Symantec

AI/ML in Cybersecurity

Andrew B. Gardner, Ph.D.

Senior Technical Director, Head of AI/ML
Center for Advanced Machine Learning (CAML)
Symantec Corporation
@andywocky
andrew_gardner@symantec.com

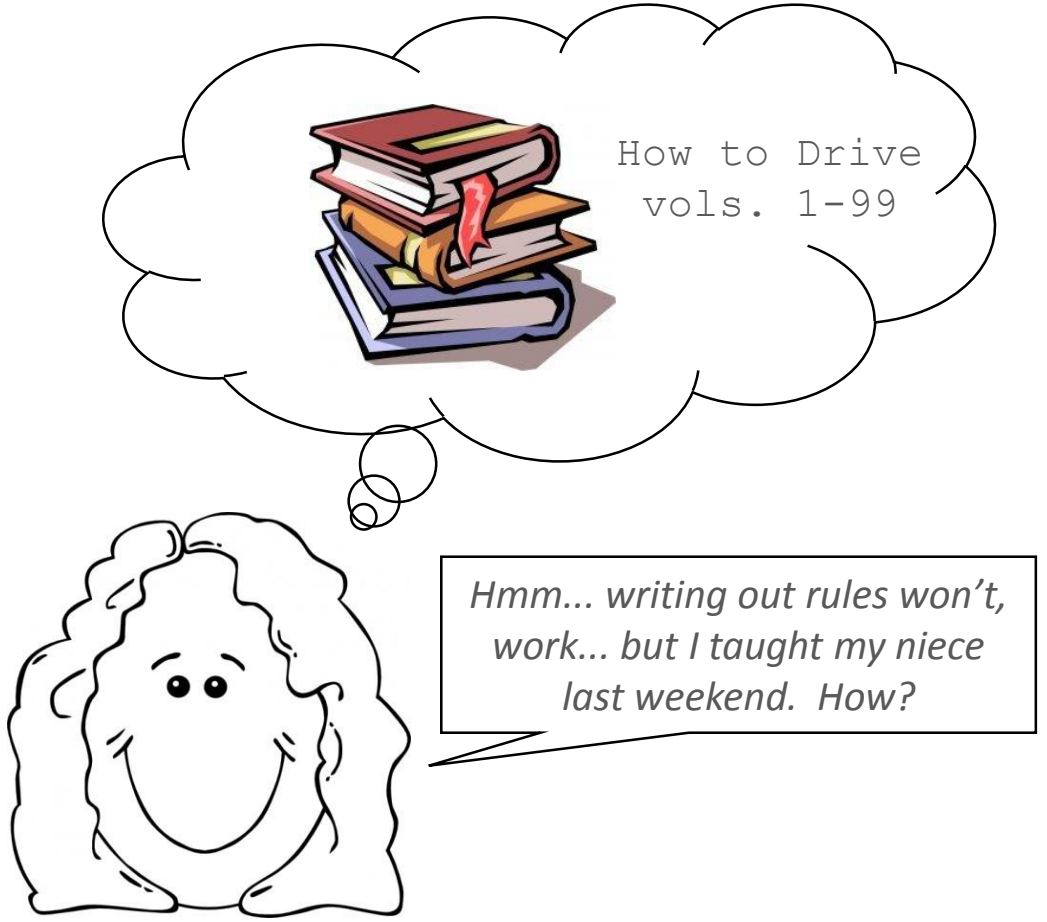
It’s a Golden Age for AI/ML

#doing-things

We aspire to create self-driving cars.

Learning By Example Changes Things

Waymo One. https://youtu.be/Eq89YGbERzs



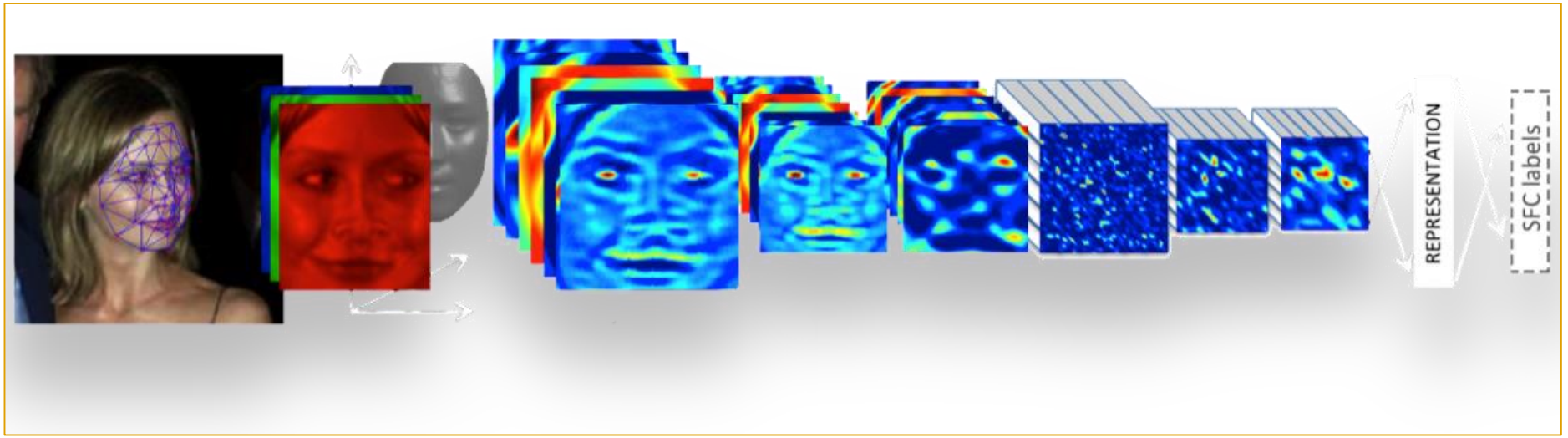
- Big data vs. many skilled developers
- Empirically trained vs. explicitly designed
- Faster, scalable ... but are there tradeoffs?

Thinking like an AI engineer.

A Cambrian Explosion of AI/ML and the 98% Rule

#perceiving-things

Computational perception – understanding humans



Y. Taigman et al., "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," https://research.fb.com/wp-content/uploads/2016/11/deepface-closing-the-gap-to-human-level-performance-in-face-verification.pdf

:

face recognition (and images, speech, text, social, video, etc.)

detect / predict at 98%+ for the win

Definitions

MACHINE LEARNING

The capability of a machine to learn without explicitly being programmed.

learning

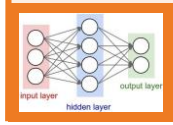
ARTIFICIAL INTELLIGENCE

The capability of a machine to imitate intelligent human behavior.

perception

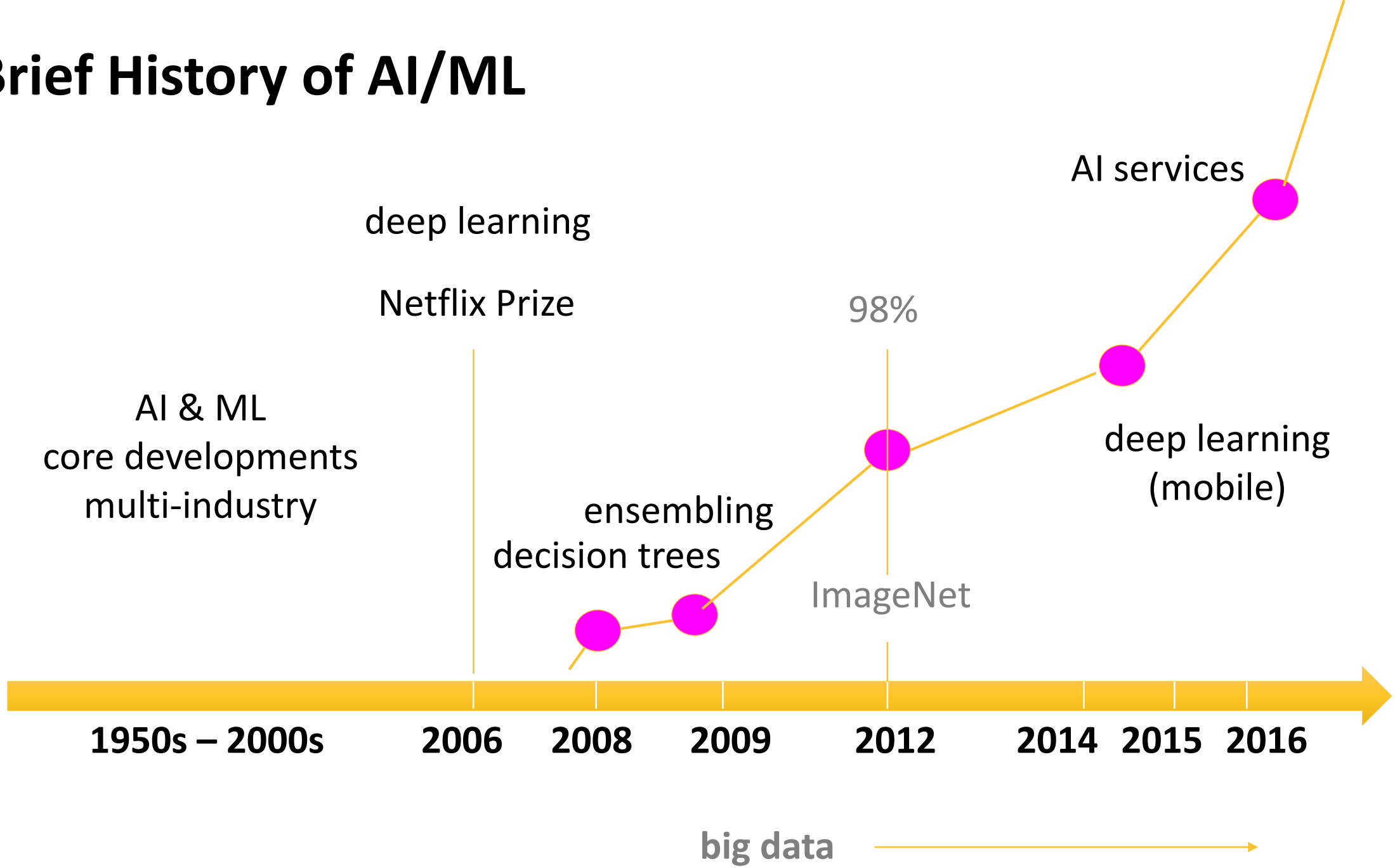
decisions

autonomy



DEEP LEARNING (DL) A powerful, popular machine learning technique loosely brain-inspired.

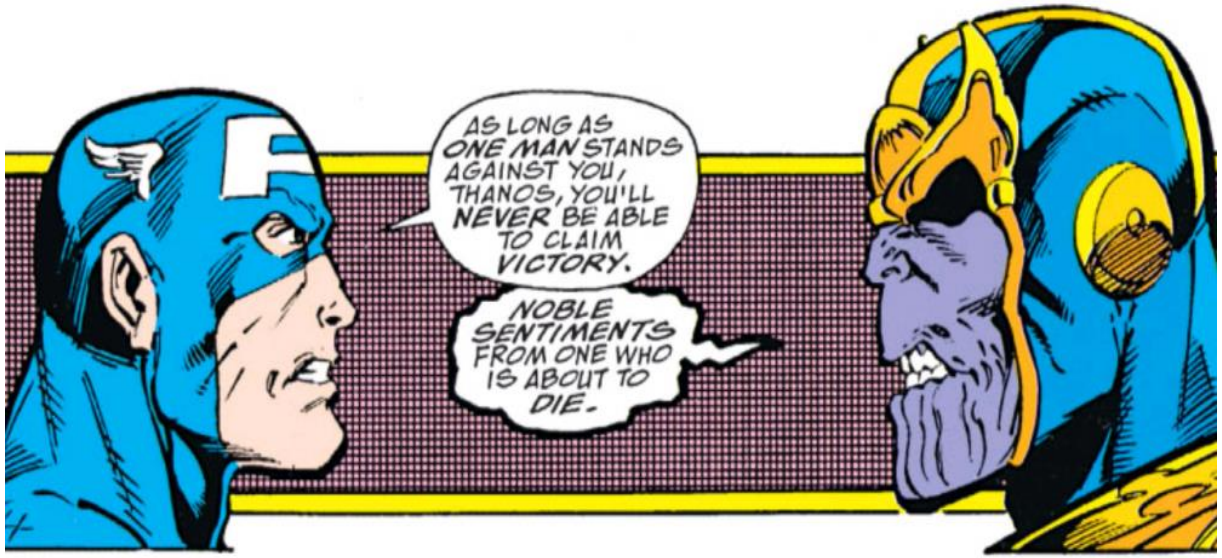
A Brief History of AI/ML



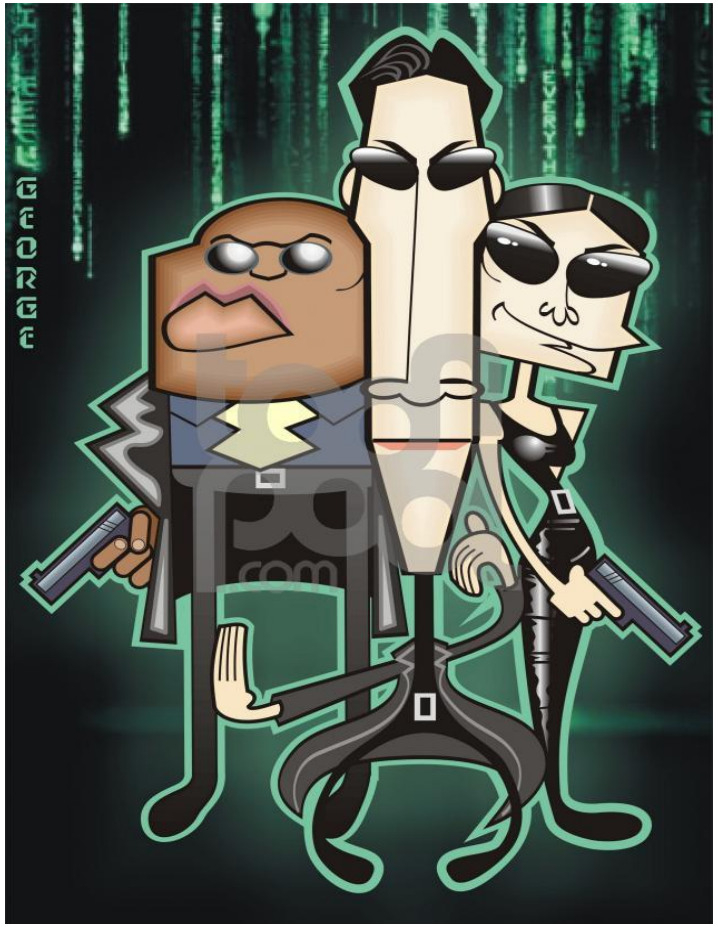
The AI/ML Security Problem: a Two-Sided Tale

1. AI/ML to protect us from bad things.

... by doing magic with data.



Avengers Infinity Wars. (c) Marvel Entertainment Group



Matrix caricature. (c) 2008 George Williams

2. Protect our AI/ML.

... from malicious data.

Magic with Cybersecurity Data is Hard

security data != cat videos

- Complex sequential data
- Not human-intuitive
  - What should a program trace or log file look like?
- Strong domain knowledge required
- Scarce | expensive labels
- Closed research models
- ➔ Slower to advance AI/ML

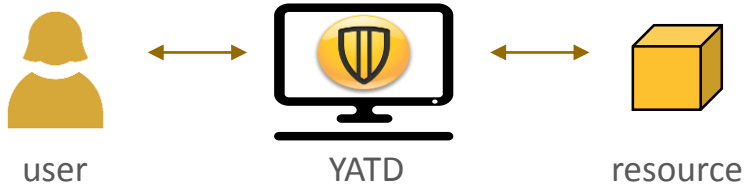
good or bad log?

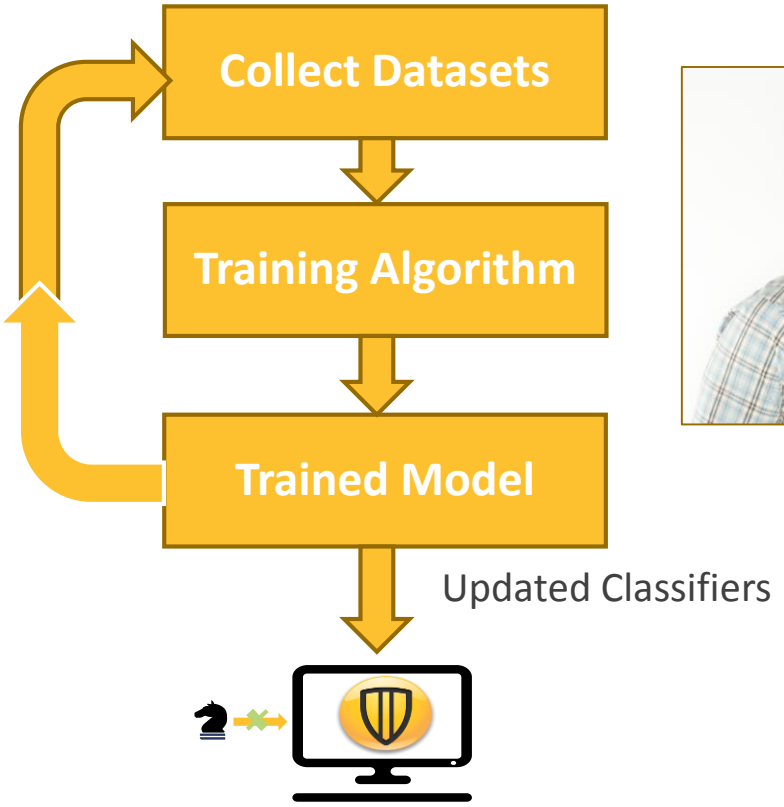

billionaire or cat?

How are ~~AI~~/ML Used in Security Today?

Yet Another Threat Detector (YATD)

- Straightforward recipe
- Data driven
- Debate about techniques
- Rely on data scientists
  - Feature engineering
  - Updates & tweaks





"Frankie"

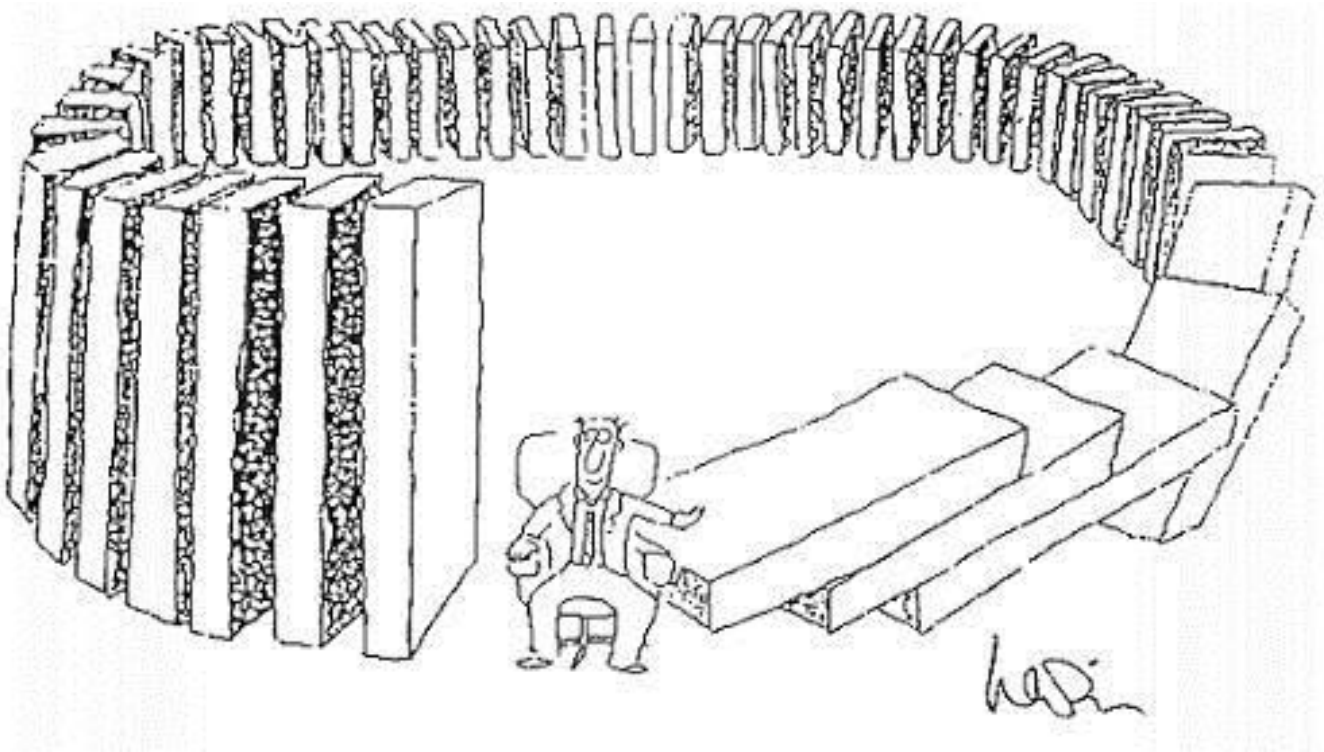
The AI/ML Manifest Destiny for Security

1. AI/ML is value neutral and will therefore result in good outcomes for everyone.
2. AI/ML should be deployed as quickly as possible, even if we don't fully understand it or its societal impacts.
3. History is uninteresting because the past has nothing to teach us.
4. AI/ML will give us new magical powers for free, like "automagic UEBA anomaly detection," for better threat protection.
5. AI/ML is all about automation and scaling.
6. More data will make it better.
7. AI/ML tilts the game in favor of the good guys.

a

(a) inspired by James Mickens 2018 Usenix Keynote, "Q: Why do keynote speakers keep suggesting that improving security is possible? A: because keynote speakers make bad life decisions and are poor role models."

What Could Possibly Go Wrong?

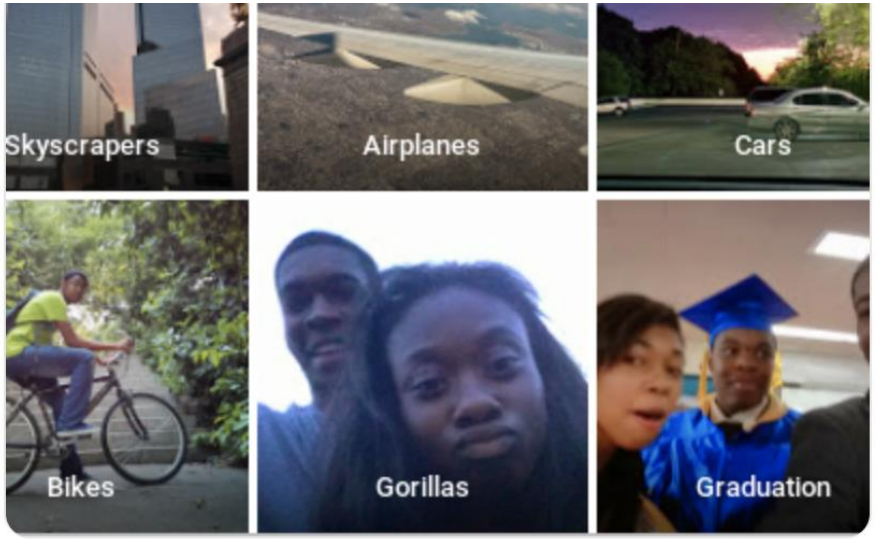


AI is Even More Challenging!

learning

Silicon Valley S4E4. (c) 2017 HBO.

decisions



Google Photo results tweeted 28 Jun 2015 (Jacky Alciné)

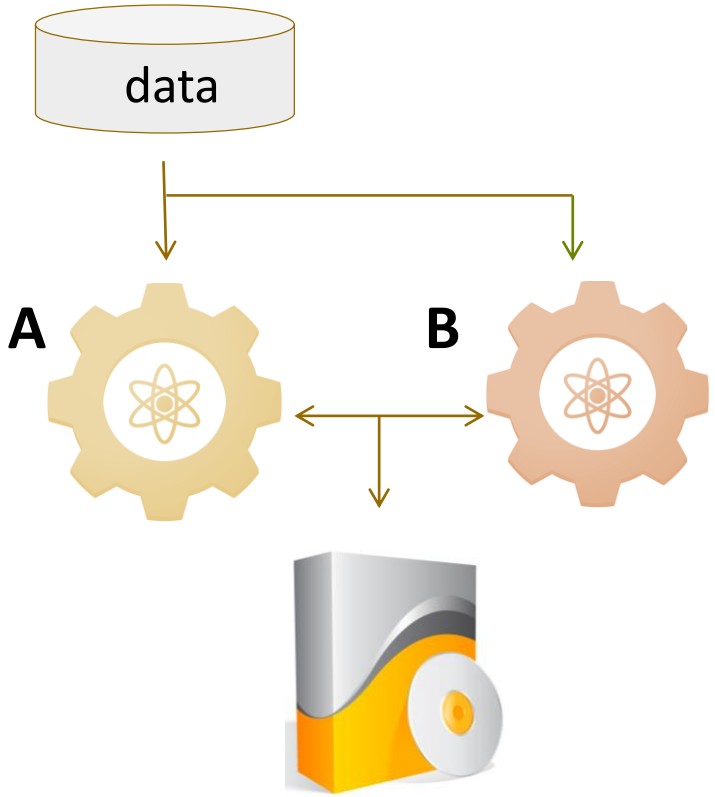
Microsoft Tay chatbot tweets Mar 24 2016 (Gerald Mellor)

autonomy

AI/ML Downsides – Unintended Side Effects

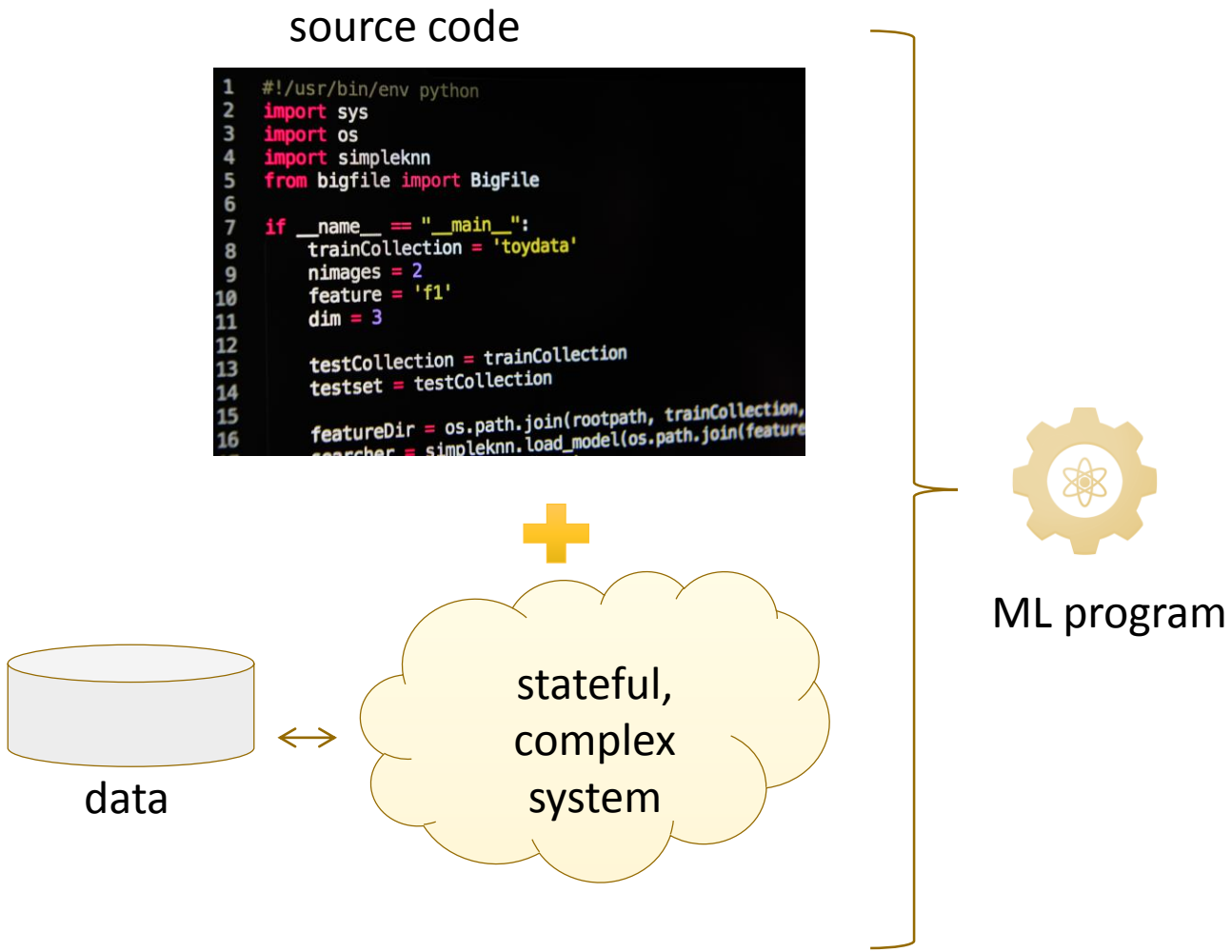
Poor architecture & unintended side effects

- Detectors A & B independent
- New system introduced
  - creates feedback between A/B
  - inadvertent, unknown?
- New sample arrives:
  - A → 2/10, B → 1/10
  - … but B sees ΔA, B → 3/10
  - … but A sees ΔB, A → 4/10
  - … and so on



AI/ML Downsides – ML Technical Debt

- Traditional software
  - Source code → program
- Machine learning software
  - Source code + **data** → program
  - Data are embedded and opaque
  - Reconstruction is hard
  - ML data versioning is hard
  - -> data and system dependencies

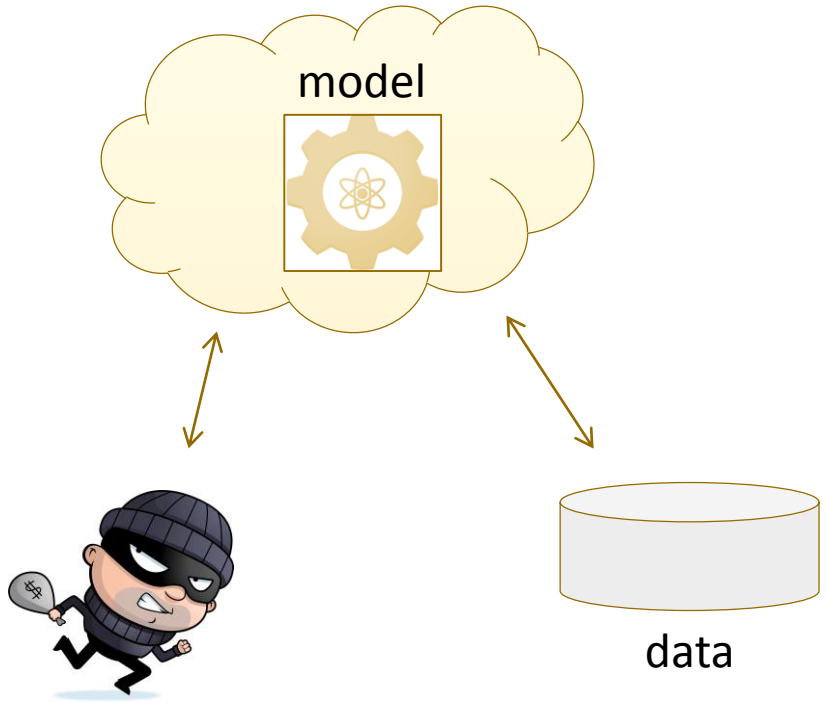


D. Sculley et al., "Machine Learning: The High Interest Credit Card of Technical Debt," https://ai.google/research/pubs/pub43146 (2014)

Adversaries Have AI/ML, Too!

Adversarial Machine Learning

- Model extraction
  - Adversary learns an approximate model using fewest possible queries
- Poisoning
  - Adversary biases machine learning model through interaction
- Adversarial examples
  - Crafting inputs to defeat ML.



AI/ML becomes a threat surface

Adversarial Samples Are Fundamentally Bad!

ADVERSARIAL SAMPLES

panda

this is normal

≠

"fake" panda

this is easily adversarially crafted to fool a classifier

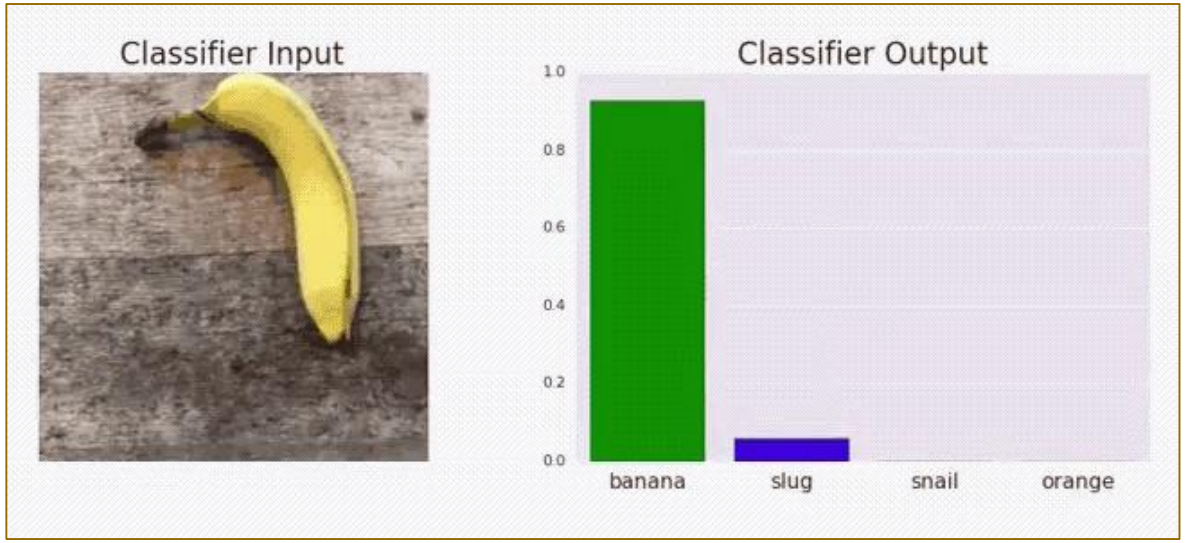
≠

gibbon

this is what the classifier incorrectly "detects"

Image credit: The Karate Kid (1984)
(c) Columbia Pictures

"No can defend!?"



- ML can be "easily" fooled***
- ... without access to the model
- ... without access to the training data
- for some types of data & models

Video credit: https://www.theverge.com/2018/1/3/16844842/ai-computer-vision-trick-adversarial-patches-google

Man or Machine? Advanced Behavioral Attacks

#deepfakes #lyrebird

- Imagine a business email compromise attack
  - you get an email to wire payment for an invoice from the CFO
- The email is written “from” your CFO
  - natural language processing from emails
- You’re suspicious and call the CFO
- But your phone (or call) is compromised
- You’re connected to an adversary who has a speechbot with your CFO’s voice
- Science fiction or possible today?

➔ **best defense:** multifactor telemetry?

Microsoft Real-Time Translation (2012)

https://www.youtube.com/watch?v=Nu-nlQqFCKg

Rethinking Security

BEFORE – Protect from bad.

- Limited, understood contexts
  - Users
  - Systems
  - Data
  - Controls
- Rigidity
- Focus on the bad

NOW – (understand) & do the right thing.

- Sophisticated, not-understood contexts
  - Hyperconnectivity
  - Unboundedness of HCI
  - Mimic human behavior
- Flexibility
- => we underestimate risk
- => focus on the possible

terrestrial

Ok to feed cute furry things.
Expected behaviors.

extraterrestrial

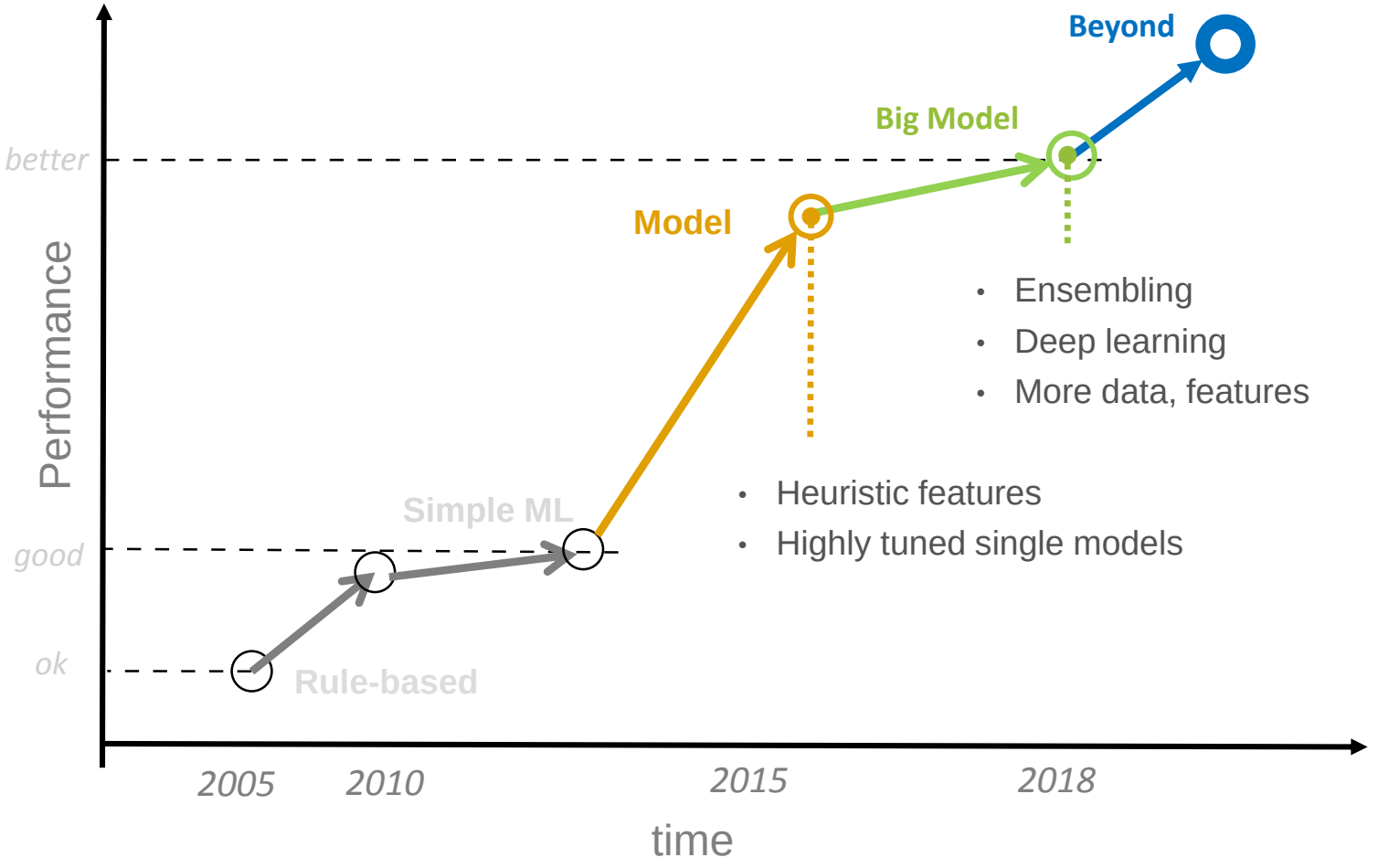
Don’t feed cute furry things.
Unexpected problems.

AI enables unfamiliar, unpredictable new contexts so we must think carefully.

Some Things Go Right: AI/ML Security Use Cases

- **Messaging – spam, BEC, phishing**
- Web – domain abuse, content filtering, bots, fraud
- **Identity, privacy and fraud**
- APTs, targeted attacks, intrusion protection
- Anti-malware, antivirus, isolated execution, detonation & quarantine, ransomware
- Policy adaptation and patch management
- Data loss prevention and information security
- Encryption
- Data center and cloud
- **Zero trust and personalized, contextual models**
- **Fake-stuff**
- **Detecting and preventing adversarial ML**

AI/ML Value: Beyond Detection Efficacy



ENDPOINT … OR ANY DETECTION

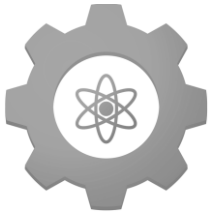
beyond efficacy

- Performance
- Compliance
- Robustness
- Enrichment
- Usability
- Autonomy

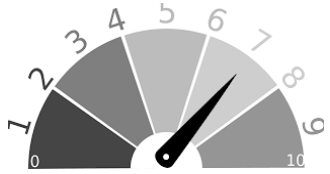
Example of Enriched Detection

Static Malware Detection with Attribution

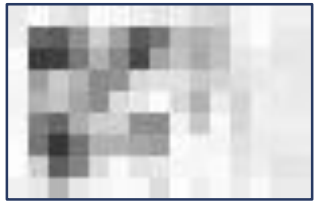
raw file



DL/ML

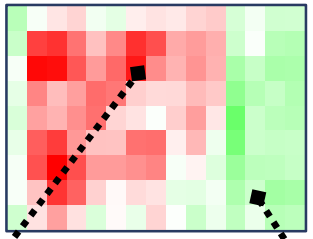


traditional score

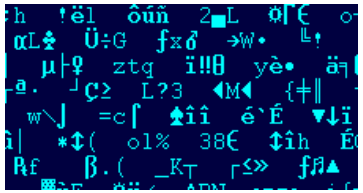


attribution - why

Learns Like Analysts

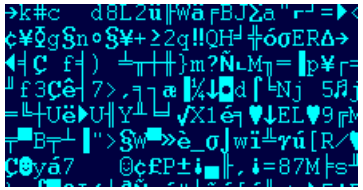


a malicious file
0.975

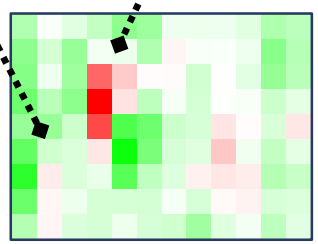


Encrypted data with padding often malicious.

“Normal” x86 code can be clean or malicious.



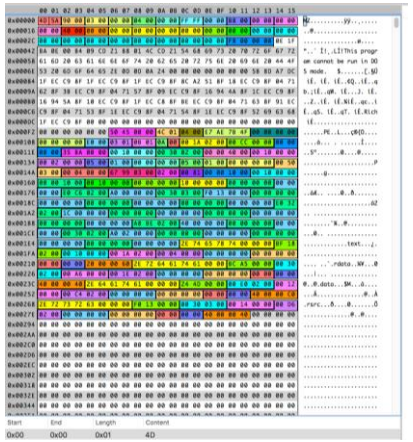
Encrypted strings more likely in malicious files.



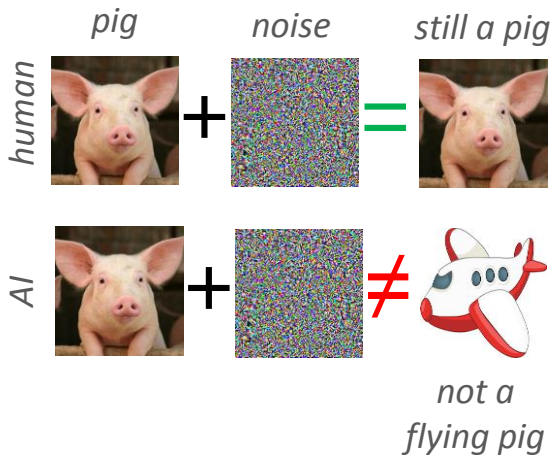
a clean file
0.011

Going Further By Tracking AI/ML Futures

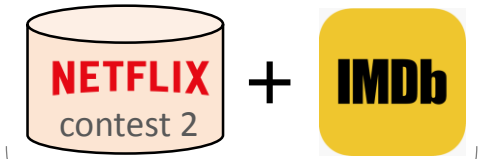
Representation Learning	Adversarial ML	Responsible ML	Relational Learning
Q: How do we represent **security data** for ML/AI?	Q: Can we detect / defend attacks **against our ML**?	Q: Is our AI/ML **unbiased, privacy-compliant** and **ethical**?	Q: How to leverage **data structure** and **expert knowledge** with ML?
ex: malware detection	ex: fooling computer vision	ex: leaking PII	ex: zero trust+ access



PE file in a hex editor



not a flying pig



identity from pseudonymized data

Thelma Arnold (identified)

What "textures" make this a clean / gray / malicious file?	*Why is AI/ML (sometimes) fooled so easily when a human isn't?*	*How do we use AI/ML without compromising privacy?*	*Is this user and context unusual, compared to others (UEBA)?*

Closing Questions

1. How important is AI/ML for Cybersecurity?
2. What should be regulated or controlled– why, how?
3. How do we measure the value of AI/ML systems?
4. How do we understand or interpret AI/ML systems?
5. What amazing ~~things can AI/ML do for you~~ risks do AI/ML pose?

Thank you!

Andrew B. Gardner, Ph.D.

For follow-ups:

Andrew B. Gardner, Ph.D.
Senior Technical Director, Head of AI/ML
Center for Advanced Machine Learning (CAML)
Symantec Corporation

andrew_gardner@symantec.com
@andywocky