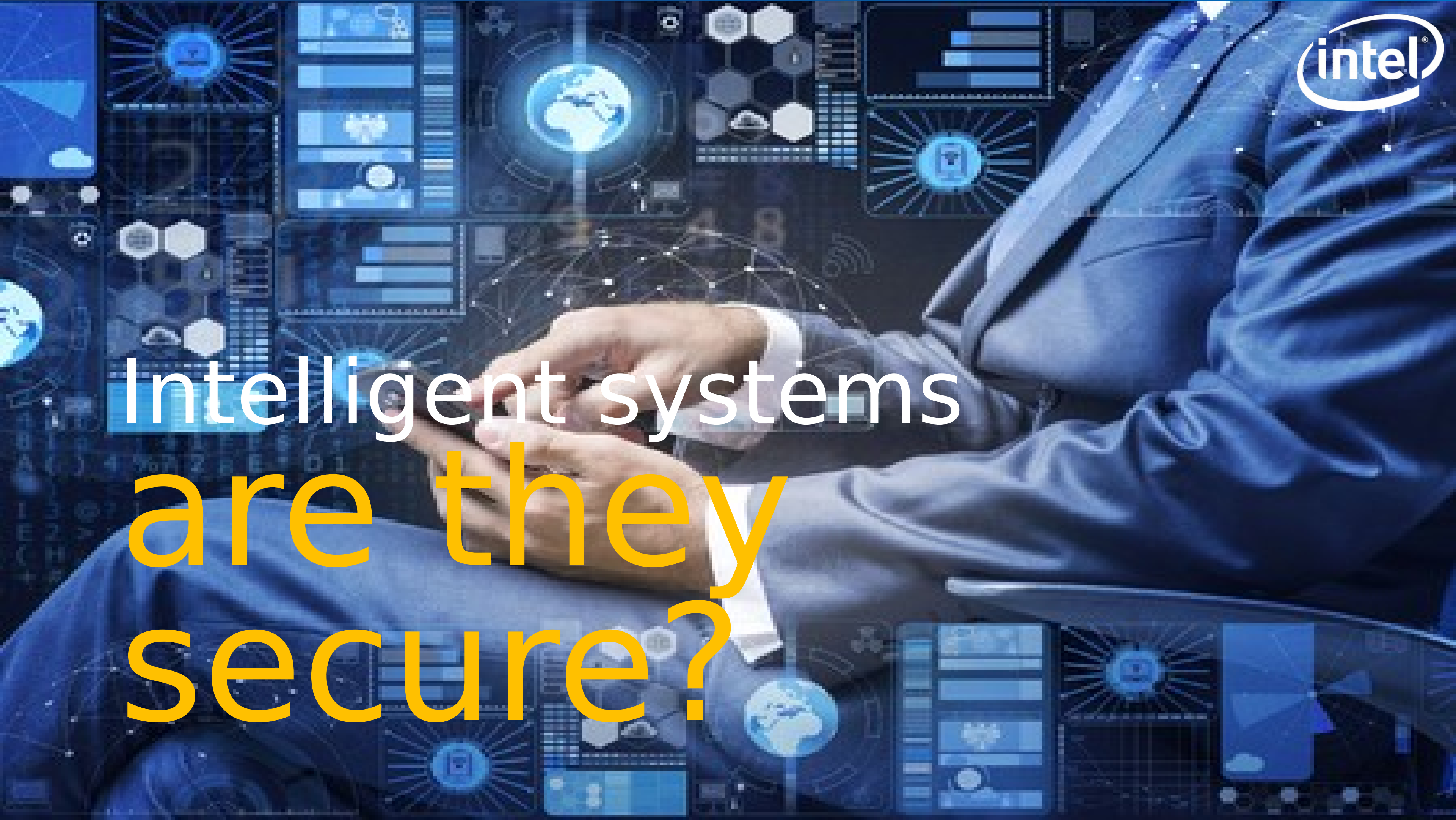


Intelligent systems
are they secure?

LEGAL DISCLAIMER

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software, or service activation. Performance varies depending on system configuration. Check with your system manufacturer or retailer or learn more at www.intel.com.

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. Consult other information and performance tests to assist with fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to www.intel.com/benchmarks.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at [intel.com].

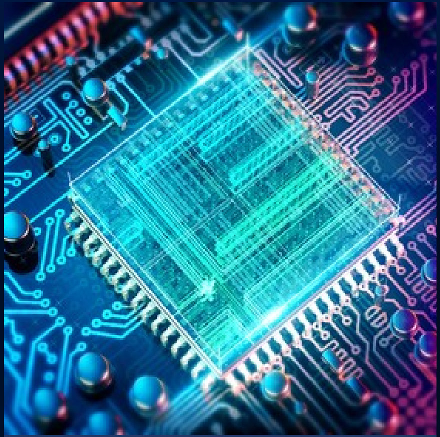
Intel provides these materials as-is, with no express or implied warranties.

Intel and the Intel logo are trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries.
*Other names and brands may be claimed as the property of others.

ARTIFICIAL INTELLIGENCE

SOFTWARE

HARDWARE



COMMUNITY

ARTIFICIAL INTELLIGENCE & SECURITY

SECURITY FOR AI

AI FOR SECURITY

intel XEON inside

intel NERVANA inside

intel STRATIX 10 inside

intel CORE i7 inside

intel MOVIDIUS inside

* Representation image only

Security for AI considerations

CYBERWEEK

Clever Hans

“We have reached the point where machine learning works, but may easily be broken”

Nicolas Papernot, Google PhD Fellow in Security
Ian Goodfellow, Research scientist at Google Brain

Artificial intelligence

Machine Learning

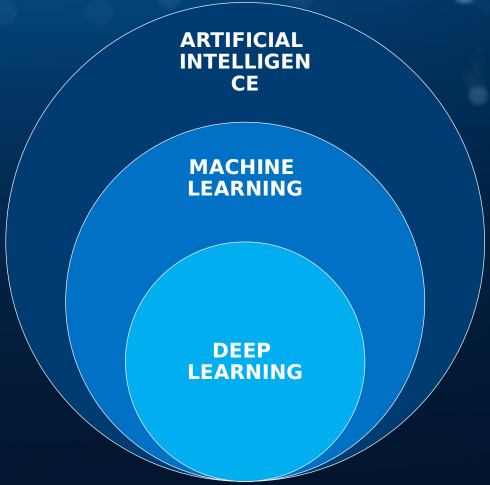
- Study many images labeled as flamingo
- Identify the flamingo in the image

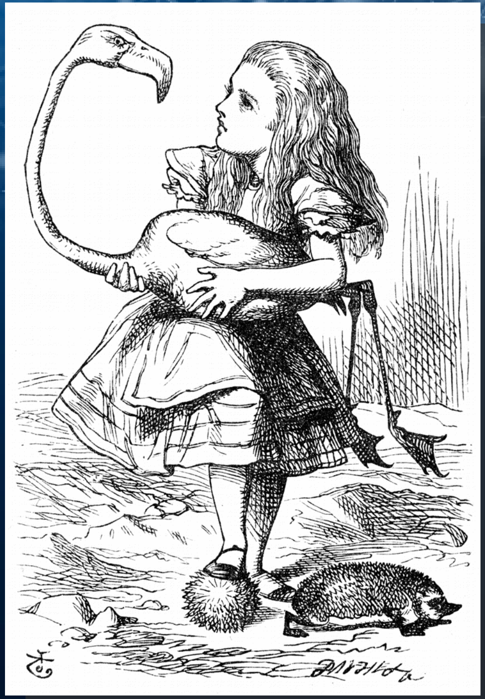
Deep Learning

- Study many images
- Identify the flamingo, hedgehog, etc.

Artificial Intelligence

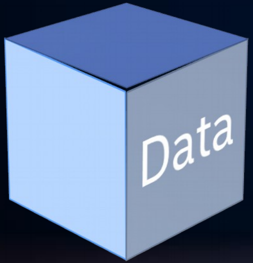
- Is she hugging the flamingo, or playing cricket?
- Is she happy, sad?



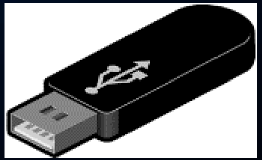


A Billion Dollar Problem

$1B



$50M



What are the issues?

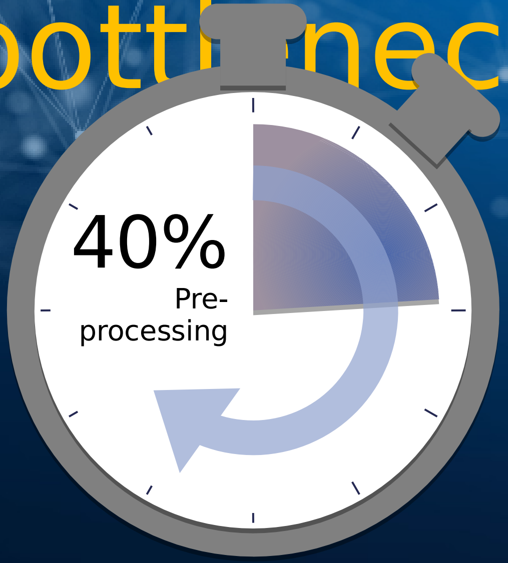
Data vs. Information

New Threat Vectors

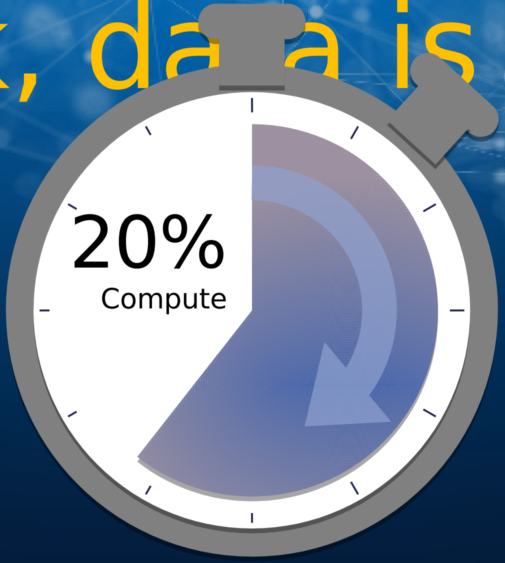
Upcoming Attacks

Unique Gaps

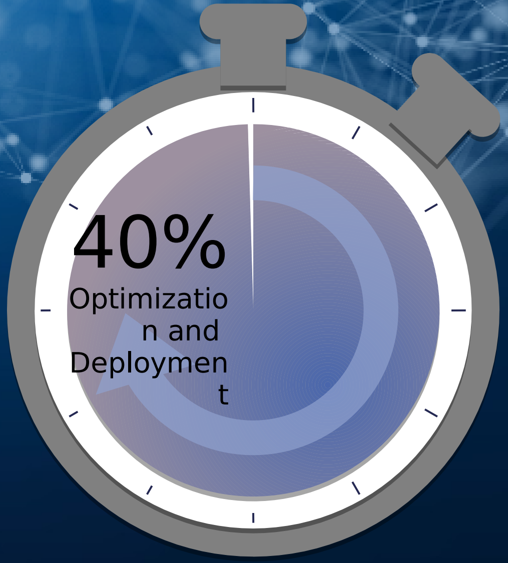
Training Compute is not the bottleneck, data is



Preparing the data for analysis, Finding the right model for the problem



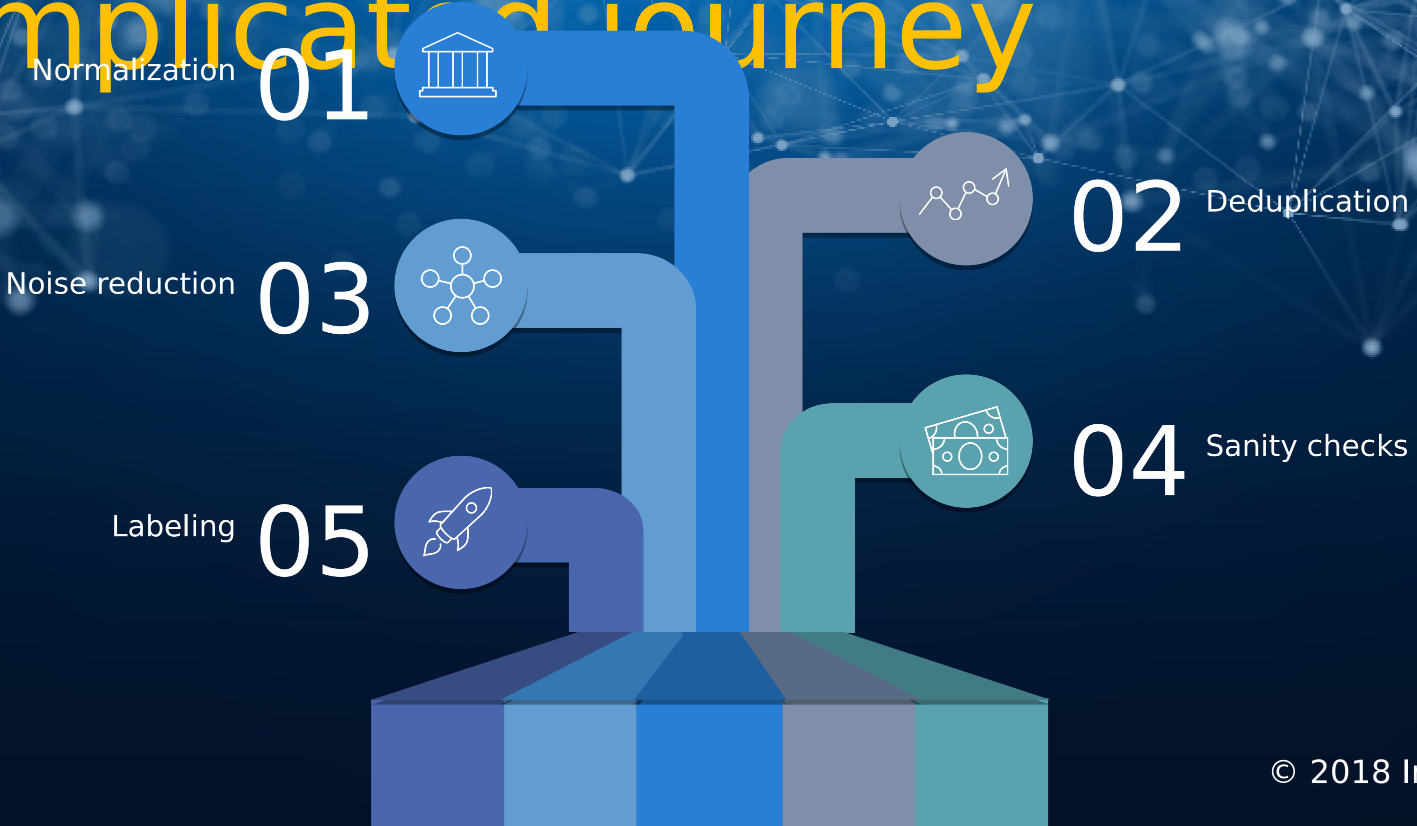
Training the model



Optimizing and deploying the model

Inference is a different story!

Pre-processing – it's a complicated journey



01 Normalization

02 Deduplication

03 Noise reduction

04 Sanity checks

05 Labeling

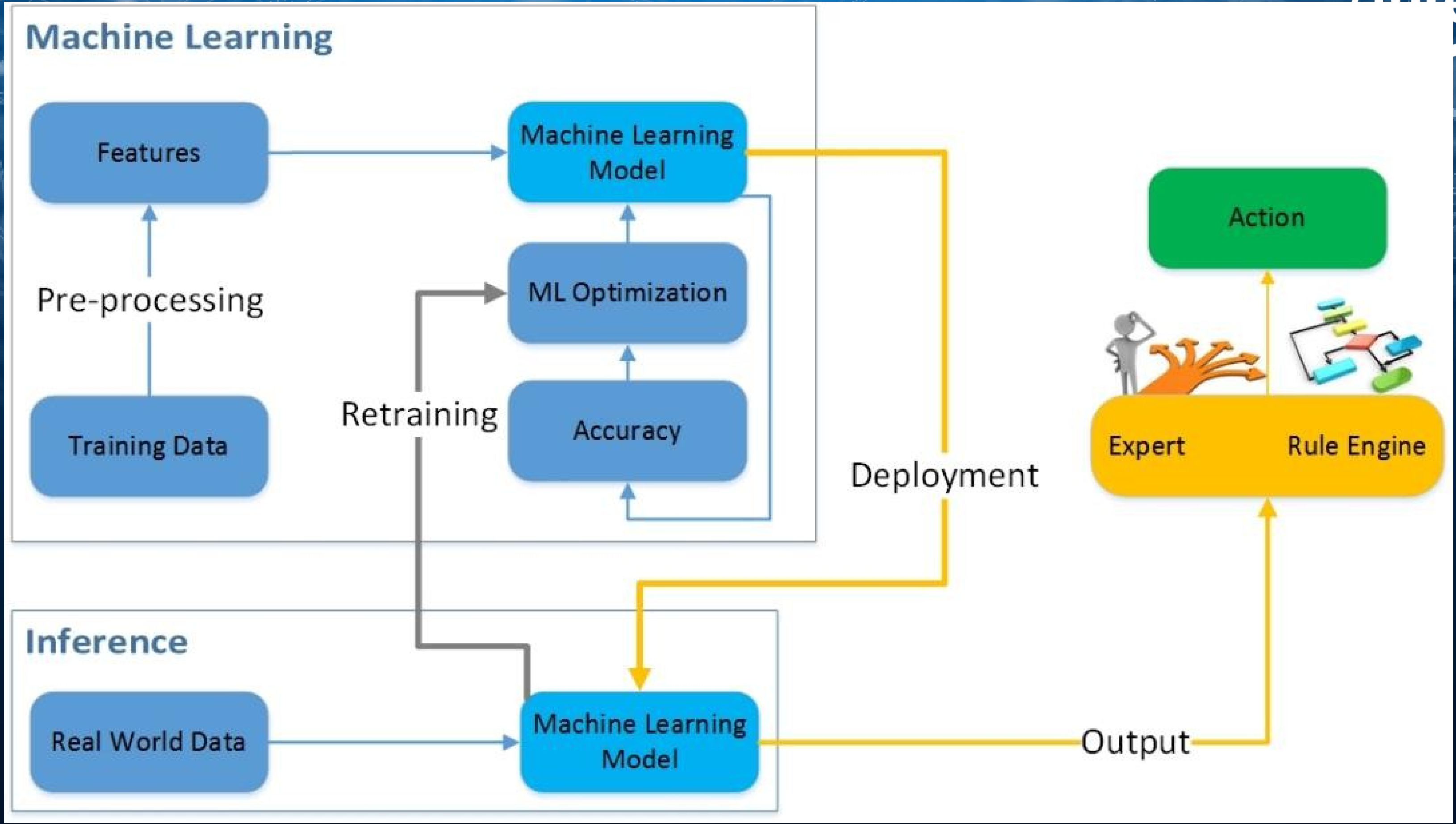
What are the issues?

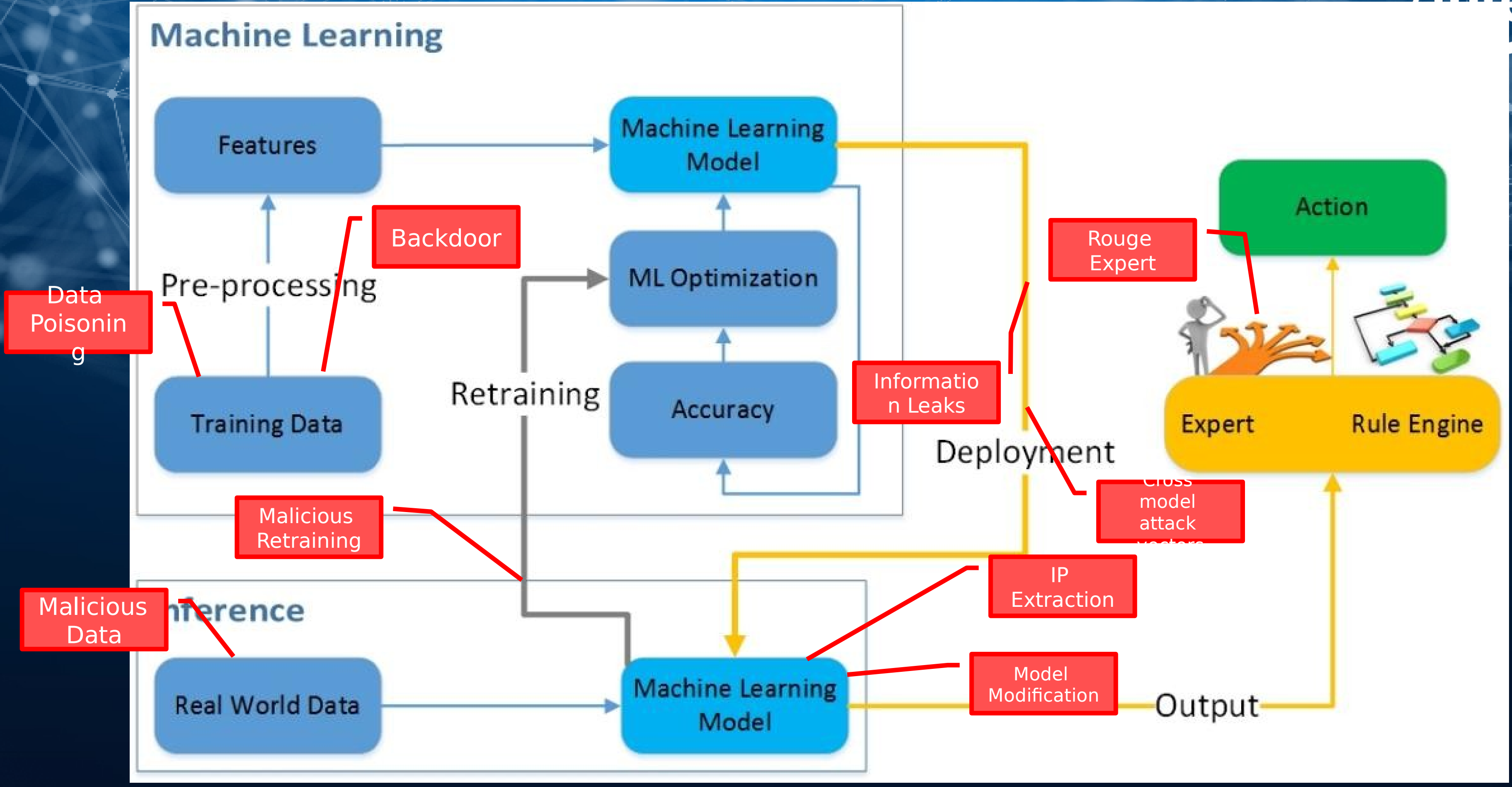
Data vs. Information

New Threat Vectors

Upcoming Attacks

Unique Gaps


Machine Learning
Features
Machine Learning Model
Pre-processing
ML Optimization
Retraining
Accuracy
Training Data
Deployment
Action
Expert
Rule Engine
Inference
Real World Data
Machine Learning Model
Output



Machine Learning

Features

Machine Learning Model

Backdoor

Pre-processing

Data Poisoning

ML Optimization

Retraining

Accuracy

Training Data

Information Leaks

Deployment

Rouge Expert

Action

Expert

Rule Engine

Cross model attack vectors

Malicious Retraining

IP Extraction

Malicious Data

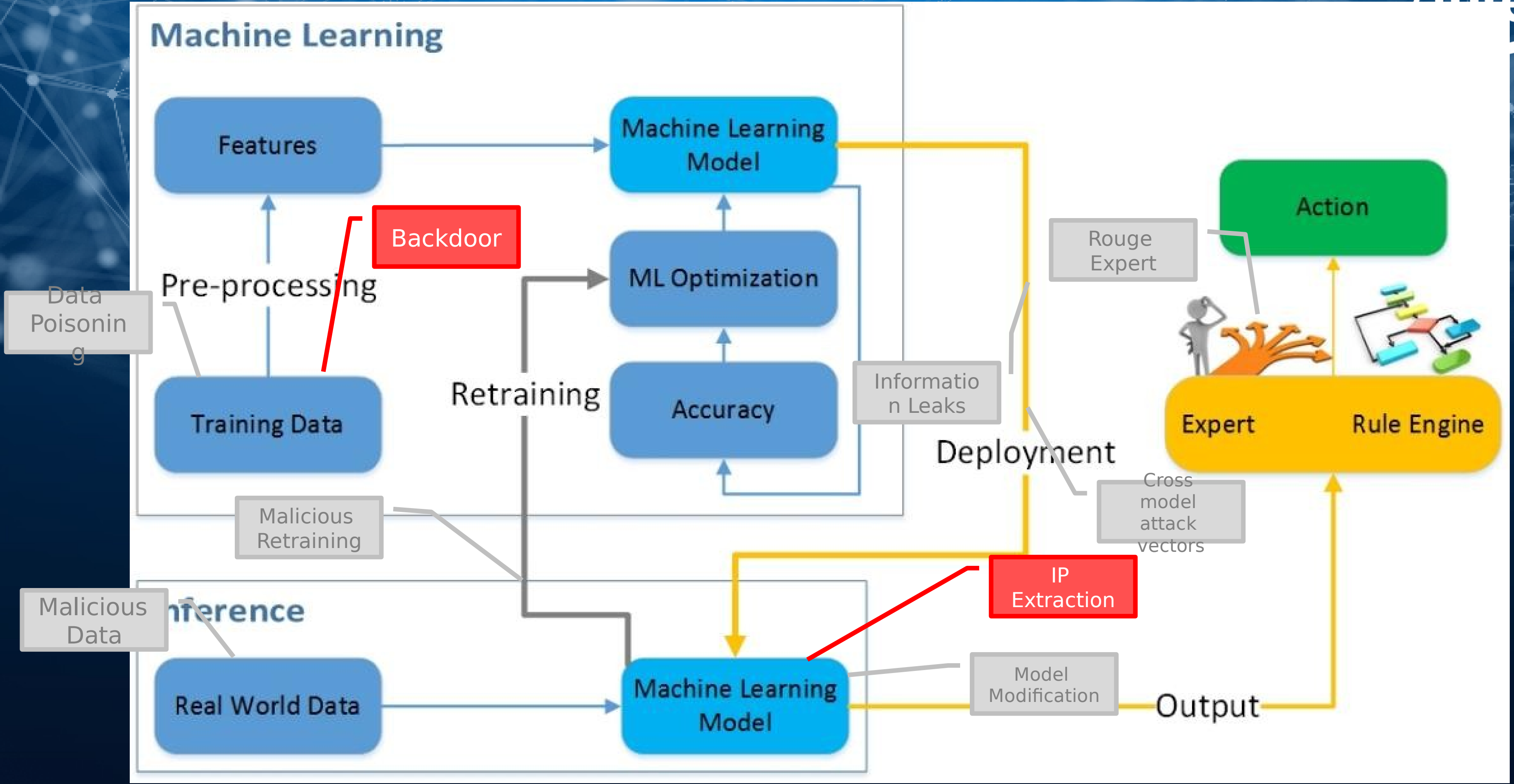
Inference

Real World Data

Machine Learning Model

Model Modification

Output



Machine Learning

Features

Machine Learning Model

Backdoor

Pre-processing

Data Poisonin g

ML Optimization

Retraining

Training Data

Accuracy

Informatio n Leaks

Rouge Expert

Action

Expert

Rule Engine

Deployment

Cross model attack vectors

Malicious Retraining

IP Extraction

Malicious Data

nference

Real World Data

Machine Learning Model

Model Modification

Output

Backdoors

Validation of ML is an open problem

We don’t have a method for detecting backdoors

- Reverse engineering, code review are not applicable to ML

IP Extraction

IP can be stolen using public APIs

Reverse engineering or device access not required

What are the issues?

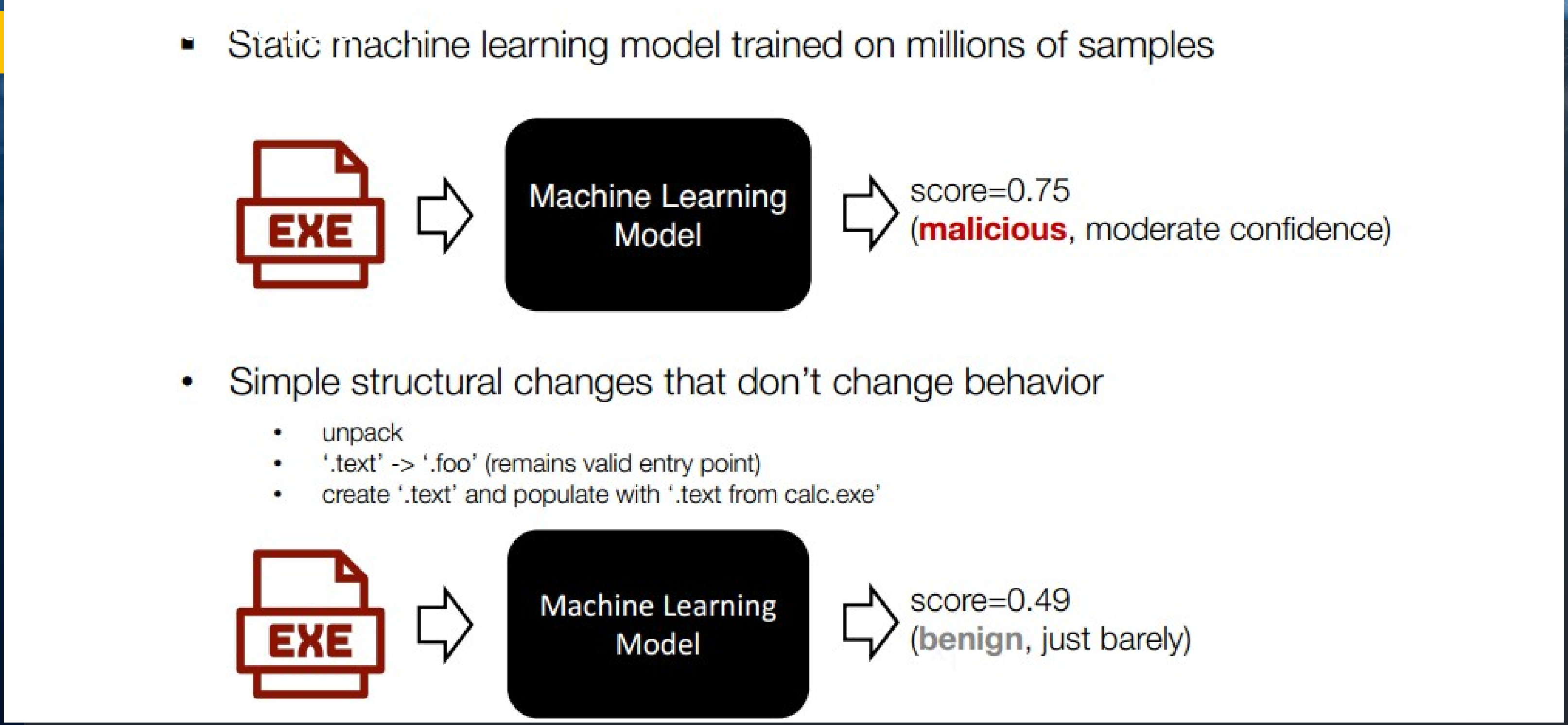
Data vs. Information

New Threat Vectors

Upcoming Attacks

Unique Gaps

Evading next generation AV u



- Static machine learning model trained on millions of samples

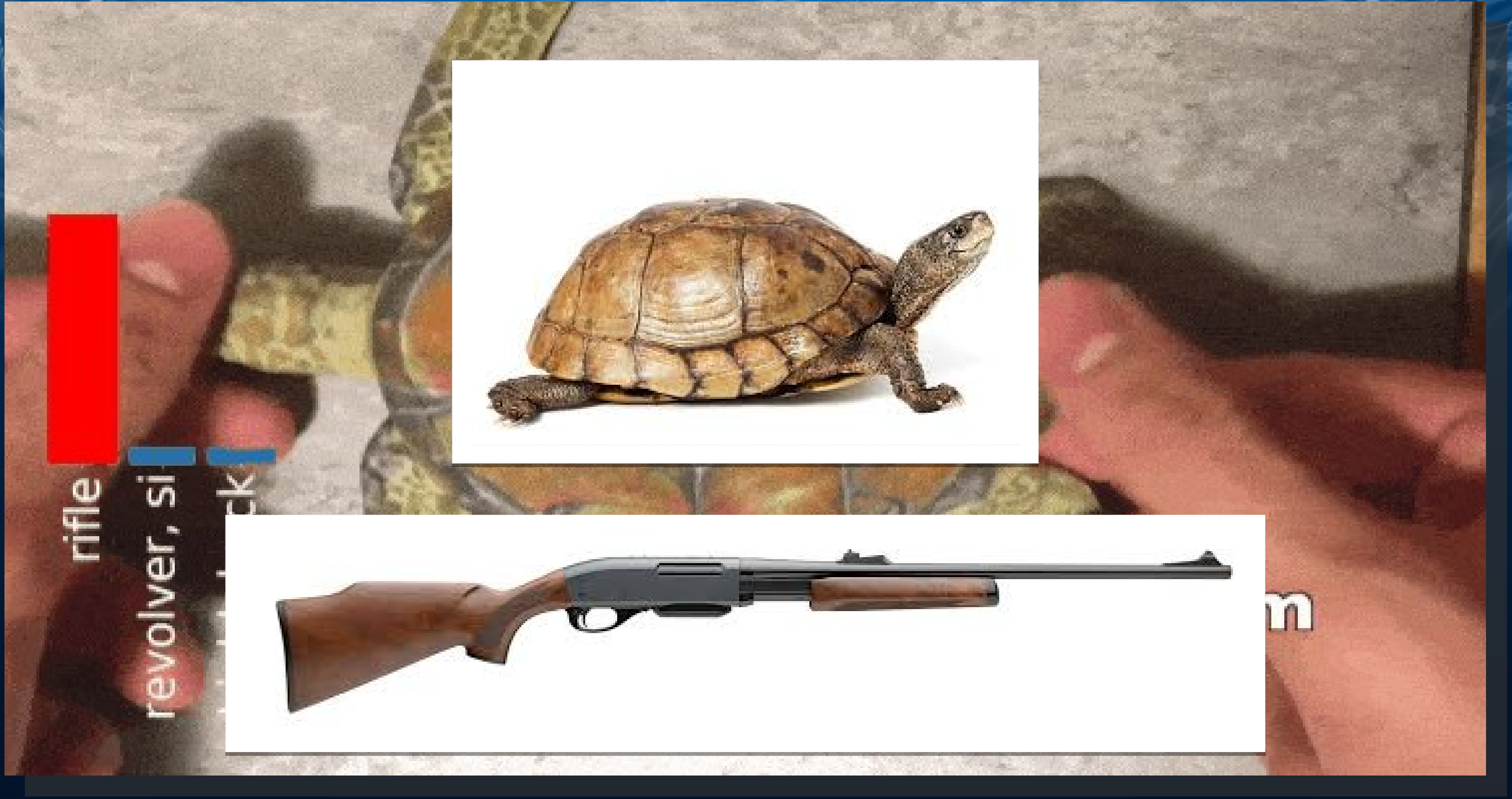
score=0.75
(**malicious**, moderate confidence)

- Simple structural changes that don't change behavior
  - unpack
  - '.text' -> '.foo' (remains valid entry point)
  - create '.text' and populate with '.text from calc.exe'

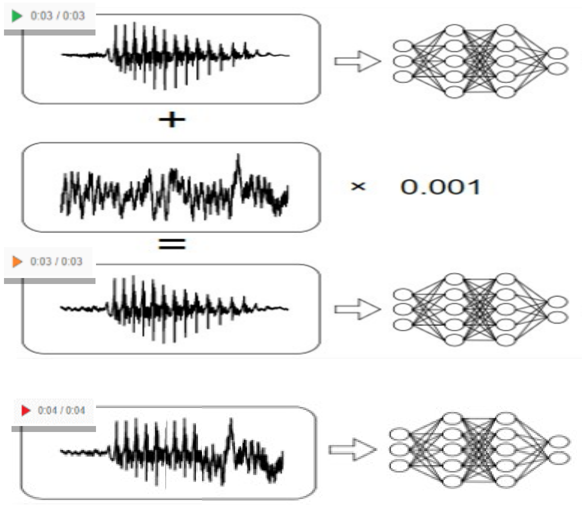
score=0.49
(**benign**, just barely)

https://www.youtube.com/watch?v=FGCIe6T0Jpc

Turtle or a Rifle?



Adversarial Audio



Adversarial Verdi's Requiem

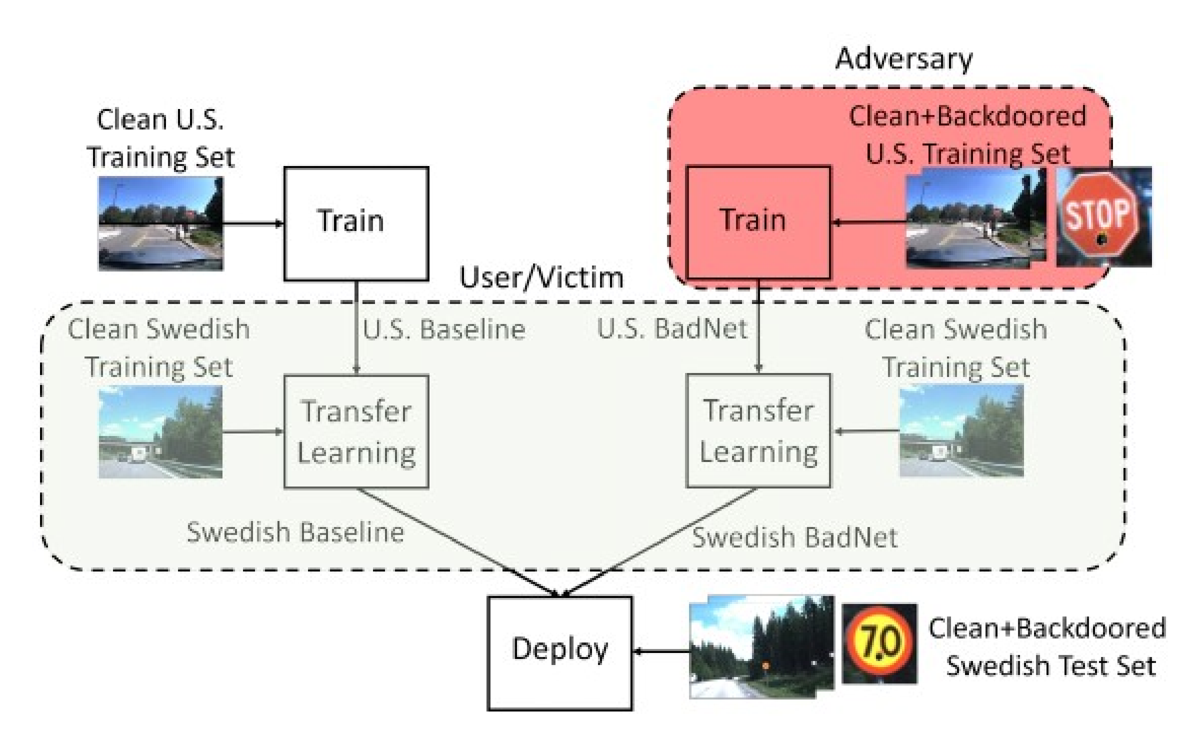
“okay google without the dataset the article is useless”

“okay google browse to evil dot com”

“okay google browse to evil dot com”

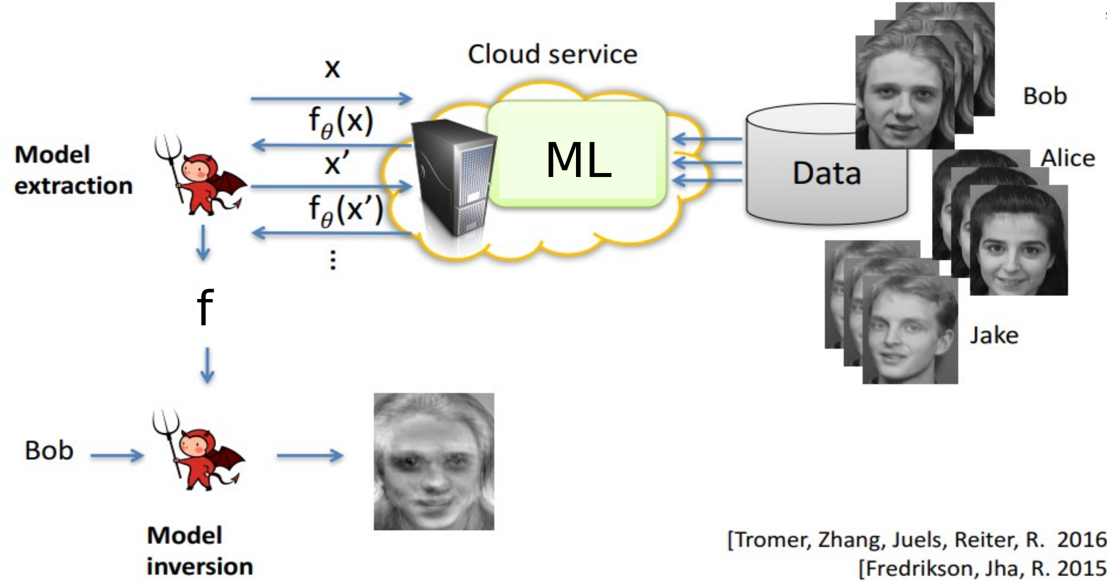
You can fool home automation, smartphones and other devices

Supply chain security – in AI



https://machine-learning-and-security.github.io/papers/mlsec17_paper_51.pdf

Information Disclosure by



[Tromer, Zhang, Juels, Reiter, R. 2016]
[Fredrikson, Jha, R. 2015]

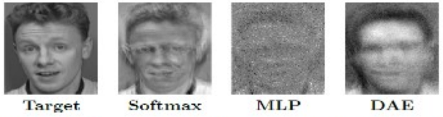
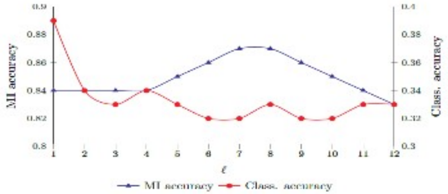
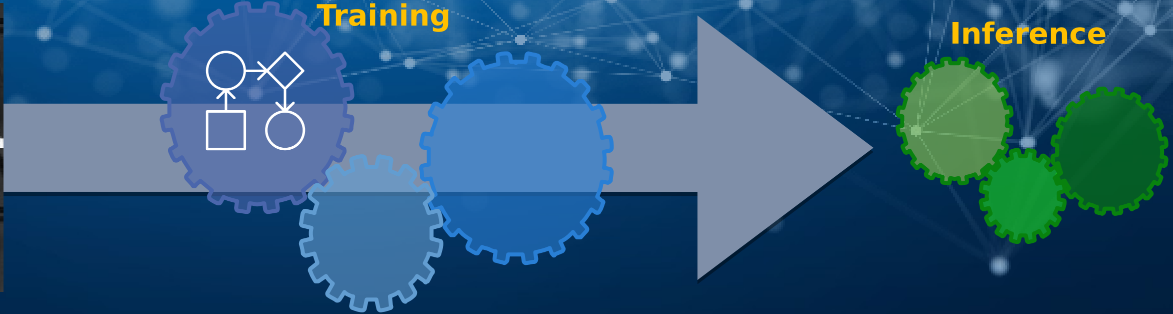


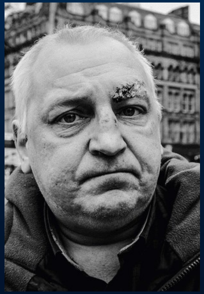
Figure 10: Reconstruction of the individual on the left by Softmax, MLP, and DAE.



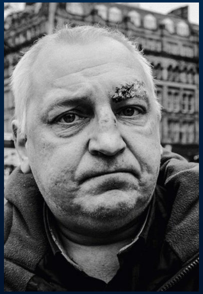
What about privacy?

Privacy leaks? Not yet, but soon...



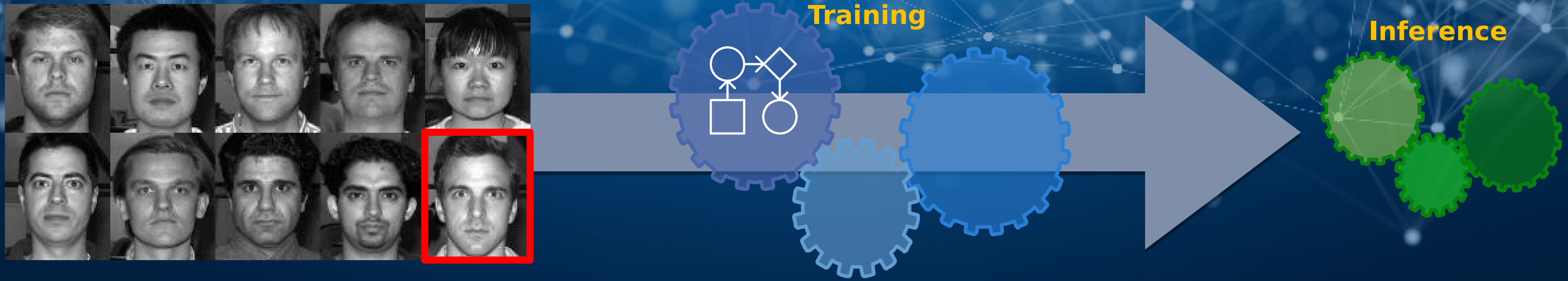


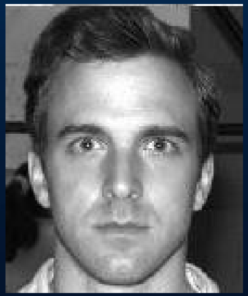
Risk: 7.4%



Risk: 35.3%

Privacy leaks? Not yet, but soon…





Risk: 96.2%

What are the issues?

Data vs. Information

New Threat Vectors

Upcoming Attacks

Unique Gaps

AI Security: Unique Gaps

IP protections are early stage (at best)

AI Security: Unique Gaps

AI Validation is a major issue

- Pretty clear if the AI does what it claims, does it do more?
- Will it fail unexpectedly?

AI Security: Unique Gaps

You shouldn’t trust the data, even if collected securely, the data might be

AI Security: dynamic systems

You may end with a different system then what you started with

AI Security: Unique Gaps

Humans in the loop pose a security risk, we don’t have sufficient controls during Machine Learning development

So, what can we do?

Our Recommendations

1. Start having conversations about Security and AI
2. Machine learning needs to be protected against attackers
3. Checks and balances, don’t trust blindly
4. Protecting AI is a journey, join us in our Keynote on Wednesday

Set an agenda in your org to discuss your AI and Data security plans

Schedule meetings with your teams to discuss your posture on AI security

Reach out to us to discuss these issues after this talk

Remember Mr. ed the talking horse?

Please reach out for more information

Guy Barnhart-Magen

guy.barnhart-magen@intel.com

@barnhartguy (twitter)