

**RSA**Conference2018

San Francisco | April 16 – 20 | Moscone Center

SESSION ID: SPO1-R12



#RSAC

# GENERATIONS OF AI IN SECURITY

## Homer Strong

Director of Data Science  
Cylance

## Colt Blackmore

Director of Product Management  
Cylance

# GENERATIONS OF AI IN SECURITY





**SAY AI AGAIN.**

**I DARE YOU.**

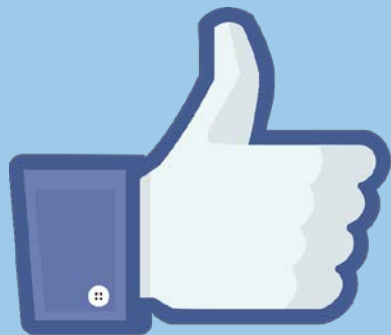
**I DOUBLE DARE  
YOU.**

**AI IS WHATEVER HASN'T  
BEEN DONE YET.**

—DOUGLAS HOFSTADTER









# CYBERSECURITY AI

link missing or broken

# AGENDA

- I. INTRODUCTION
- II. THE ART OF DATA CURATION
- III. MAN VERSUS MACHINE
- IV. SECURING SECURITY SYSTEMS
- V. CONCLUSION



# LEARN ...

- I. The present and future of AI security technologies
- II. How to evaluate the maturity of AI systems
- III. The risks and opportunities of AI for security

# I. THE ART OF DATA CURATION

or: garbage in, garbage out

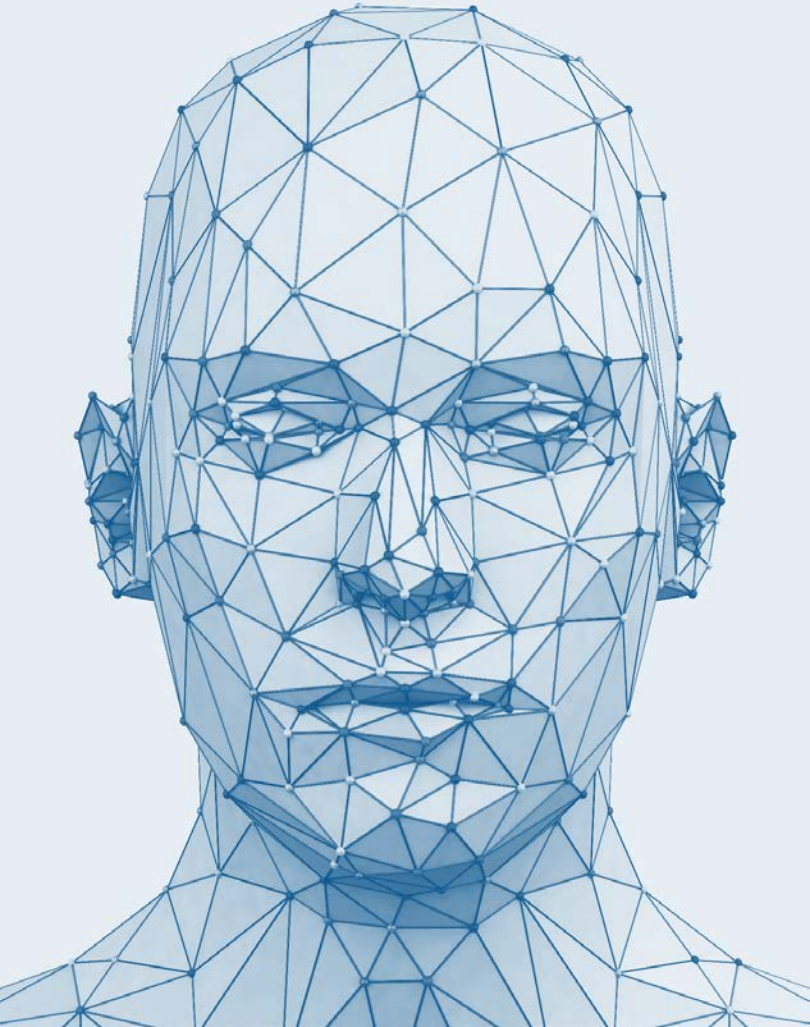


security is a problem of



SCALE

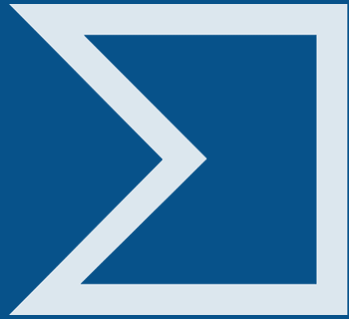




# IMAGE RECOGNITION



THE FIRST RULE  
OF DATA SCIENCE IS:  
YOU **DO NOT** TALK  
ABOUT DATA SCIENCE.



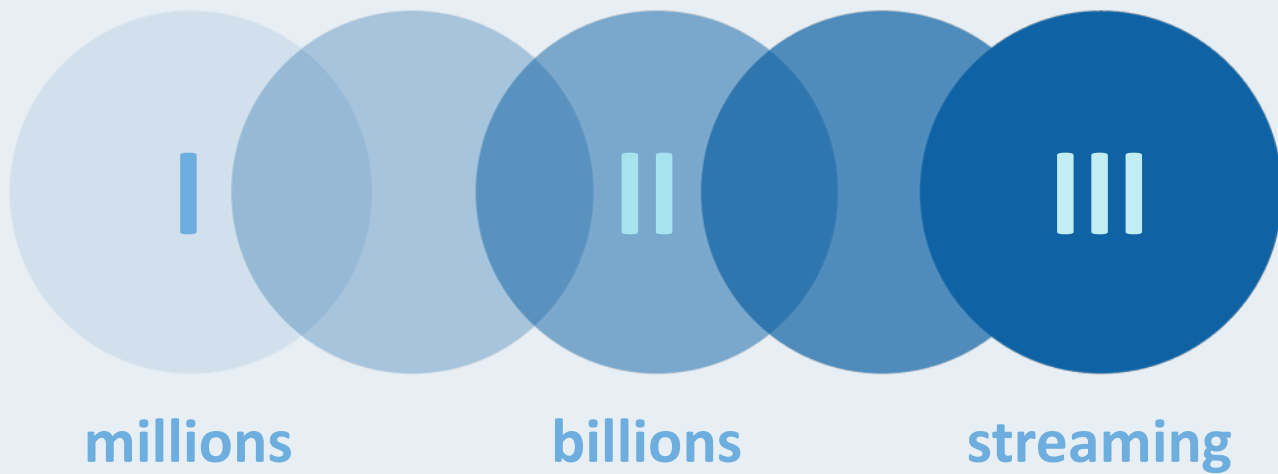
VirusTotal



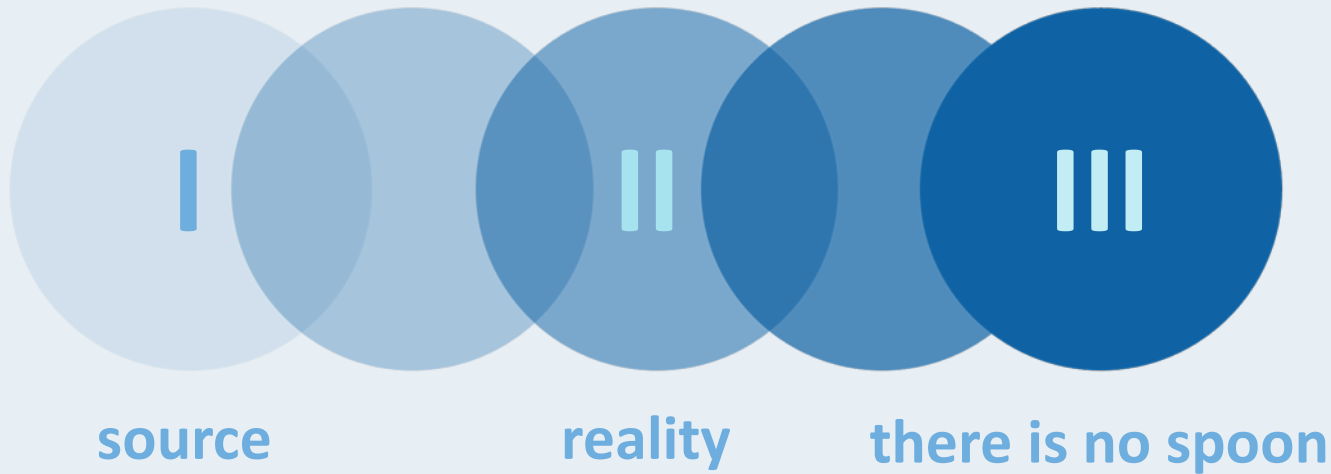
The image is a standard Emergency Broadcast System (EBS) test pattern. It features a grid of colored squares: a top row of white, yellow, cyan, magenta, red, and blue; a bottom row of blue, black, magenta, black, cyan, black, and white; and a bottom-most row of grayscale steps. A black horizontal bar is positioned across the middle of the grid. The text "EMERGENCY BROADCAST SYSTEM" is written in white, uppercase, sans-serif font on this bar. A white crosshair is centered on the green square in the top row.

EMERGENCY BROADCAST SYSTEM

# GENERATIONS: VOLUME



# GENERATIONS: DIVERSITY



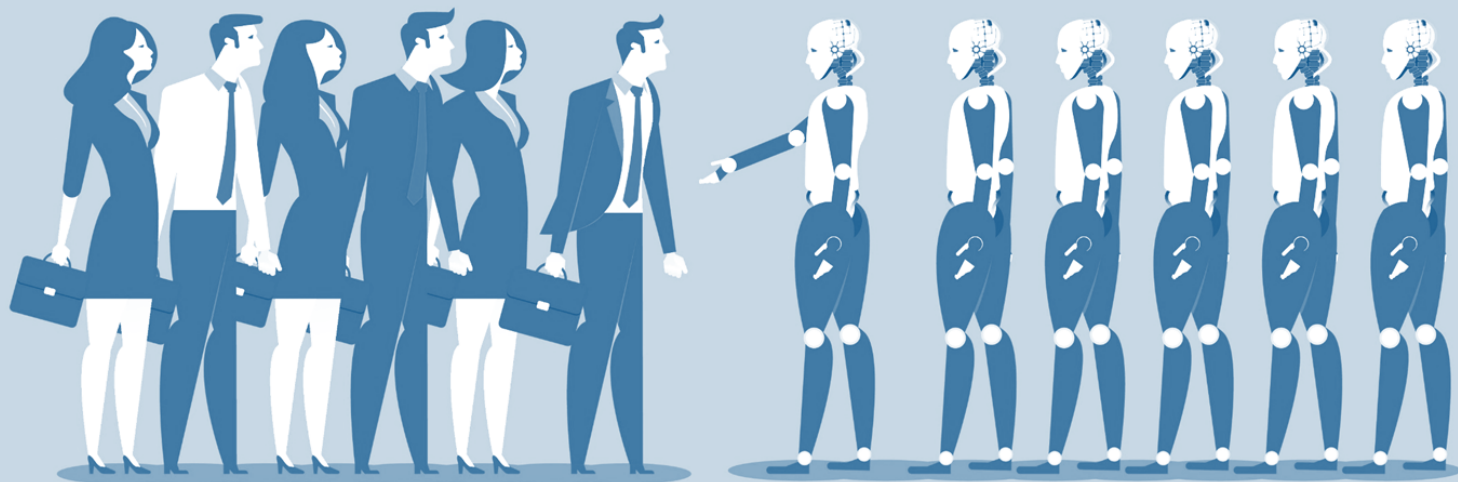
# GENERATIONS: HYGIENE



# II. MAN VERSUS MACHINE

## or: better, stronger, faster

(we have the technology)





X	fast	✓
X	scalable	✓
X	reliable	✓
X	cheap	✓





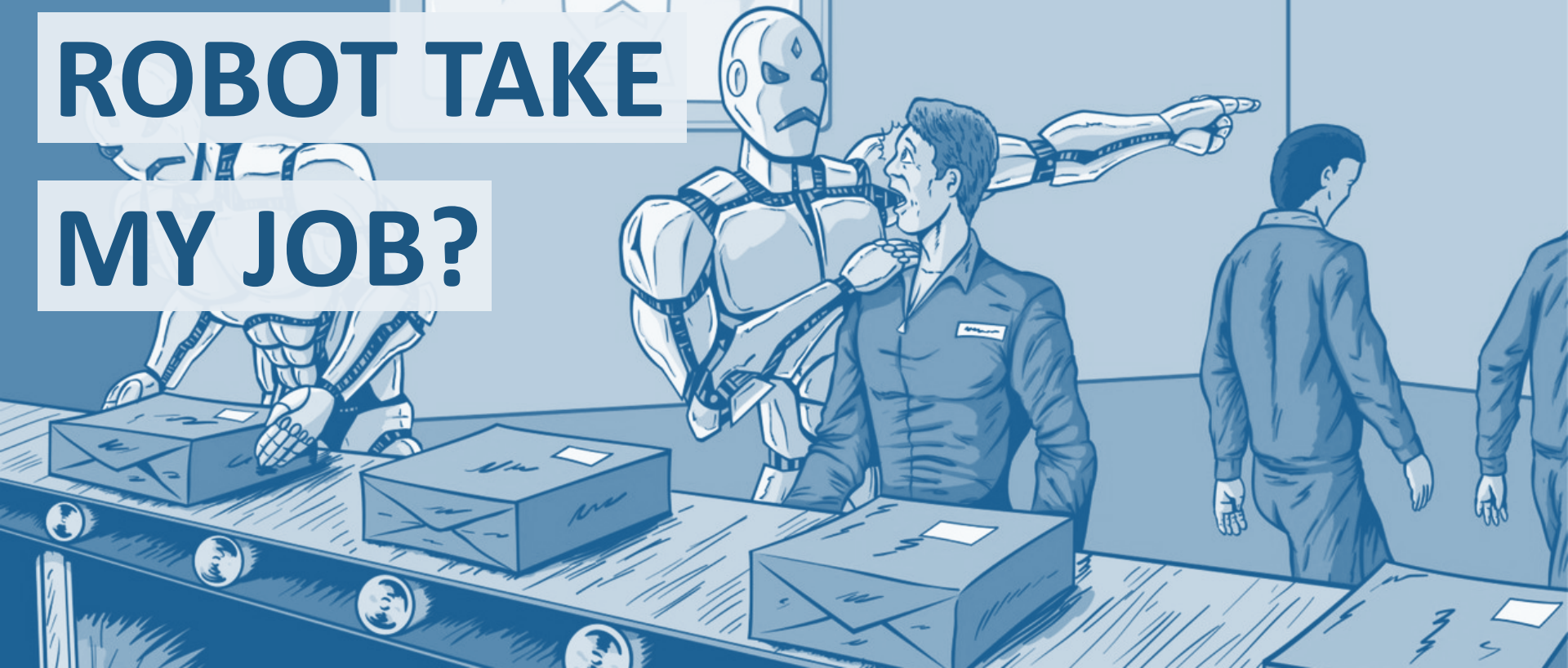
A stylized illustration on a blue background. On the left, a white robot head in profile is shown with internal mechanical details like gears and a circuit board. On the right, a white human head in profile is shown with dark blue hair. A dark blue thought bubble originates from the robot's head and contains a list of four traits, each preceded by a red 'X' and followed by a green checkmark.

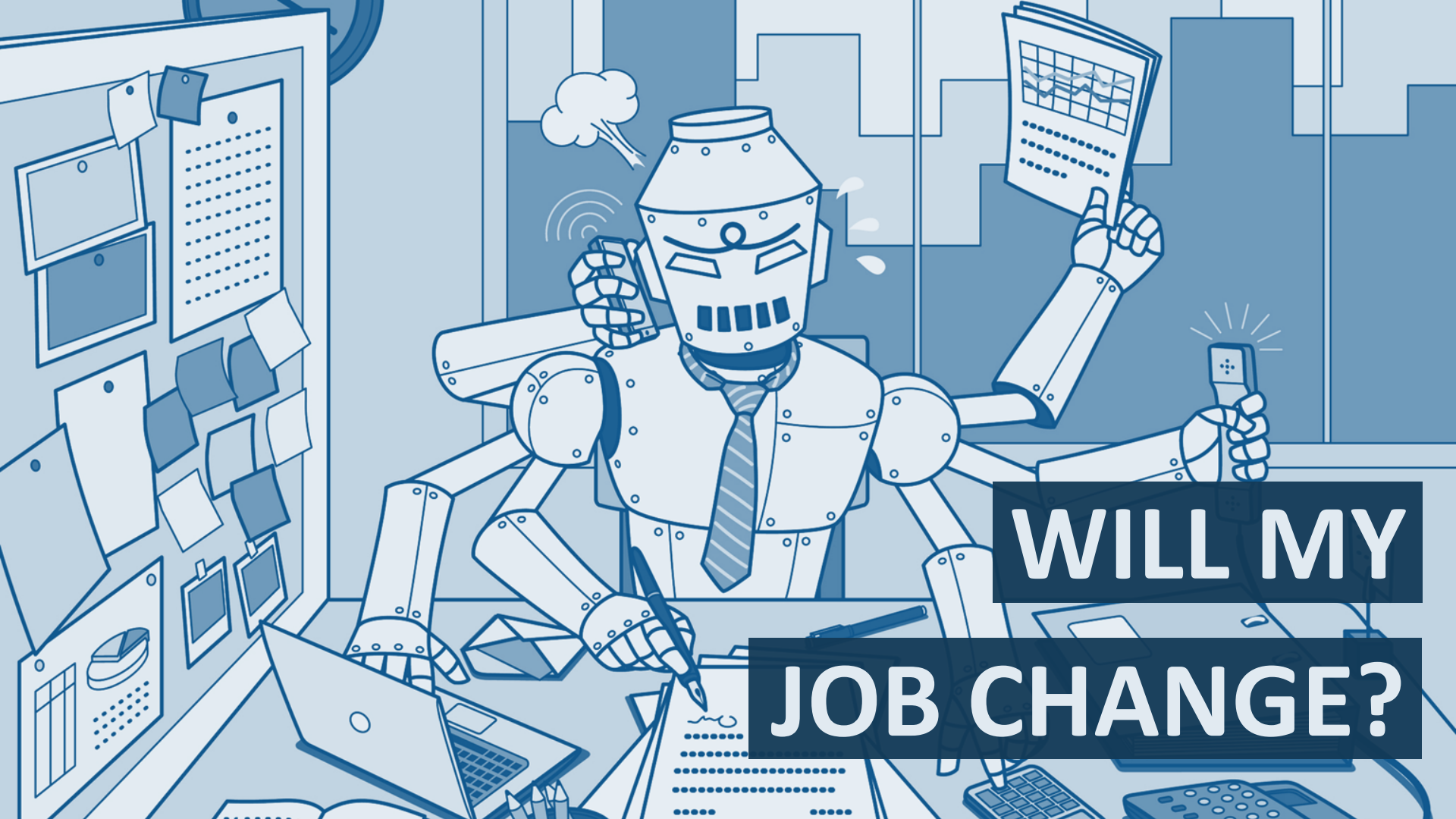
X	creative	✓
X	diverse	✓
X	daring	✓
X	ethical	✓

**WILL A**

**ROBOT TAKE**

**MY JOB?**





**WILL MY**

**JOB CHANGE?**

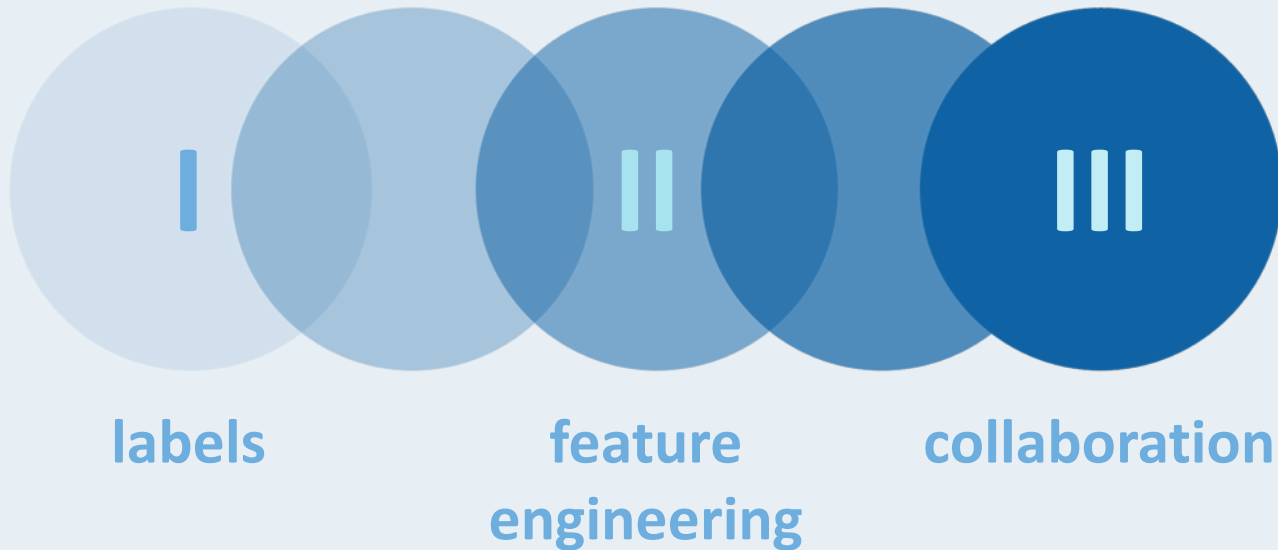
A group of colorful, blocky robots in a futuristic setting. The robots have large, square heads with circular eyes and small, rectangular mouths. They are standing on a metallic platform. The background is a dark, industrial environment with various mechanical parts and structures. The overall style is a vibrant, stylized illustration.

# HOW TO TRAIN YOUR AI

**“WHAT WE HAVE  
HERE IS A FAILURE  
TO COMMUNICATE.”**

—MAJOR PAYNE

# GENERATIONS: **AI** SYMBIOSIS





# III. SECURING SECURITY SYSTEMS

## or: who watches the watchers?



# FLAVORS OF ATTACK

poisoning



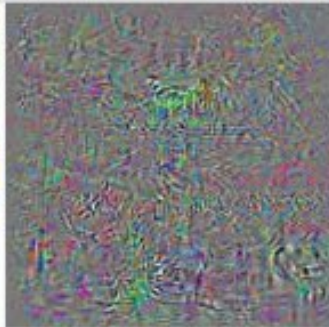
stealing



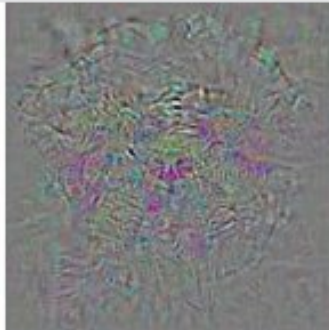
perturbation



# MASK: OSTRICH



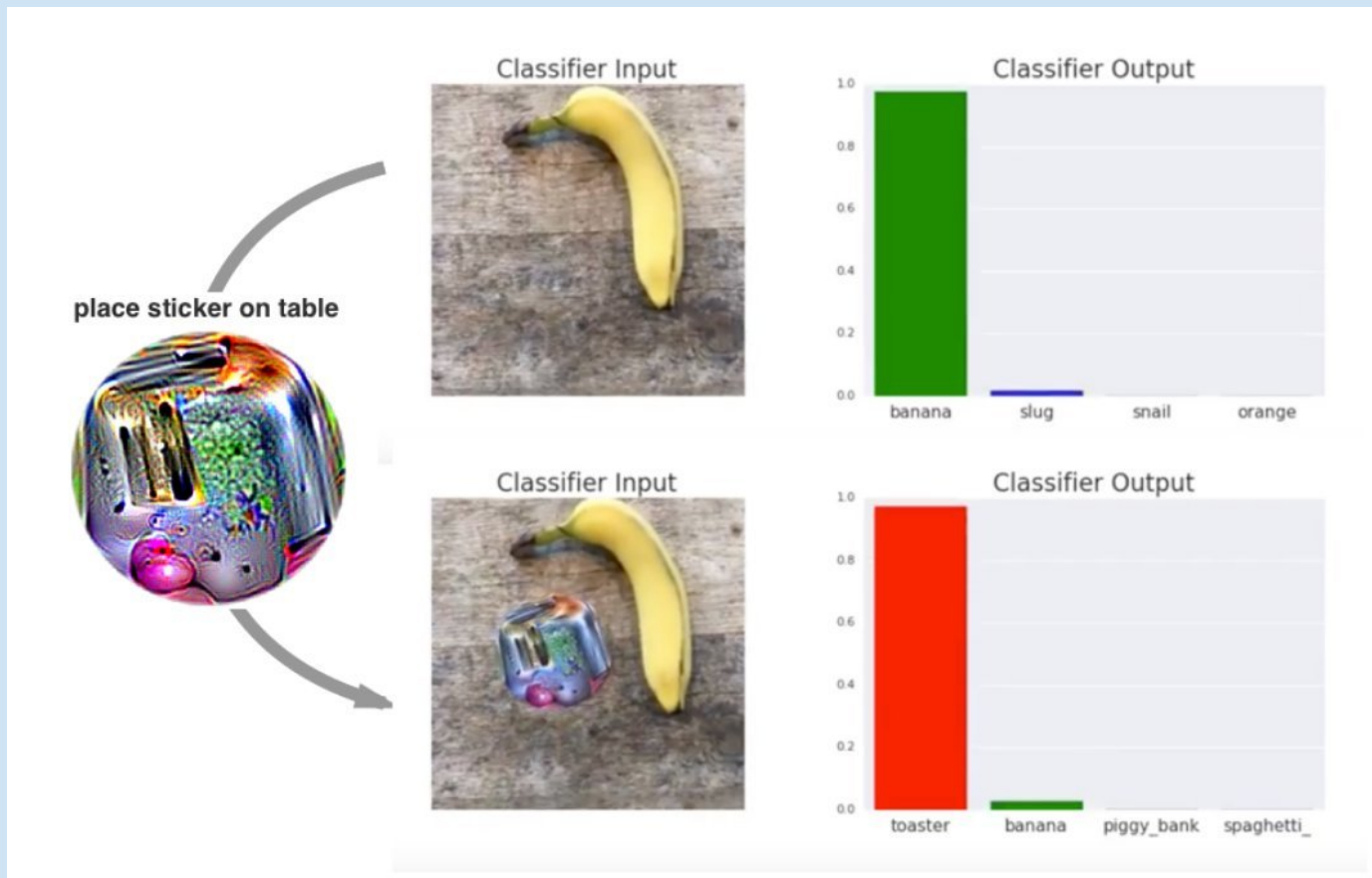
School bus + tiny adversarial perturbation = "ostrich"



Dog + tiny adversarial perturbation = "ostrich"

Adversarial input can fool a machine-learning algorithm into misperceiving images.

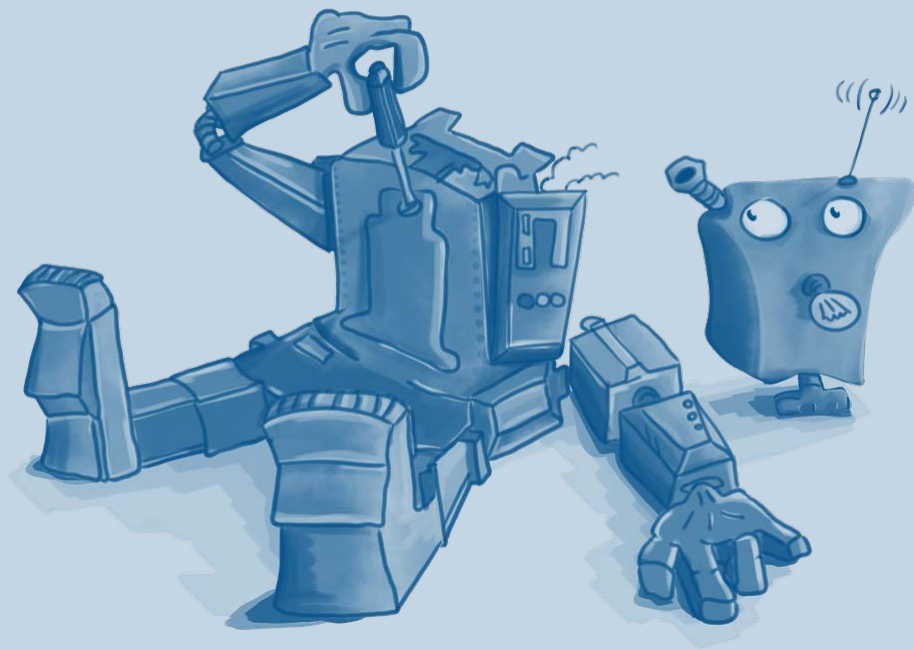
# PATCH: TOASTER



# TOY: TURTLE

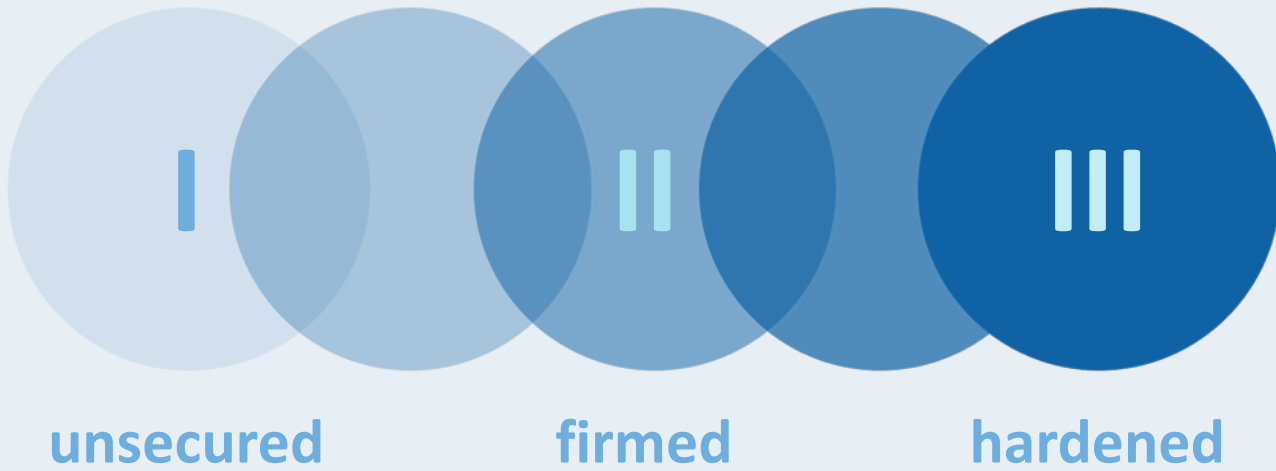


# IS SECURITY **AI** BROKEN?





# GENERATIONS: SECURING **AI**



# CONCLUSION: ASK ...

- I. Is the model really used to make decisions?
- II. How do you know when the model needs to be retrained?
- III. Who curates data?
- IV. What happens when the model is wrong?
- V. Who are the security experts who reviewed the features?
- VI. Who would notice if there were an attack?



**THANKS!**

Check out the whitepaper:  
[Generations of Machine Learning  
in Cybersecurity](#)