RSAConference2018
San Francisco | April 16 – 20 | Moscone Center

SESSION ID: SPO1-R12

GENERATIONS OF AI IN SECURITY

Homer Strong

Director of Data Science
Cylance

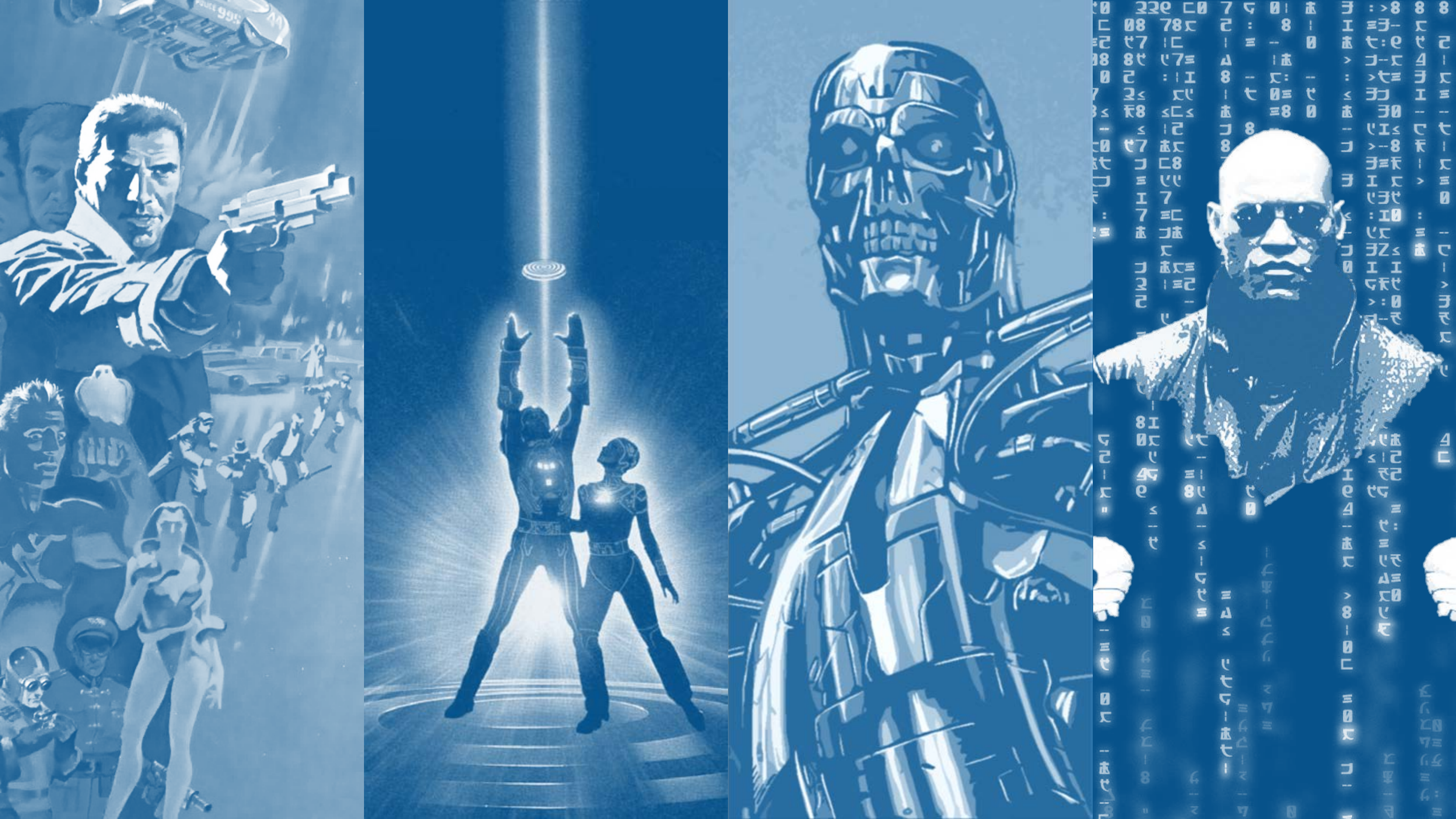
Colt Blackmore

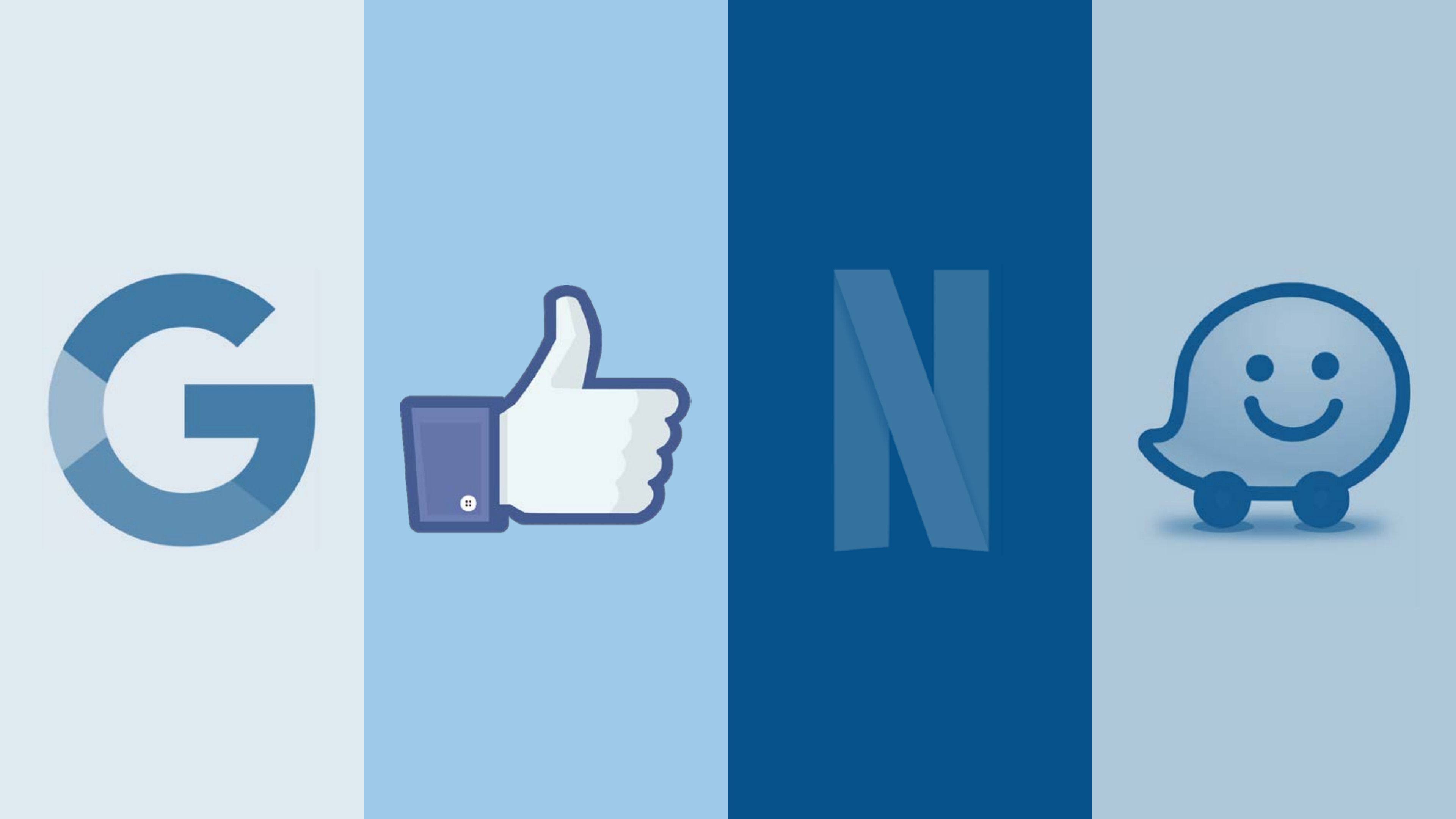
Director of Product Management
Cylance

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

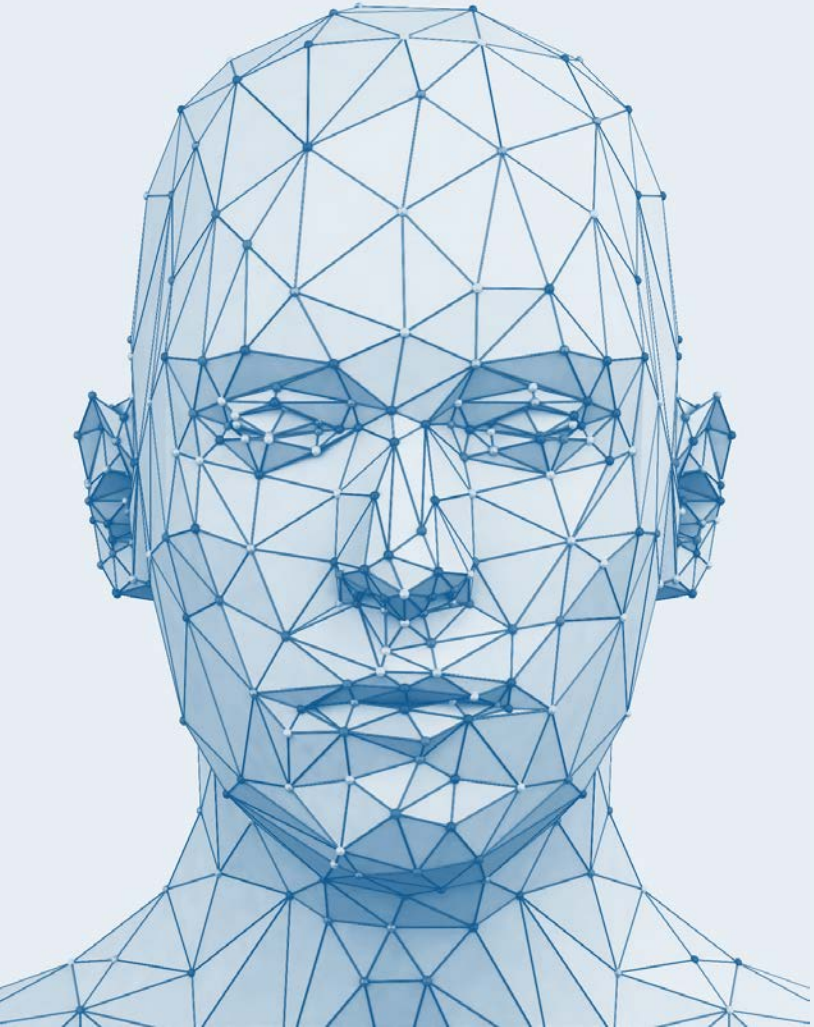


IMAGE RECOGNITION

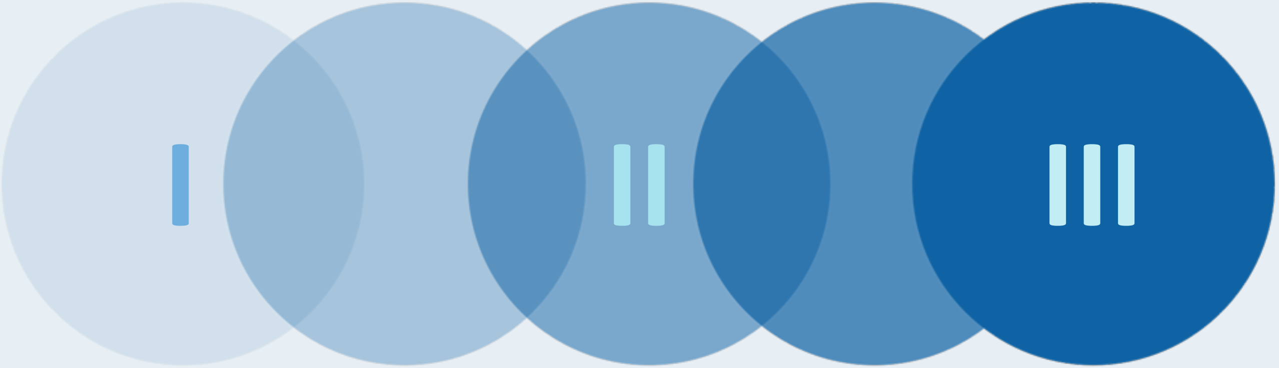
THE FIRST RULE
OF DATA SCIENCE IS:
YOU **DO NOT** TALK
ABOUT DATA SCIENCE.

VirusTotal

EMERGENCY BROADCAST SYSTEM

GENERATIONS:

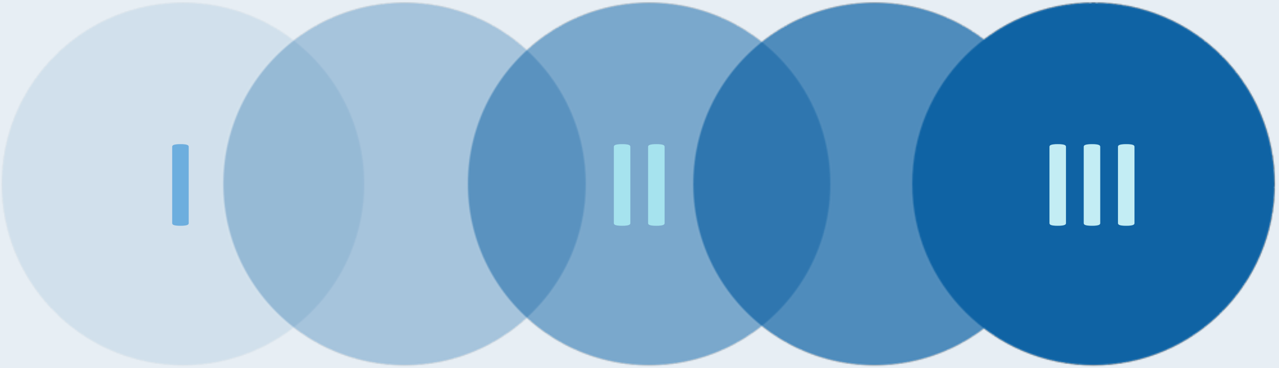
VOLUME



millions billions streaming

GENERATIONS:

DIVERSITY



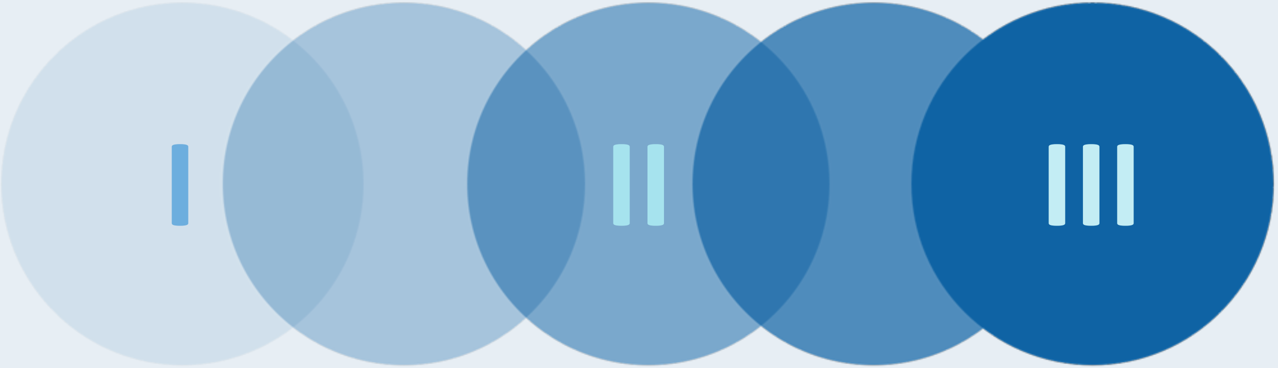
source

reality

there is no spoon

GENERATIONS:

HYGIENE



human, heuristic

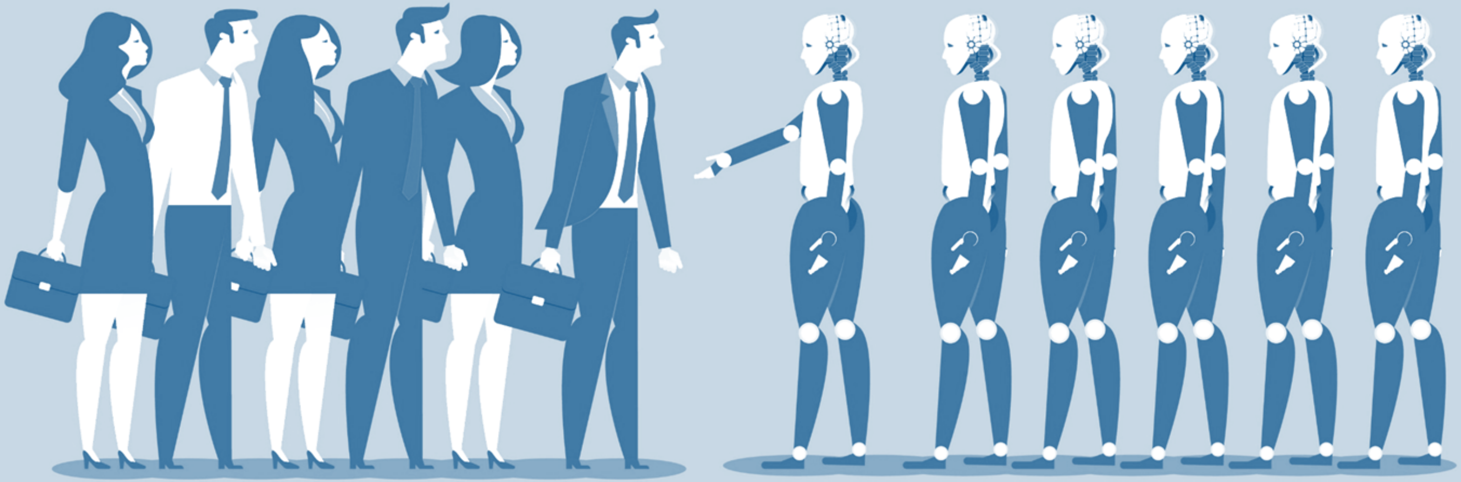
clustering, guided

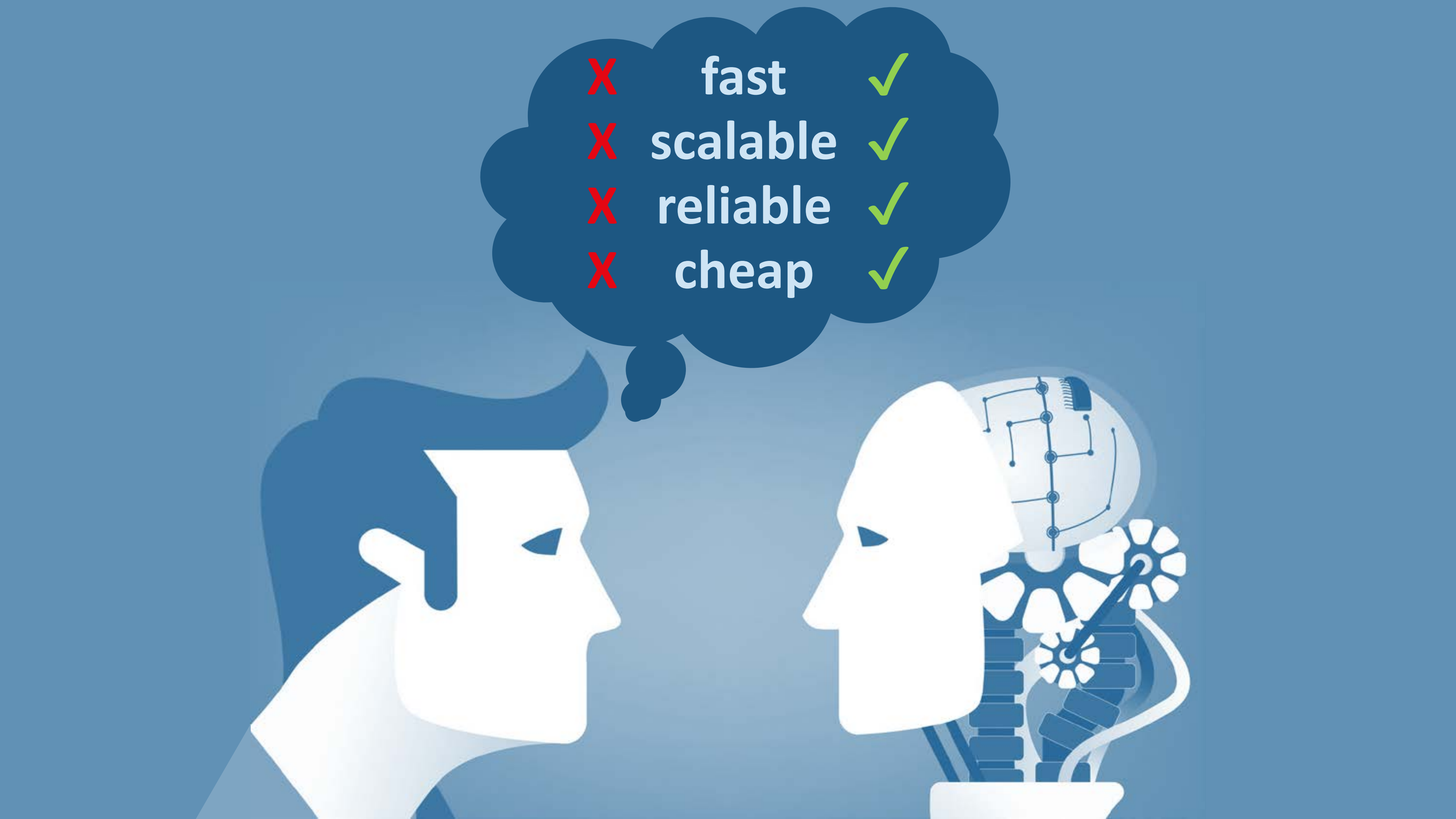
active learning

II. MAN VERSUS MACHINE

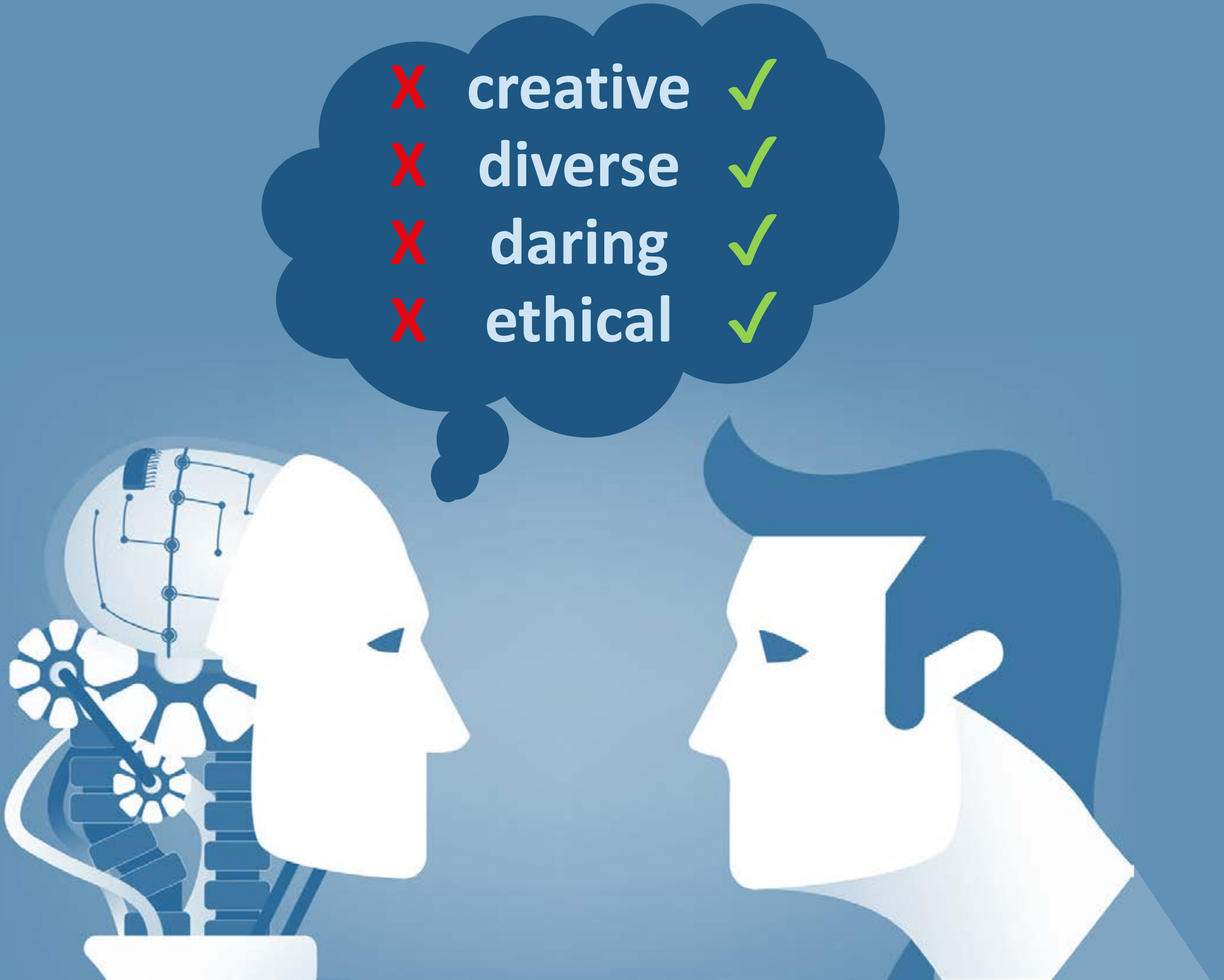
or: better, stronger, faster

(we have the technology)





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



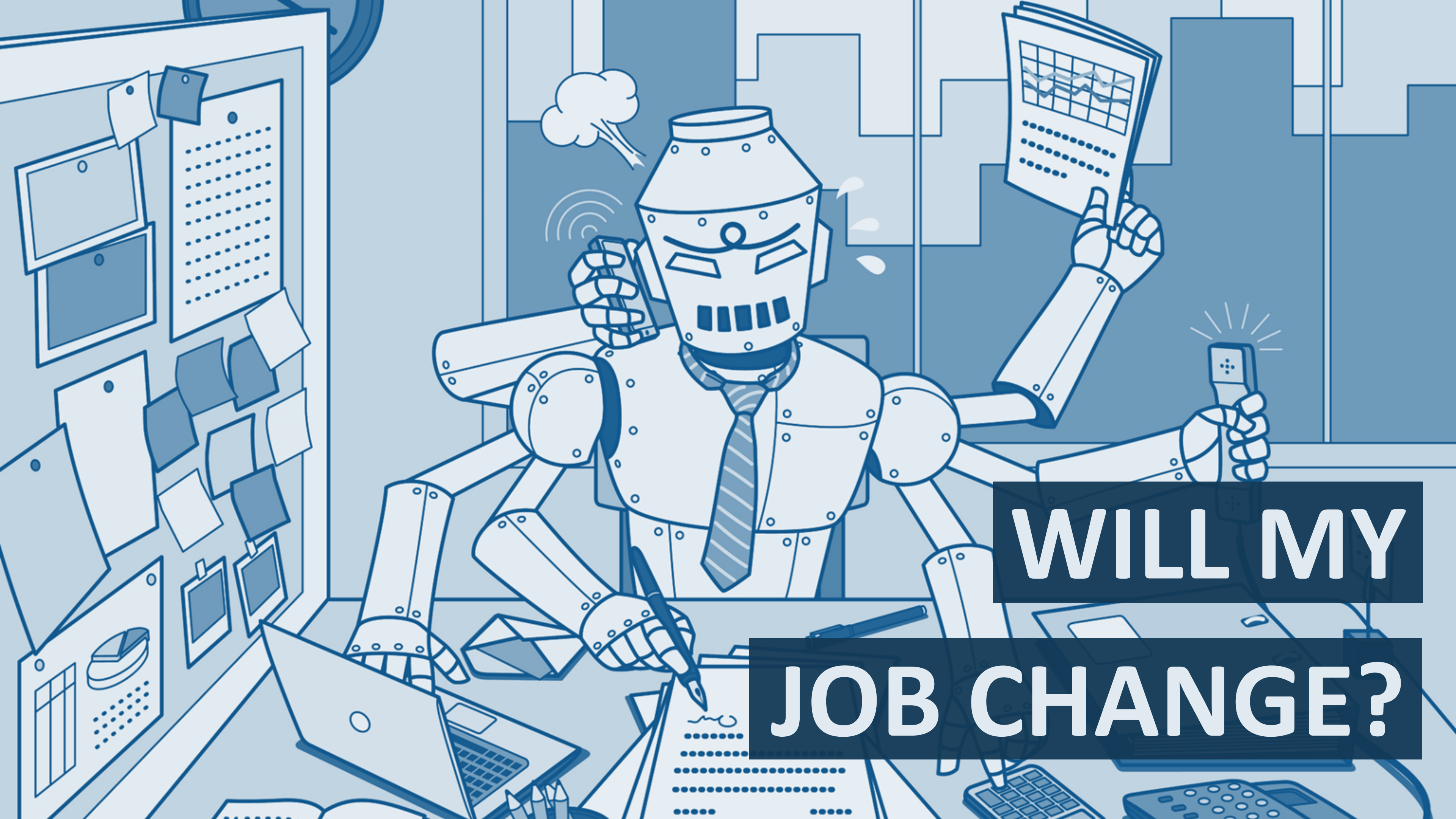
X creative ✓

X diverse ✓

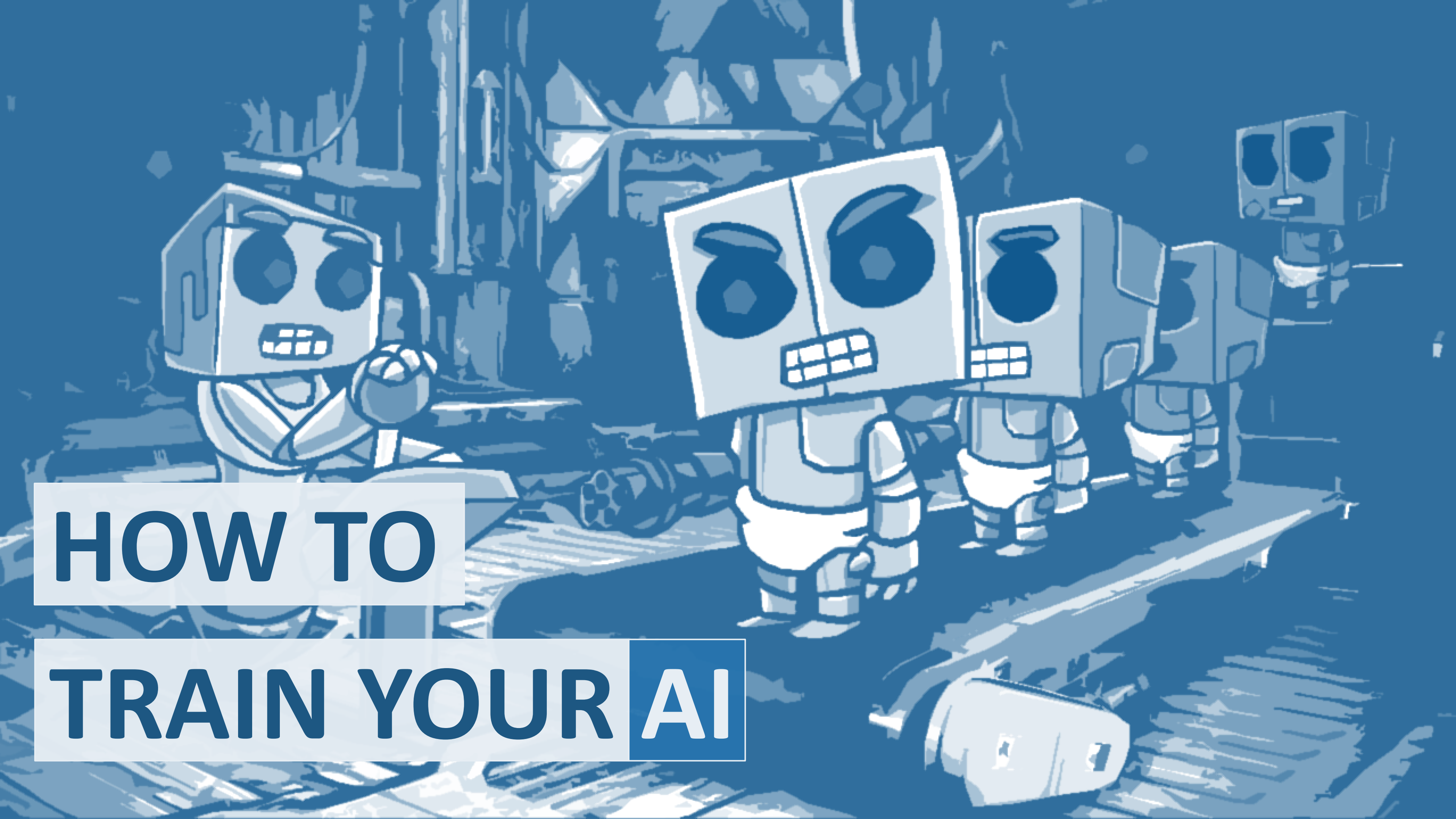
X daring ✓

X ethical ✓

WILL A ROBOT TAKE MY JOB?



WILL MY JOB CHANGE?



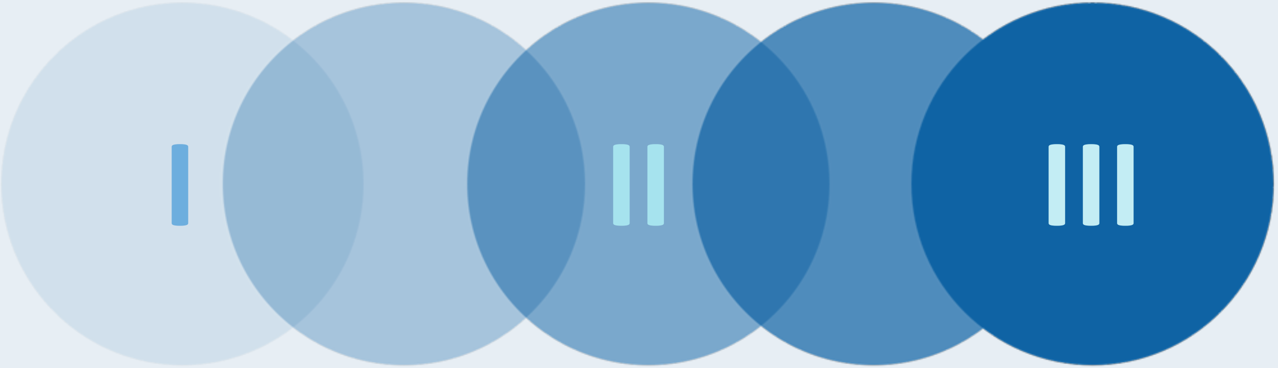
HOW TO TRAIN YOUR AI

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

—MAJOR PAYNE

GENERATIONS:

AI SYMBIOSIS



I

II

III

labels

feature engineering

collaboration

III. SECURING SECURITY SYSTEMS

or: who watches the watchers?

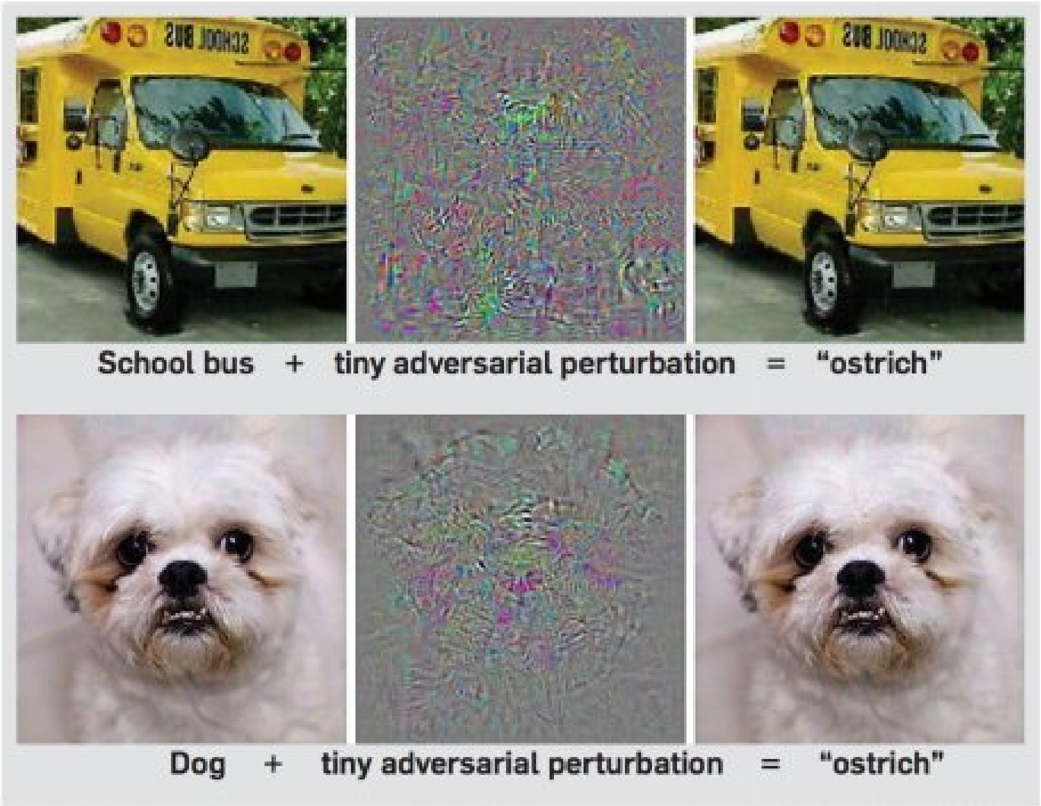
FLAVORS OF ATTACK

poisoning

stealing

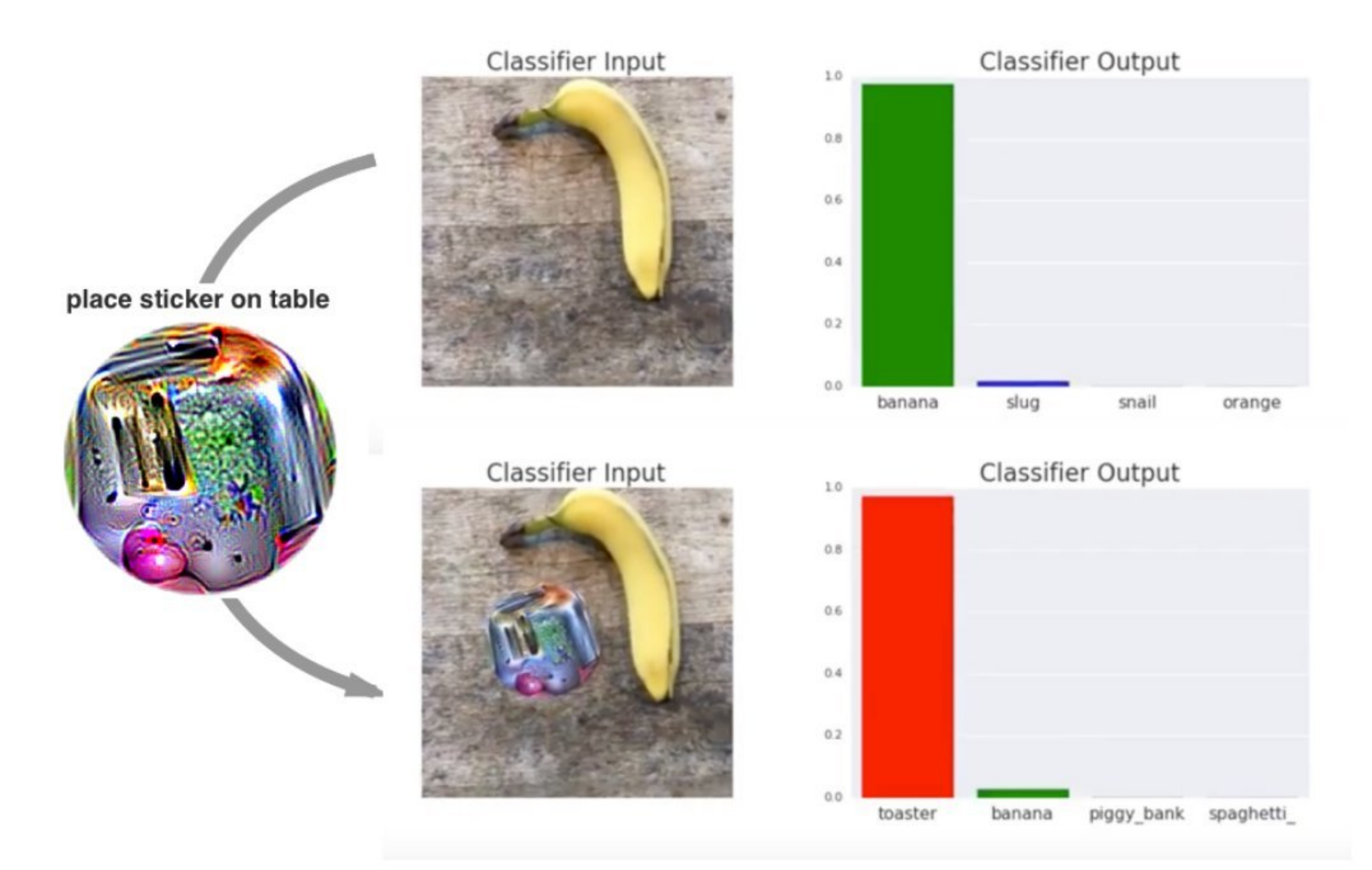
perturbation

MASK: OSTRICH

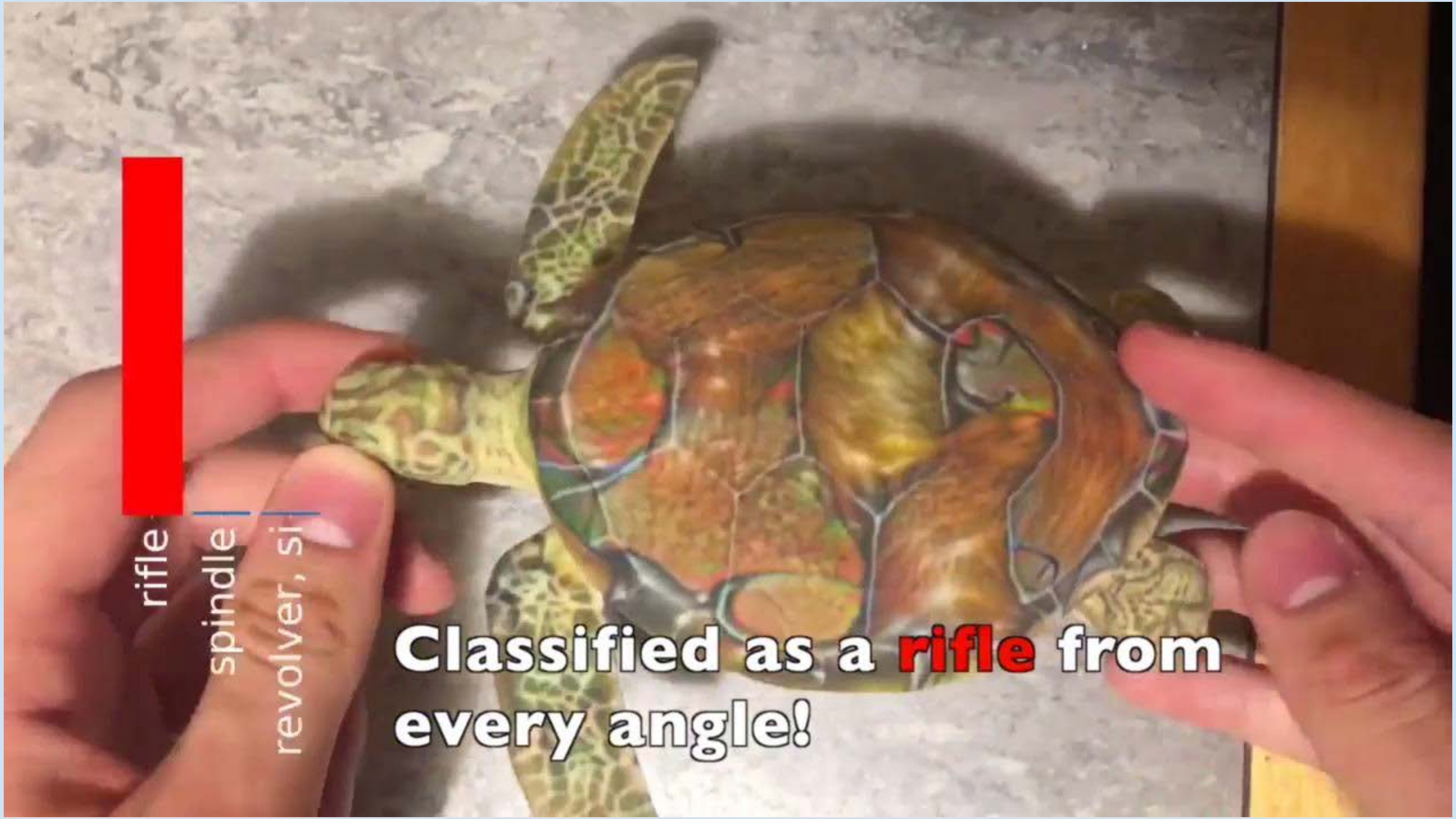

School bus + tiny adversarial perturbation = “ostrich”
Dog + tiny adversarial perturbation = “ostrich”

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

PATCH: TOASTER

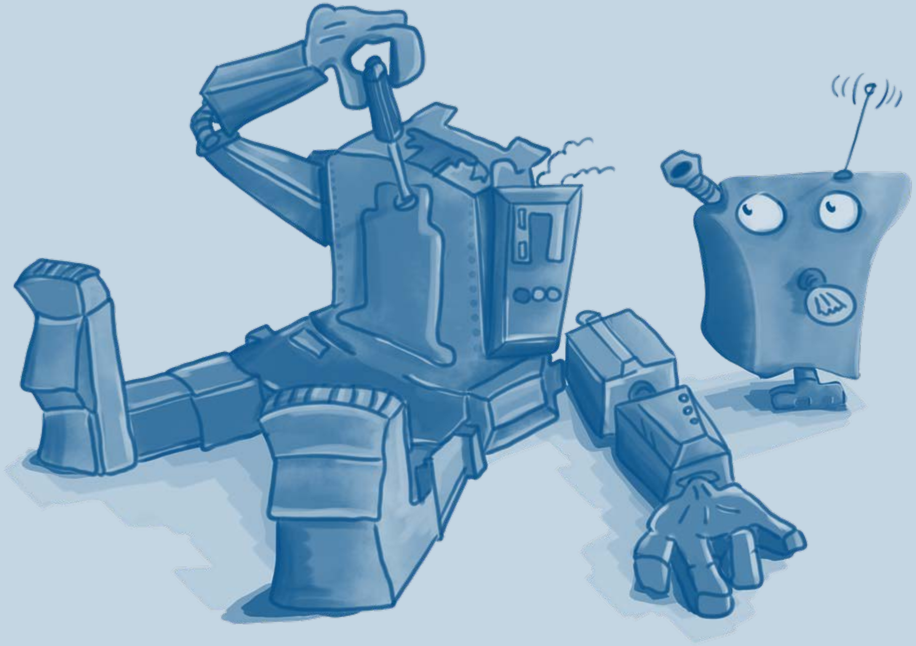


TOY: TURTLE



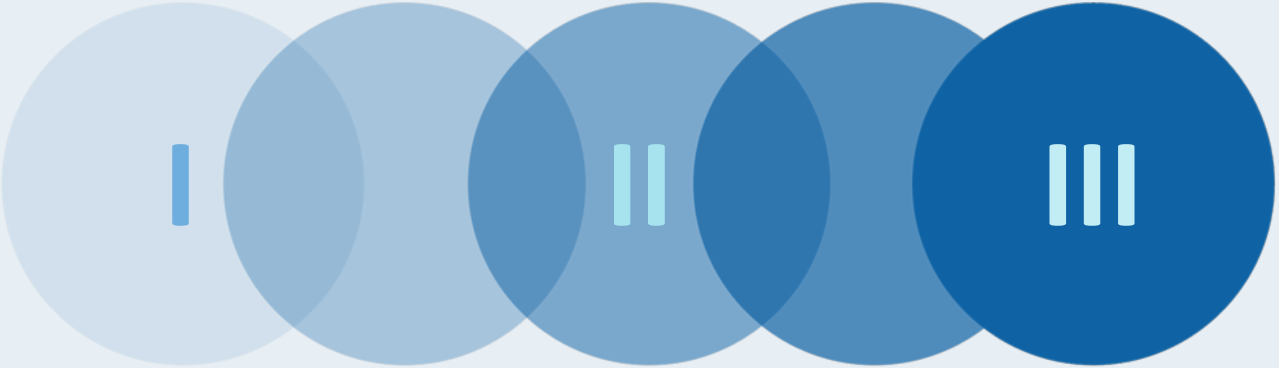
rifle
spindle
revolver, si
Classified as a rifle from every angle!

IS SECURITY AI BROKEN?



GENERATIONS:

SECURING AI



unsecured firmed hardened

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