expert beat traditional drug discovery machine learning algorithms, deep learning-based malware detection algorithms can more adeptly detect malware—even zero day attacks. In other words, deep

learning could be used to classify malicious code without an expert creating rules that explicitly define malicious code.

Once the neural network learns to identify malicious code, it can identify unknown files as nefarious or benign with extremely high accuracy—in real-time (unlike sandboxing). Not only would deep learning enable prediction, but prevention could also be incorporated. The moment a malicious file was detected, it could be quarantined.

Training the malware detection neural network would involve analysis of many millions of malicious and legitimate files for accurate classification. While not-trivial, the system would be simpler and more precise than gathering cybersecurity experts (for each file type) to extract features, as in traditional machine learning.

Once the neural network was trained, a small agent would be deployed on endpoint devices. The process would be similar to deploying the Siri app on iPhones. The agent could work with or without network connectivity. The malware recognition neural network could improve each day, and the agents would receive updates as needed when connectivity was present.

The goal of malware is to infect a system and evade detection by morphing itself. This trait is the main reason why pattern matching techniques fail. By creating a network of applications and file modifications, DeepInsight can be used to distinguish between malware and benign software.

Identification of Insider Threats

The website of the National Counterintelligence and Security Center says that "Over the past century, the most damaging U.S. counterintelligence failures were perpetrated by a trusted insider with ulterior motives."

Given a set of labeled examples, a deep learning based system could determine patterns for every network, device, and individual user. This data would be correlated to identify subtle discrepancies that indicate threats in progress. Behavior of users, key business systems, and applications could be modeled. Employees' demographic data and/or behavioral patterns could also extracted from access logs, traffic, etc. to indicate fraudulent activity.

The system would have the added benefit of identifying APTs, not just insiders. Server and firewall logs, Security Information and Event Management (SIEM) events, etc. would provide important sources of training material for the neural network.

Such a system could identify actions like uploading numerous files to Dropbox or other cloud-based services. Thus, the system could incorporate data loss prevention features.

A rapid change in the usage and communication pattern of an individual can be a sign of an upcoming insider attack. DeepInsight can be used to identify both these changes and how individual users relate to the rest of the user base. The latter

can be used to reduce false positives by learning about certain benign relationships as more data is gathered.

Advanced Intrusion Detection & Network Anomaly Detection

Current Intrusion Detection Systems (IDS) and firewalls work through a combination of misuse detection techniques or anomaly detection techniques. Heuristics created by experts guide the system to make the appropriate predictions. Be they deployed at the network or host level, deep learning based systems can be created to make better predictions of intrusions. Smart filtering could adaptively point nefarious actors to honeypots for further analysis.

The system could be coupled with packet analysis techniques—including deep packet inspection—to deter Distributed Denial of Service (DDoS) attacks, the presence of botnets and zombie attacks, and detection of nefarious scans.

DeepInsight can analyze network communications to find unusual behavior. It can establish a baseline for individual node behavior for comparison with future logs. It can also be used to determine the likelihood of future communications between nodes. If a communication occurs which has a low likelihood, it would register as anomalous.

SUMMARY

The deep learning sub-field of AI is gaining an ever increasing amount of attention. It holds the promise of attaining many of the long-standing goals for AI. This paper has examined deep learning in general and specifically considered use cases for cybersecurity. Over the coming years, advances in unsupervised deep learning (interpreting unlabeled data) will propel AI's success.

While the promise of deep learning is extraordinary, the current state of the art is still complex.

The barrier to entry for non-trivial applications of the approach is high. Technica is uniquely positioned to bring AI information and solutions to our DoD, IC, and Federal clients.

Technica envisions AI as a core component of the next-generation IT platform that allows for disruptive innovation. Other core components like HPC/ GPU acceleration can work in coordination with AI to propel cybersecurity effectiveness, especially as IoT comes on-line in the next decade.

Technica provides professional services, products, and innovative technology solutions to the Federal Government. We specialize in network operations and infrastructure; cyber defense and security; government application integration; systems engineering and training; and product research, deployment planning, and support.

Technica®
22970 Indian Creek Drive, Suite 500
Dulles, VA 20166
703.662.2000