AI & Cybersecurity

Separating Fact from Fiction

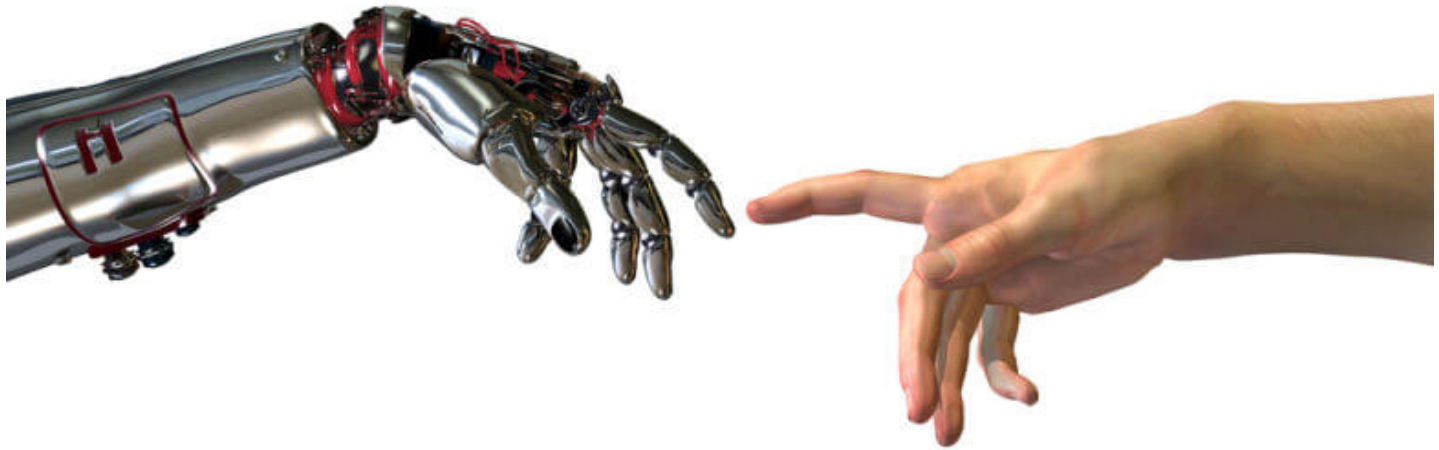
SECURITY INNOVATION

Agenda

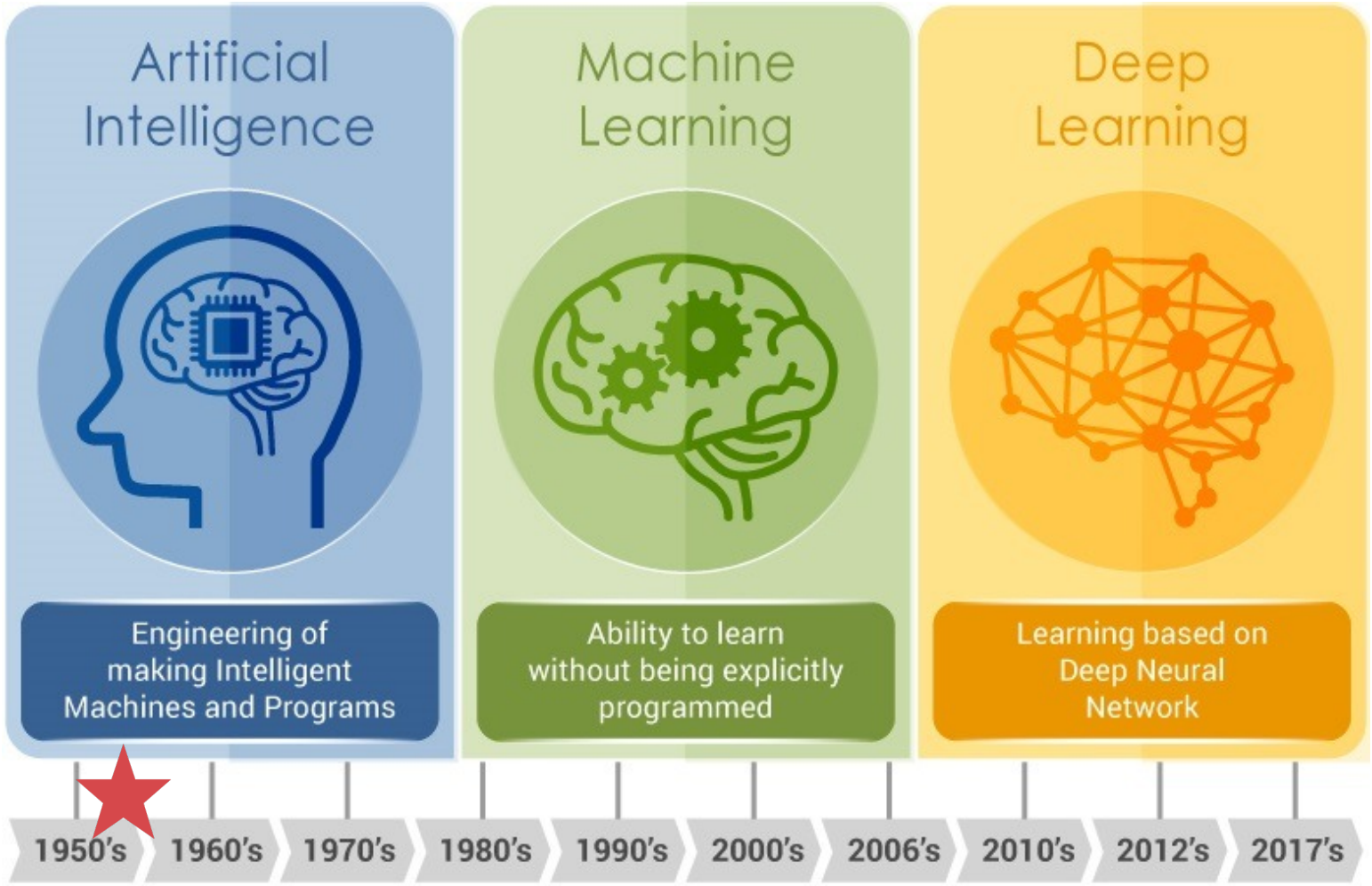
- AI 101
  - History, Overview, Real World Examples
- Defending
  - Spam Filters, Fraud Detection, Anti-virus
  - Shortcomings
- Attacking
  - Data Pipeline
  - Attacks, Tools, Remediations
- Future
  - Automation, AI Assisted Attacks, Singularity?

AI 101

Are you optimistic or pessimistic about the rapid advancements in AI?



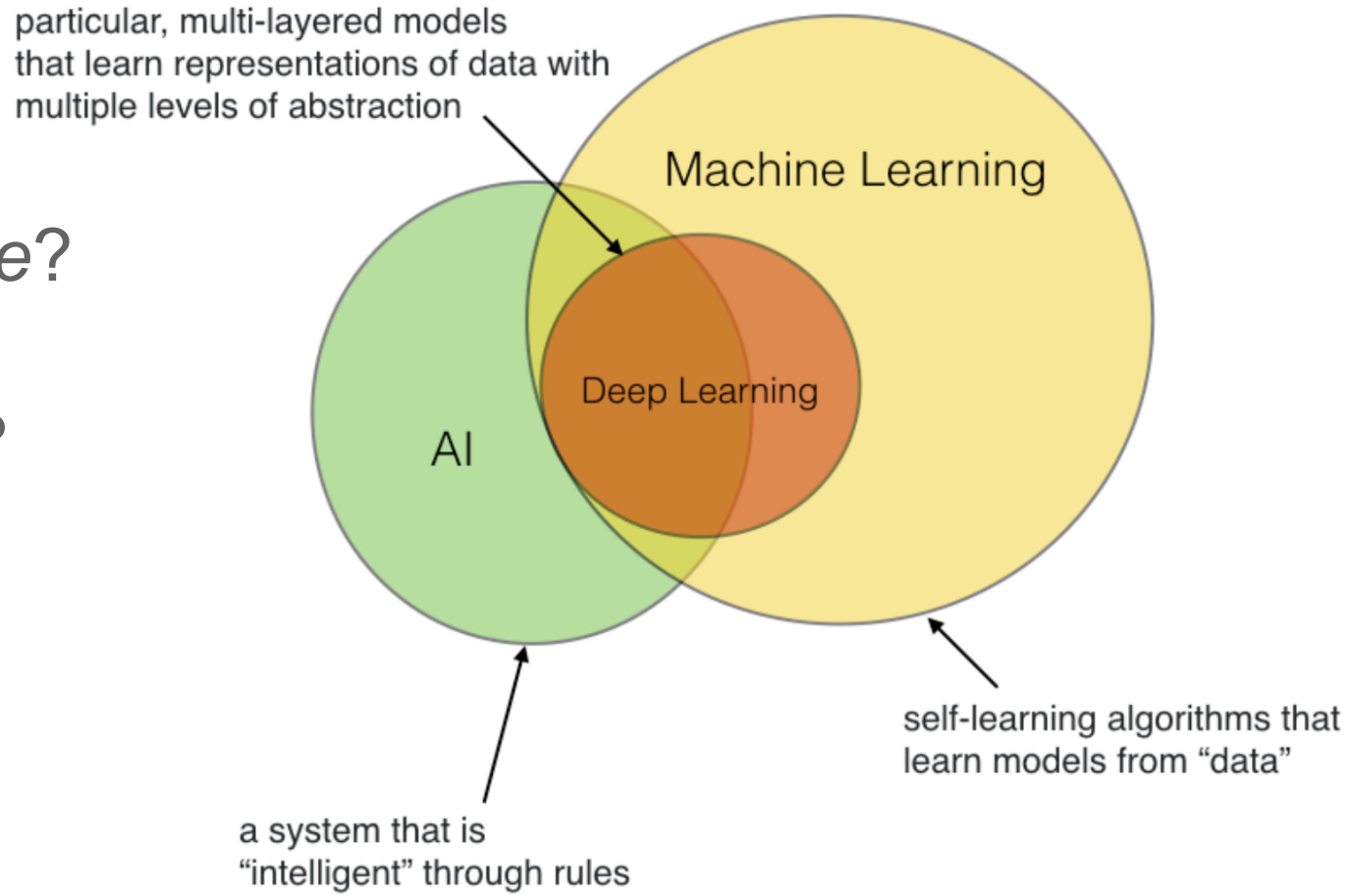
History



SECURITY INNOVATION

Overview

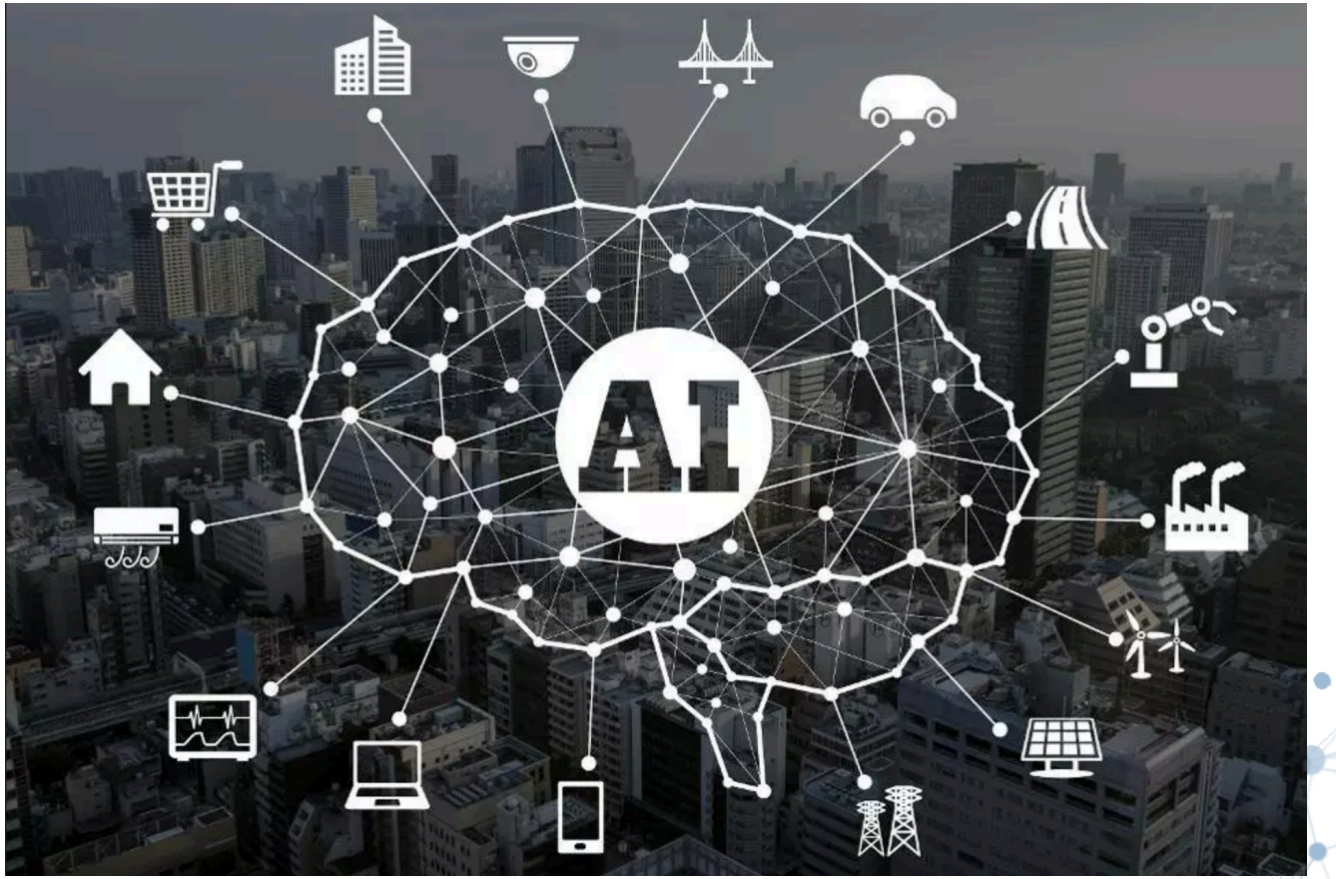
- What is *Artificial Intelligence*?
  - A science
- What is *Machine Learning*?
  - An approach
- What is *Deep Learning*?
  - A set of techniques



Real World Examples

AI is transforming industries

- Autonomous vehicles
- Recommendation engines
- Facial recognition
- Fraud detection



Does AI really matter for cybersecurity?

Detection

Automation

Scalability

Accuracy

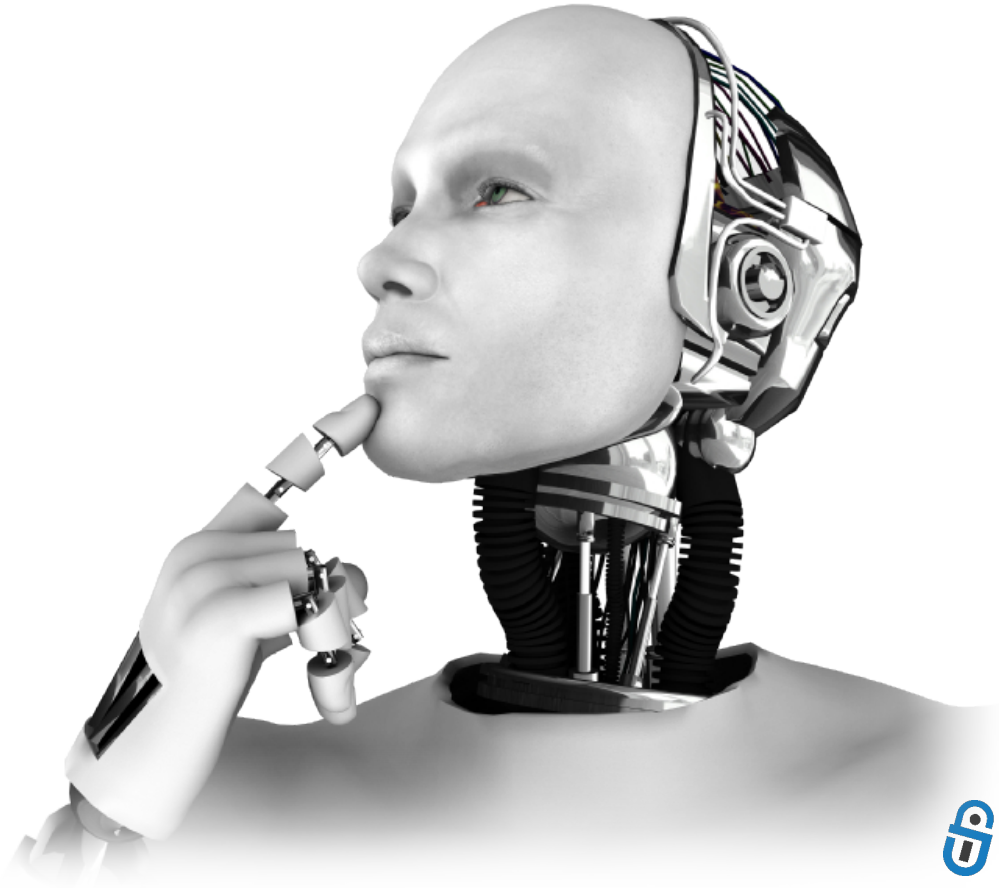
Prevention

Analysis

Defending

Fact or Fiction

Can AI predict security incidents?



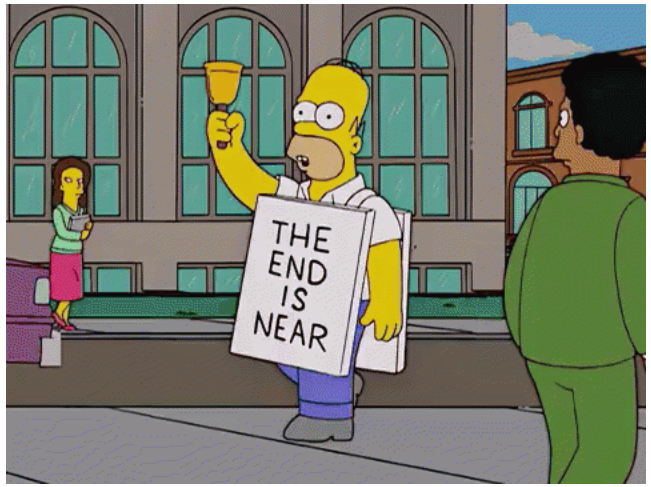
Getting Defensive

Even though AI can't predict cybersecurity incidents, it is being used in a wide variety of contexts that can help defend an organization.

- Spam Filtering
- Fraud Detection
- Malware Detection & Anti-Virus
- IDS/IPS
- Vulnerability Management

Fact or Fiction

As a cybersecurity professional, AI is on the verge of taking my job and potentially cutting off my head?



SECURITY INNOVATION

AI: Augmented Intelligence

AI is changing cybersecurity staffing requirements due to improved efficiencies in patching networks, eliminating false positives, and investigating vulnerabilities.

Labor spent containing cyber exploits (hours/week)	Not facilitated by AI	Facilitated by AI	Difference (% reduction)
Time wasted by security staff members chasing erroneous or false positives	400.83	41.42	**89.67%**
Cleaning, fixing and/or patching networks, applications and devices (i.e. endpoints)damaged/infected by cyber exploits or malware	212.89	39.63	**81.38%**
Investigating and detecting application vulnerabilities	195.88	70.48	**64.02%**

* Study conducted by Ponemon Institute and further information can be found here

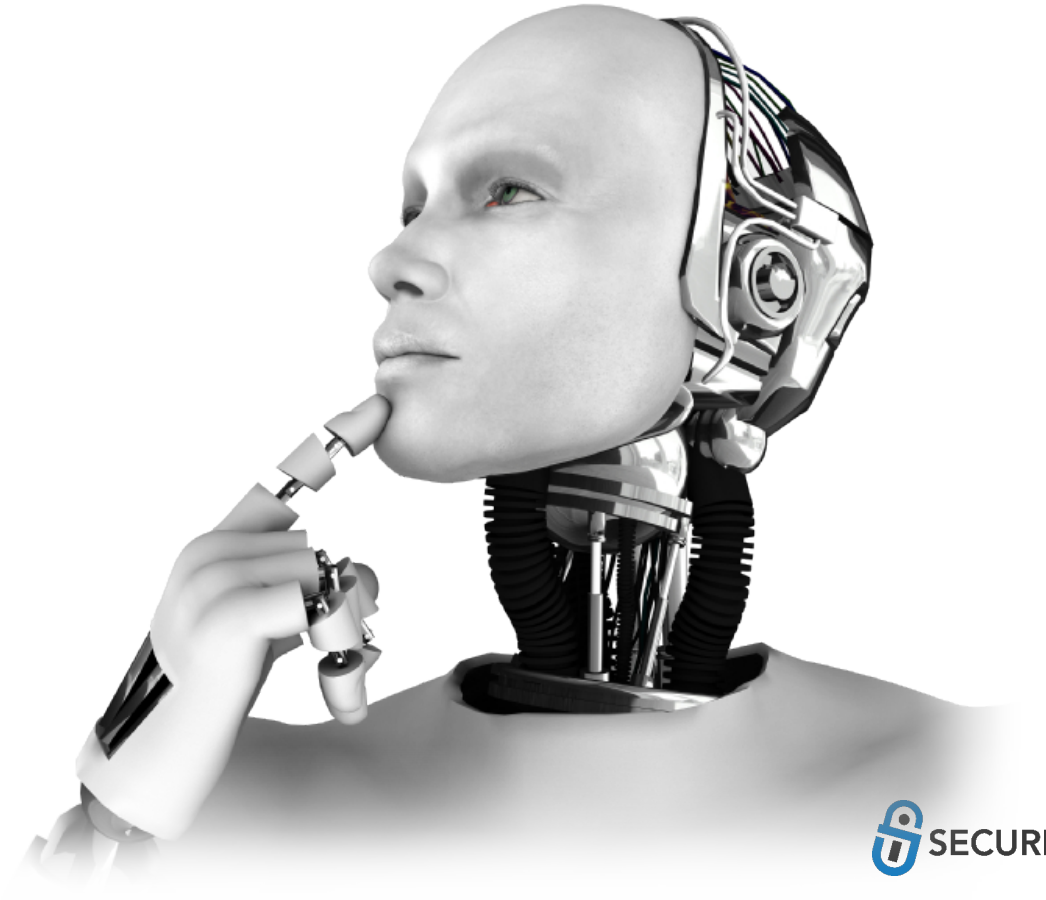
Fact or Fiction

As an organization, AI or even any automated service is not a cybersecurity silver bullet?

Attacking

Fact or Fiction

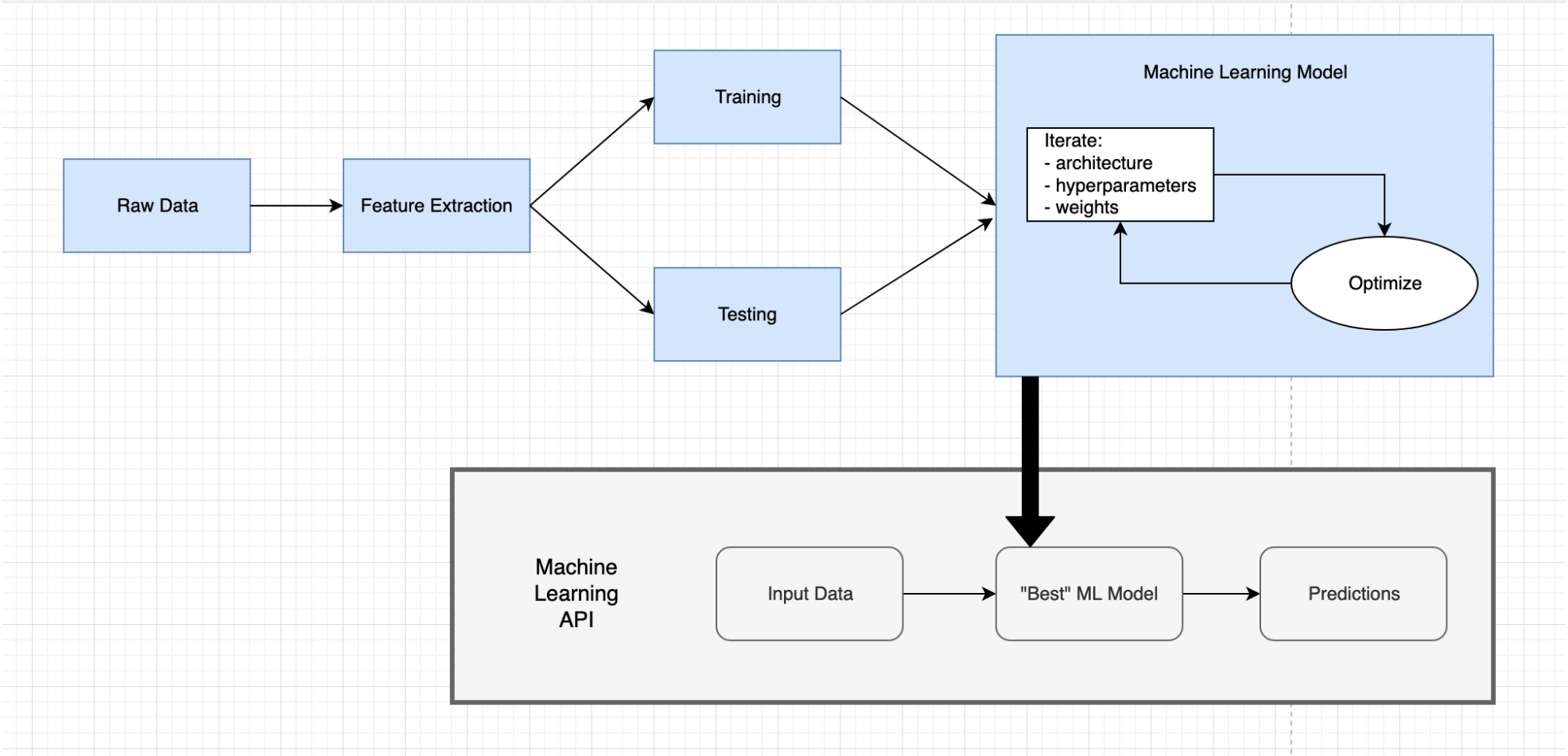
It’s not possible to hack an AI model?



Wait.. AI can be hacked? Yes!

- Autonomous vehicles
  - Attacking lane recognition function to drive cars off highway
- Recommendation engines
  - Reverse engineering an algorithm to extend reach of content
- Facial recognition
  - Wearing special glasses to misidentify the wearer
- Fraud detection
  - Modifying a check to bypass computer vision mechanism

ML Pipeline


Raw Data
Feature Extraction
Training
Testing
Machine Learning Model
Iterate:
- architecture
- hyperparameters
- weights
Optimize
Machine Learning API
Input Data
"Best" ML Model
Predictions

Attacks

Type	Description	Data	Model	Prediction	Output
Adversarial	Attempting to fool a model through malicious inputs			X	Force incorrect predictions
Data Exfiltration	Stealing data from output or other leakage	X			Steal data
Data Poisoning	Poisoning data to make the system misbehave	X			Force incorrect data predictions
Linkage Attack	De-anonymizing data by joining datasets or identifiers	X			Steal data
Model Poisoning	Feeding adversarial data to the classifier during training			X	Force incorrect predictions
Model Theft	Rebuilding a model or retrieving data that was used to train the original model		X		Steal model
Transfer Attack	Adversarial examples that affect one model often affect another model		X		Force incorrect predictions

Fact or Fiction

AI is the cutting edge and there is no way to prevent or remediate these kinds of attacks?

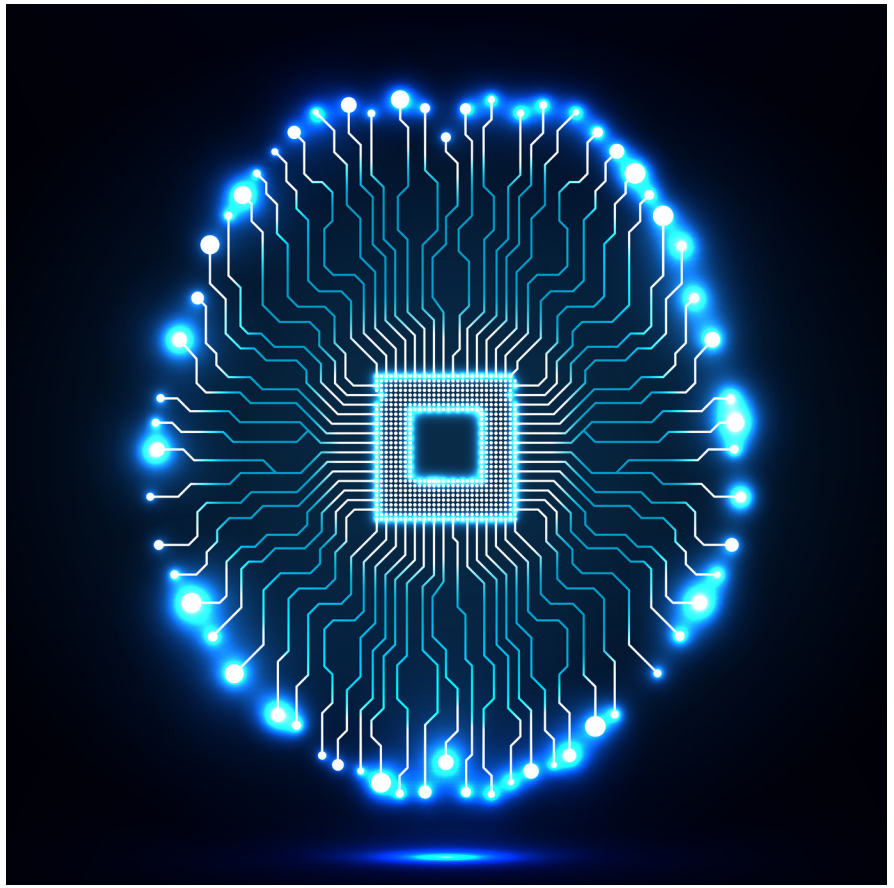
Remediations

- Model hardening
- De-noising
- Introspection
- Infrastructure attack simulation
- Cloud security configuration review

Future

Future

- Defense
  - Furthered automation of threat detection and response
  - User Entity Behavior Analytics (UEBA)
- Offense
  - Weaponized AI
    - Malware
  - AI assisted phishing



Thank You!

www.securityinnovation.com

bstewart@securityinnovation.com

SECURITY INNOVATION

Sources

- https://www.bing.com/images/search?view=detailV2&ccid=TbU%2bPIYp&id=E12D22719AD923B0F7F2BA6F25F5B2DF28721295&thid=OIP.TbU-PIYpZoni_IOPtwRb0QHaDX&mediaurl=https%3a%2f%2fmarkmanson.net%2fwp-content%2fuploads%2f2016%2f04%2frobot-and-human-hands-touching-780x356.jpg&exph=355&expw=780&q=artificial+intelligence+overlord+meme&simid=608035293549756810&selectedIndex=145&ajaxhist=0
- https://www.bing.com/images/search?view=detailV2&ccid=lN2zLZwX&id=73B8C81DB903643873F5BA7C4C097A4E4871ABCA&thid=OIP.lN2zLZwXC4nztkuki9WGKQHaE2&mediaurl=https%3a%2f%2fj3fau2wsnso1qkiarz7keon5-wpengine.netdna-ssl.com%2fwp-content%2fuploads%2f2018%2f05%2fHistory_of_artificial_intelligence.jpg&exph=512&expw=781&q=history+of+ai&simid=608009244594867021&selectedIndex=24&ajaxhist=0
- https://www.bing.com/images/search?view=detailV2&ccid=rBkRq%2fQ3&id=71E0A8197A4DAFD6969DDA18146B1EE9963CAA0D&thid=OIP.rBkRq_Q33br1FjnBbVt7AwHaHa&mediaurl=https%3a%2f%2fbellacaledonia.org.uk%2fwp-content%2fuploads%2f2012%2f05%2fyes2.jpg&exph=1024&expw=1024&q=yes&simid=608004966790136904&selectedIndex=3&ajaxhist=0
- https://www.bing.com/images/search?view=detailV2&ccid=MBEpMKv1&id=320B8FCFCA3E6A7E3804A0B82F4C711081A97B2B&thid=OIP.MBEpMKv17i44PST10or2pgHaFj&mediaurl=https%3a%2f%2fmedia.giphy.com%2fmedia%2feXo5eC1tK7cas%2fgiphy.gif&exph=360&expw=480&q=the+end+is+near+meme&simid=608007118546667172&selectedIndex=176&ajaxhist=0
- https://www.bing.com/images/search?view=detailV2&ccid=jZJAQZJZ&id=83E127B90CC971246EF4062C9458872D78DE90F5&thid=OIP.jZJAQZJZWqctgO_moEtbgAHaGo&mediaurl=http%3a%2f%2fwww.intensityanalytics.com%2fimages%2fslider%2frobot-thinking.png&exph=600&expw=670&q=robot+thinking&simid=608040374526151613&selectedIndex=3&ajaxhist=0
- https://www.bing.com/images/search?view=detailV2&ccid=qnGCOism&id=CA36A3157BD33E951F2EC9DFC4B734C100DC343E&thid=OIP.qnGCOismfhtAnvqrMom2yQHaE8&mediaurl=https%3A%2F%2Fi1.wp.com%2Fsecurityaffairs.co%2Fwordpress%2Fwp-content%2Fuploads%2F2017%2F11%2FArtificial-Intelligence.jpg&exph=640&expw=960&q=artifiicial+intelligence+cybersecurity&simid=608033644288869498&selectedindex=8&ajaxhist=0&vt=1