What's new in security? February, 2020

- Iowa Caucus App woes

  "The software was the handiwork of Shadow Inc., a maker of digital campaign tools for Democrats that received more than $63,000 from the Iowa Democratic Party in November and December. On Tuesday, the company expressed "regret" about the confusion without disclosing what went wrong." Feb 4, 2020
  (https://www.politico.com/news/2020/02/04/iowa-app-caucuses-2020-election-110710)

  Apparently, issues with poor planning, lack of proper testing (i.e. technical issues), and inadequate training caused problems with the app not functioning properly. ***(my review)***

Oldie but a goodie …

Source: https://hackaday.com/2014/04/04/sql-injection-fools-speed-traps-and-clears-your-record/

Cyber Security and AI + ML

Amherst Security Group (@AmherstSec)

February 12, 2020

Robert Hurlbut

RobertHurlbut.com • @RobertHurlbut

Robert Hurlbut

Threat Modeling Architect, Trainer

Microsoft MVP – Developer Security 2005-2009, 2015-2020

(ISC)2 CSSLP 2014-2020

Co-host with Chris Romeo – Application Security Podcast

MeetUp Leader: Boston .NET Architecture Group, Amherst Security Group

Contacts

Web Site: https://roberthurlbut.com

Twitter: @RobertHurlbut, @AppSecPodcast

AI … what?

"Repeat any word enough times, and it eventually loses all meaning, disintegrating like soggy tissue into phonetic nothingness.

For many of us, the phrase "artificial intelligence" fell apart in this way a long time ago.

AI is everywhere in tech right now, said to be powering everything from your TV to your toothbrush, but never have the words themselves meant less."

Source: "The State of AI in 2019" by James Vincent (Jan 28, 2019)

https://www.theverge.com/2019/1/28/18197520/ai-artificial-intelligence-machine-learning-computational-science

Definitions

Artificial Intelligence (AI)

Google's definition:

Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart".

Definitions

Machine Learning (ML)

Google's definition:

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed.

Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.

Definitions (Contd.)

Data Analytics

Data Analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making.

Definitions (Contd.)

Cyber Security

Cyber Security is the body of technologies, processes and practices designed to protect networks, computers, programs and data from attack, damage or unauthorized access.

In a computing context, **security** includes both **cyber security** and physical **security**.

Definitions (Finally.)

AI + ML + Data Analytics + Cyber Security

Machine learning has been quickly adopted in cyber security for its potential to automate the detection and prevention of attacks.

One prominent example is next-generation antivirus (NGAV) products. ML models in NGAV have fundamental advantages compared to traditional AV, including higher likelihood of identifying zero-day attacks and targeted malware.

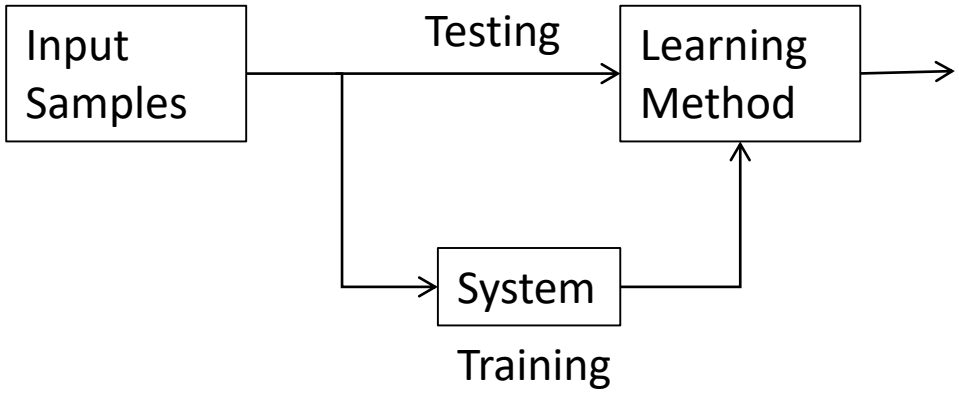
AI + ML

- As mentioned, **machine learning** is a branch of **artificial intelligence**, concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.

- As intelligence requires knowledge, it is necessary for the computers to acquire knowledge.

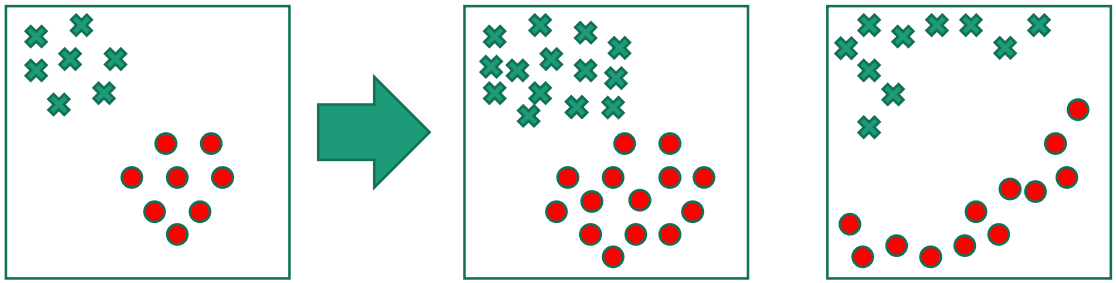
- That means Data – lots of Data

Learning system model



Training and testing

- Training is the process of making the system able to learn.
- No free lunch rule:
  - Training set and testing set come from the same distribution
  - Need to make some assumptions or bias



Performance

- There are several factors affecting the performance:
  - **Types of training** provided
  - The form and extent of any initial **background knowledge**
  - The **type of feedback** provided
  - The **learning algorithms** used
- Two important factors:
  - Modeling
  - Optimization

Algorithms

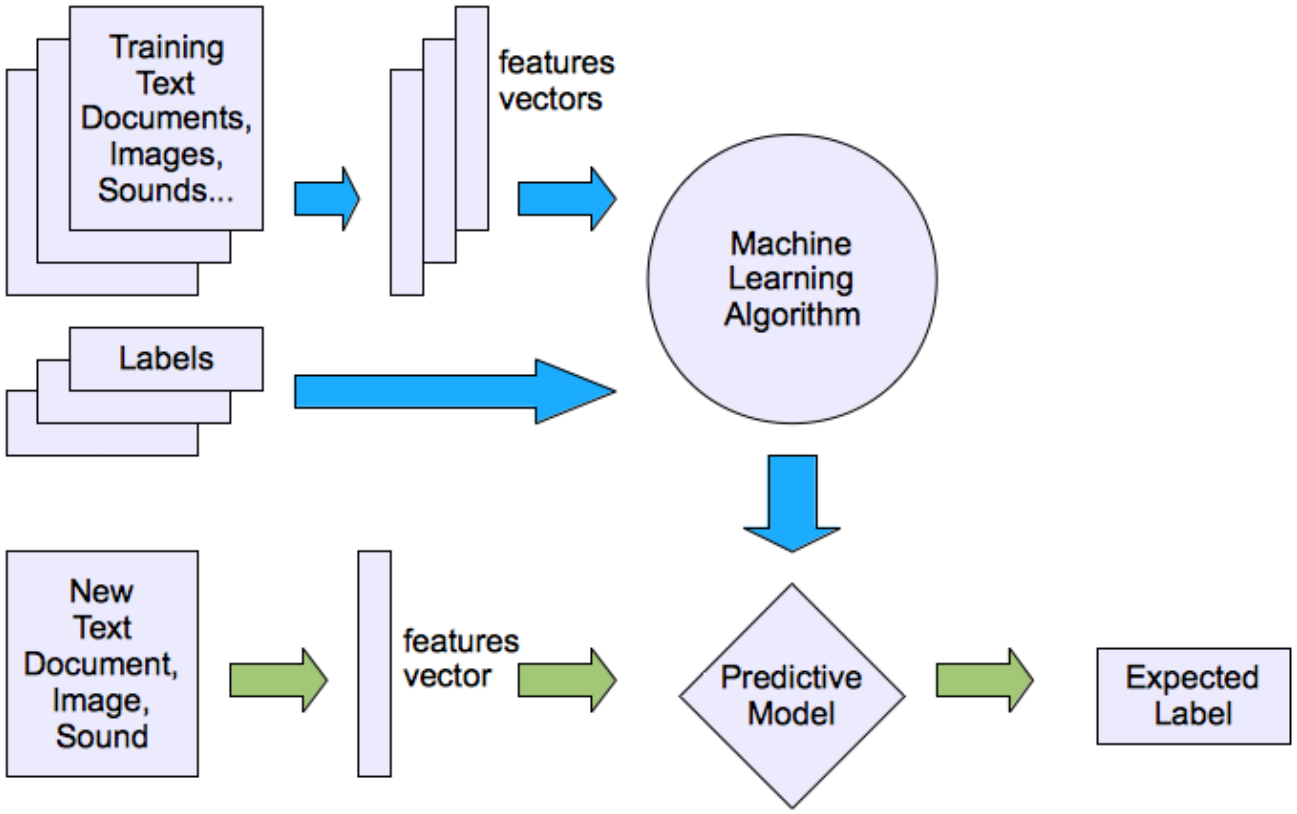
- The success of machine learning system also depends on the algorithms.
- The algorithms control the search to find and build the knowledge structures.
- The learning algorithms should extract useful information from training examples.

Algorithms

- **Supervised learning**
  - Prediction
  - Classification (discrete labels), Regression (real values)
- **Unsupervised learning**
  - Clustering
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction
- **Semi-supervised learning**
- **Reinforcement learning**
  - Decision making (robot, chess machine)

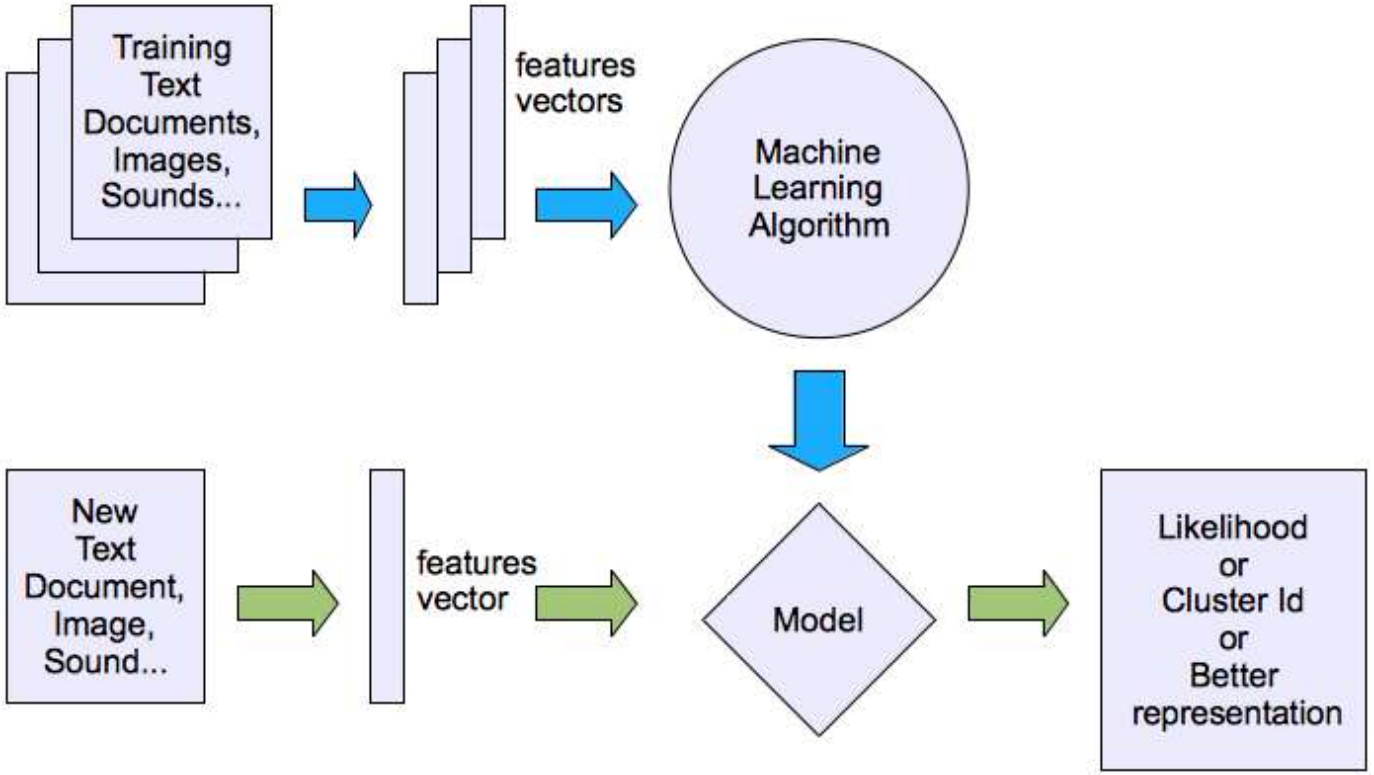
Machine learning structure

- Supervised learning



Machine learning structure

- Unsupervised learning



Examples of ML in Cyber Security

- Spam Mitigation
- Malware detection
- Mitigating Denial of Service Attacks
- Reputation in Cyber Space
- User Identification
- Detecting Identity Theft
- Information Leakage Detection and Prevention
- Social Network Security
- Detecting Advanced Persisted Threats
- Detecting Hidden Channels

Not perfect …

"We are a long way from machines that are as intelligent as humans—or even rats.

So far, we've seen only 5% of what AI can do."

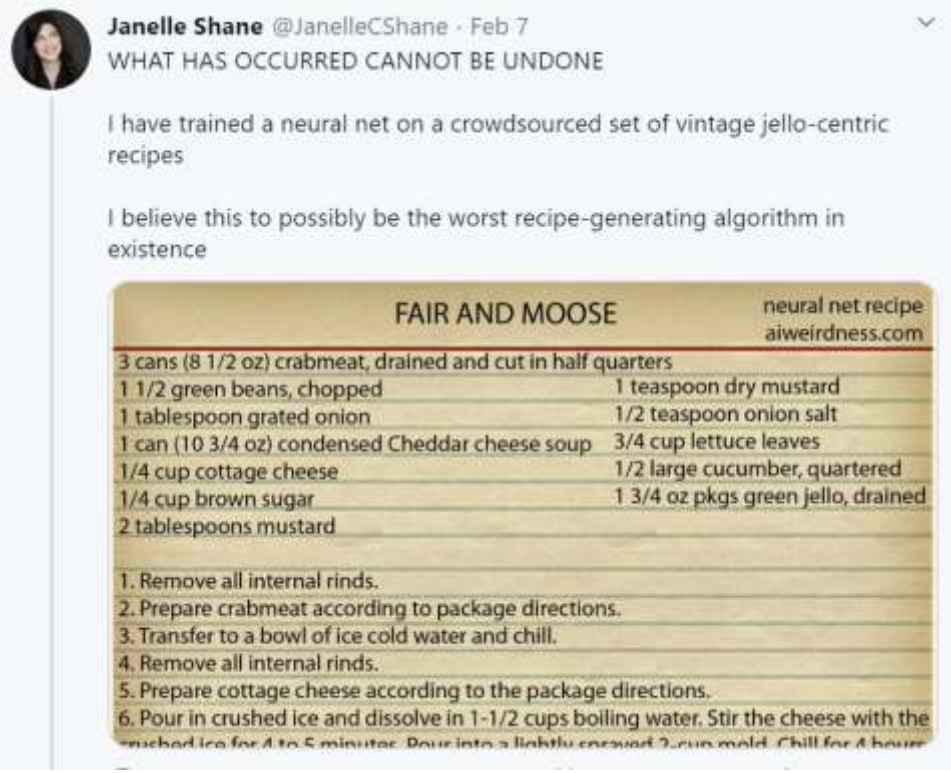
- Yann LeCun, VP and Chief AI Scientist, Facebook

Source: What's Next for Artificial Intelligence? June 14, 2016

https://www.wsj.com/articles/whats-next-for-artificial-intelligence-1465827619

“Today's AI is much closer in brainpower to an earthworm than to a human. It can pattern-match but doesn't understand what it's doing.”

Janelle Shane Feb 8, 2020 @JanelleCShane



Janelle Shane @JanelleCShane · Feb 7

WHAT HAS OCCURRED CANNOT BE UNDONE

I have trained a neural net on a crowdsourced set of vintage jello-centric recipes

I believe this to possibly be the worst recipe-generating algorithm in existence

FAIR AND MOOSE	neural net recipe aiweirdness.com
3 cans (8 1/2 oz) crabmeat, drained and cut in half quarters	
1 1/2 green beans, chopped	1 teaspoon dry mustard
1 tablespoon grated onion	1/2 teaspoon onion salt
1 can (10 3/4 oz) condensed Cheddar cheese soup	3/4 cup lettuce leaves
1/4 cup cottage cheese	1/2 large cucumber, quartered
1/4 cup brown sugar	1 3/4 oz pkgs green jello, drained
2 tablespoons mustard	

1. Remove all internal rinds.
2. Prepare crabmeat according to package directions.
3. Transfer to a bowl of ice cold water and chill.
4. Remove all internal rinds.
5. Prepare cottage cheese according to the package directions.
6. Pour in crushed ice and dissolve in 1-1/2 cups boiling water. Stir the cheese with the

#MLsec

"Just a reminder that #MLsec is about security OF ML rather than ML FOR security. Building security into ML is a thing."

"One analog might be helpful in thinking about this. Not that security software (AV, intrusion detection, and the like) is not software security (building all kinds of software to be secure) #swsec #MLsec"

Gary McGraw, Jan 31, 2020 @cigitalgem

ML Security Risks

- Someone fooling a machine learning system by presenting malicious input of data that causes a system to make a false prediction or categorization.
- Attacker intentionally manipulates data being used by machine learning – which could compromise an entire system ("*adversarial example*").
- Data confidentiality – information meant to be protected could be, through subtle means, extracted from a machine learning model containing that data

Adversarial example

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$

$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$=$

$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

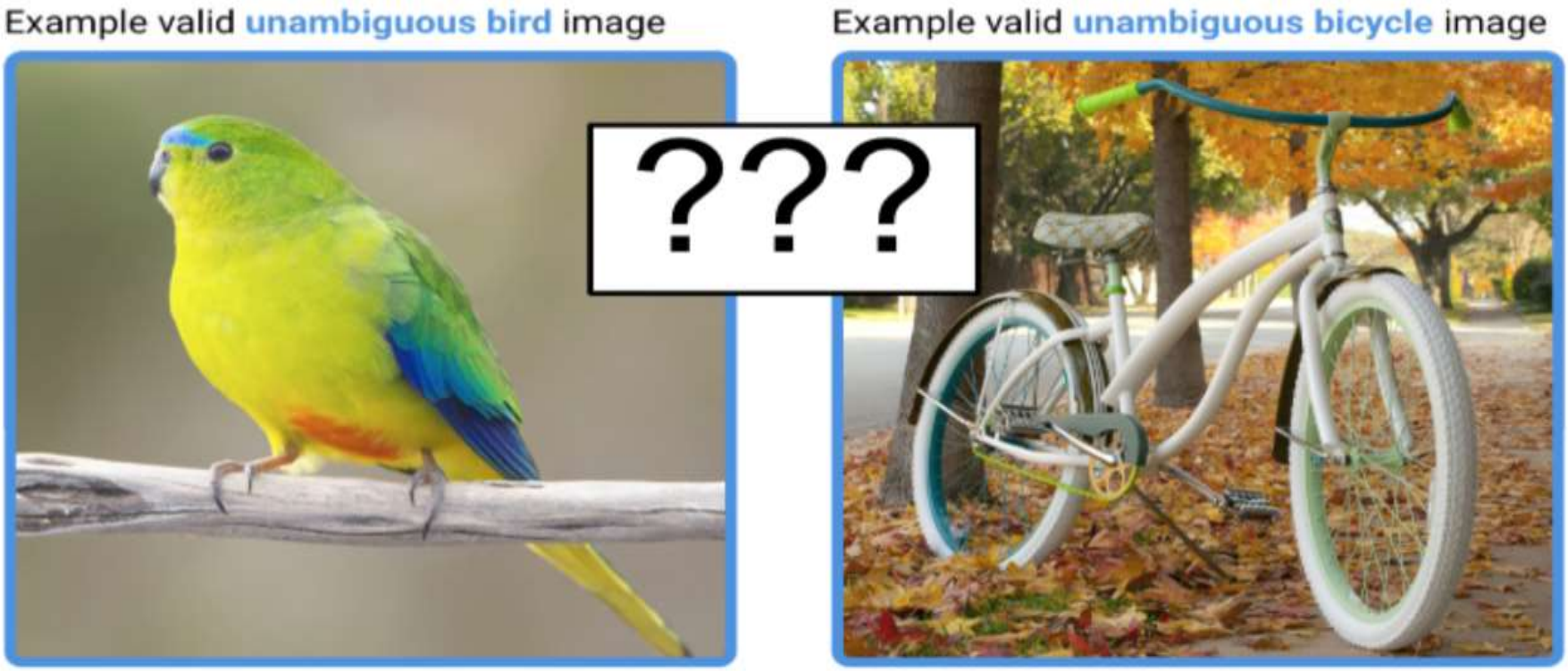
"gibbon"
99.3 % confidence

(Goodfellow 2018)

After changing each pixel a tiny bit, the new image is incorrectly classified with extremely high confidence.

Source: "Unsolved research problems vs. real-world threat models" by Catherine Olsson (March 26, 2019) (https://medium.com/@catherio/unsolved-research-problems-vs-real-world-threat-models-e270e256bc9e)

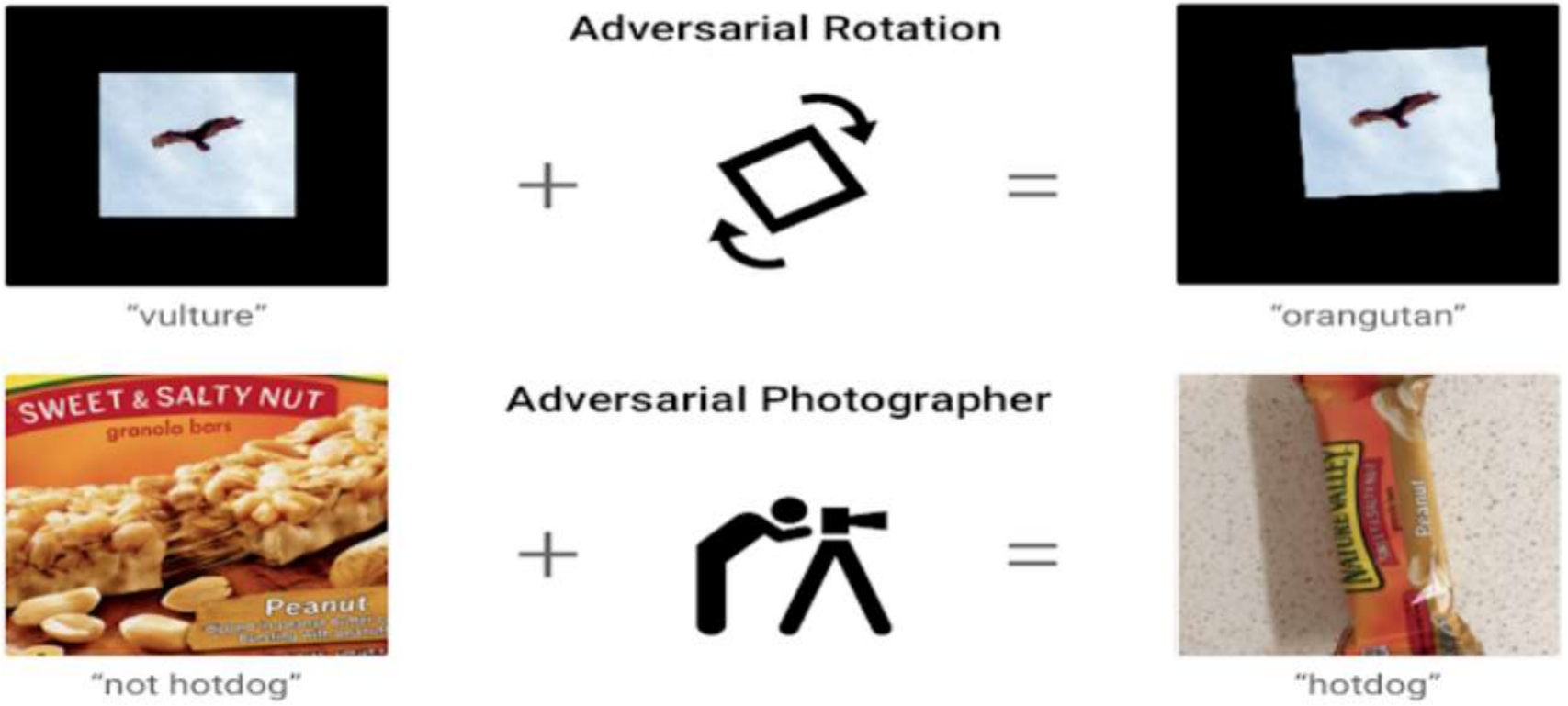
Ambiguous examples



Current image classifiers cannot reliably distinguish between unambiguous bird and bicycle images. [Unrestricted Adversarial Examples Challenge].

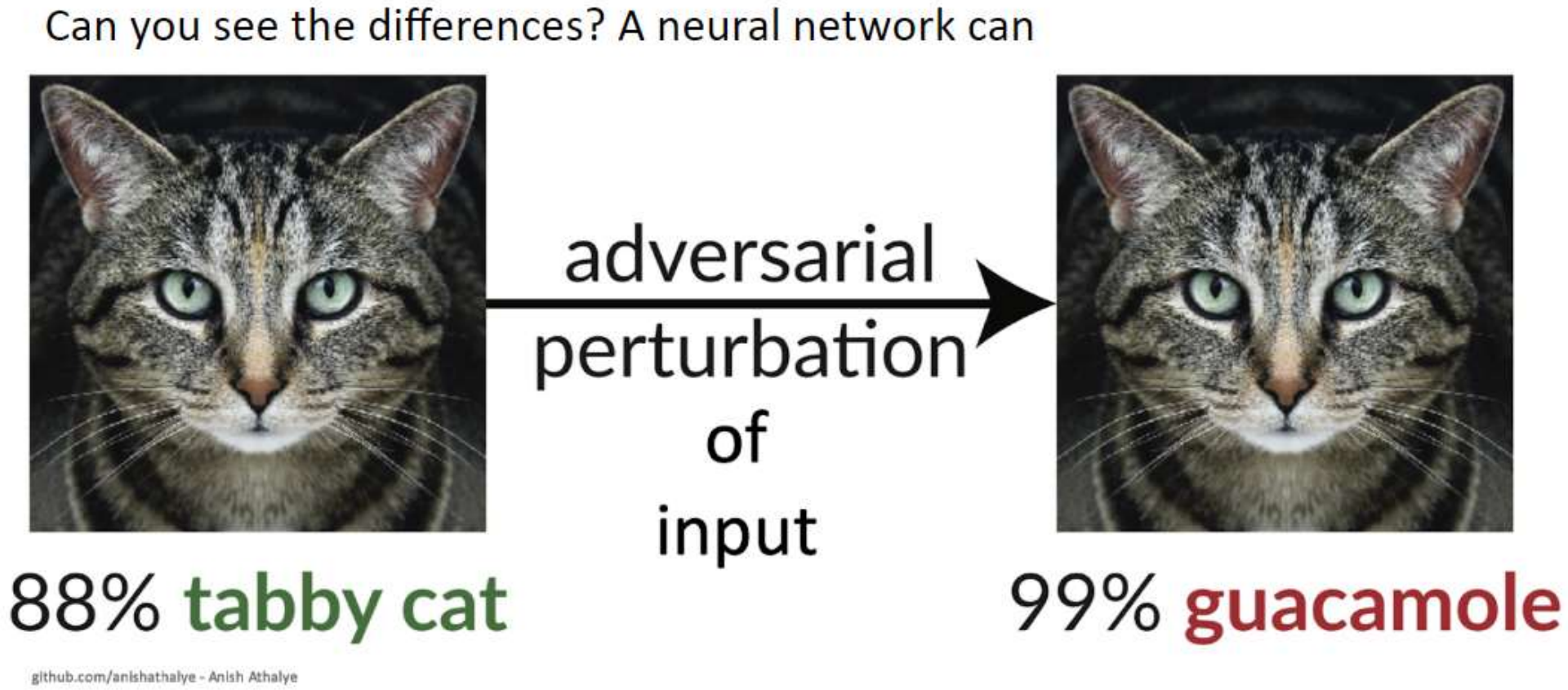
Source: “Unsolved research problems vs. real-world threat models” by Catherine Olsson (March 26, 2019) (https://medium.com/@catherio/unsolved-research-problems-vs-real-world-threat-models-e270e256bc9e)

More adversarial examples



Source: "Unsolved research problems vs. real-world threat models" by Catherine Olsson (March 26, 2019) (https://medium.com/@catherio/unsolved-research-problems-vs-real-world-threat-models-e270e256bc9e)

Neural Networks are Inherently Unpredictable



Source: "Introduction to Feed-forward Neural Networks (aka Deep Learning)" by Abraham King (https://docs.google.com/presentation/d/1a_kUtLZthsDdd_NGfBUBnsjrCKGljsYCWXb5dAjtQK4/edit#slide=id.g5815686d6f_1_49)

AI/ML-specific Threats

- #1: Adversarial Perturbation
  - Variant #1a: Targeted misclassification
  - Variant #1b: Source/Target misclassification
  - Variant #1c: Random misclassification
  - Variant #1d: Confidence Reduction
- #2: Data Poisoning
  - #2a Targeted Data Poisoning
  - #2b Indiscriminate Data Poisoning
- #3 Model Inversion Attacks
- #4 Membership Inference Attack
- #5 Model Stealing
- #6 Neural Net Reprogramming
- #7 Adversarial Example in the Physical domain (bits->atoms)
- #8 Malicious ML providers who can recover training data
- #9 Attacking the ML Supply Chain
- #10 Backdoor Machine Learning
- #11 Exploit software dependencies of the ML system

Source: “Threat Modeling AI/ML Systems and Dependencies” Microsoft (https://docs.microsoft.com/en-us/security/threat-modeling-aiml)

AI/ML-specific Threat / Mitigation

- #1: Adversarial Perturbation
  - Variant #1a: Targeted misclassification
- Mitigations
  - **Reinforcing Adversarial Robustness using Model Confidence Induced by Adversarial Training** [19]: The authors propose Highly Confident Near Neighbor (HCNN), a framework that combines confidence information and nearest neighbor search, to reinforce adversarial robustness of a base model. This can help distinguish between right and wrong model predictions in a neighborhood of a point sampled from the underlying training distribution.
  - **Attribution-driven Causal Analysis** [20]: The authors study the connection between the resilience to adversarial perturbations and the attribution-based explanation of individual decisions generated by machine learning models. They report that adversarial inputs are not robust in attribution space, that is, masking a few features with high attribution leads to change indecision of the machine learning model on the adversarial examples. In contrast, the natural inputs are robust in attribution space.

Source: "Threat Modeling AI/ML Systems and Dependencies" Microsoft (https://docs.microsoft.com/en-us/security/threat-modeling-aiml)

AI/ML-specific Threat / Mitigation

- #1: Adversarial Perturbation
  - Variant #1a: Targeted misclassification
- Mitigations
  - Attribution-driven Causal Analysis [20]:

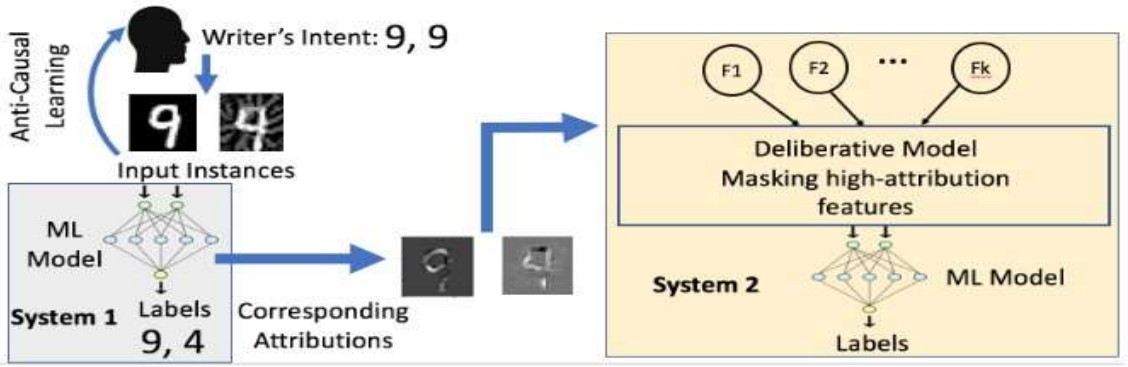


Figure 2: The architecture of the proposed approach motivated by the two level Kahneman's decomposition of cognition. Typical machine learning models for classification perform anti-causal learning to determine the label from the input instance. As noted by (Chalasani et al., 2018), such anti-causal reasoning lacks the natural continuity of causal mechanisms and is often not robust. But we view this model as System 1 and use attribution methods (Integrated Gradient in our experiments) to obtain features with positive and negative attributions. In this example with the MNIST dataset, we see that the adversarial perturbation that causes misclassification of 9 into 4 also significantly changes the attributions. For example, the top part of the perturbed 9 (misclassified as 4) has negative attribution. In deliberative System 2, we perform reasoning in the causal direction, and mask the high attribution features (pixels in this case) to obtain a number of input instances in the causal neighborhood of the original image. The original attributions are robust but the adversarial attributions are not robust which causes the model to assign different labels to images in the causal neighborhood of adversarial examples.

Source: "Threat Modeling AI/ML Systems and Dependencies" Microsoft (https://docs.microsoft.com/en-us/security/threat-modeling-aiml)

Learn more …

black hat USA 2020

REGISTER NOW

AUGUST 1 - 6, 2020

MANDALAY BAY / LAS VEGAS

ATTEND | TRAININGS | BRIEFINGS | ARSENAL | FEATURES | SCHEDULE | BUSINESS HALL | SPONSORS | PROPOSALS

BRIEFINGS TRACKS

AI, ML, & DATA SCIENCE

The focus of the AI, ML, and Data Science track is the use and impact of AI/ML and its sub-disciplines on the security domain. This track welcomes both an offensive and defensive perspective. Relevant content would be in the use of AI for offensive activities, attacks against systems implementing AI, defending systems implementing AI, and the use of AI in solving security challenges. The content for the track should have a heavy focus on practical and applied concepts related to the topic area where the AI/ML functionality plays a key role.

TRACK LEAD
NATHAN HAMIEL

Resources

- "Unsolved research problems vs. real-world threat models" by Catherine Olsson (March 26, 2019)

  https://medium.com/@catherio/unsolved-research-problems-vs-real-world-threat-models-e270e256bc9e

- "The Inherent Insecurity in Neural Networks and Machine Learning Based Applications" by Abraham Kang (May 15, 2019)

  https://towardsdatascience.com/the-inherent-insecurity-in-neural-networks-and-machine-learning-based-applications-2de4c975bbbc

- Berryville Institute of Machine Learning (BIML) – soon to release 78 particular risks in ML systems

  https://berryvilleiml.com/

Resources

- Microsoft

  "Threat Modeling AI/ML Systems and Dependencies" (11/11/2019)
  https://docs.microsoft.com/en-us/security/threat-modeling-aiml

  "AI/ML Pivots to the Security Development Lifecyle Bug Bar" (11/11/2019)
  https://docs.microsoft.com/en-us/security/bug-bar-aiml

  "Failure Modes in Machine Learning" (11/11/2019)
  https://docs.microsoft.com/en-us/security/failure-modes-in-machine-learning

BIML's review: https://berryvilleiml.com/2020/01/16/on-recent-microsoft-and-nist-ml-security-documents/ (January 16, 2020)

Resources

Twitter

- AI + ML + Security experts on Twitter

  Gary McGraw (BIML/#MLSec) @cigitalgem

  Matthew Rosenquist (AI/Privacy) @Matt_Rosenquist

- AI + ML experts and influencers on Twitter (lists)

  https://medium.com/@oleksii_kh/learn-in-your-tweet-15-top-ai-experts-to-follow-on-twitter-5c68681af90a

  https://medium.springboard.com/30-twitter-influencers-you-have-to-follow-for-ai-machine-learning-977587b6406e

Questions?

Contacts

Web Site:
https://roberthurlbut.com

Twitter: @RobertHurlbut