Remember our history, Least We are Doomed to repeat it ! Malfeasant/AI

Gregory Akers
Executive Technology Consultant
Former Cisco SVP, Advanced Security Research/Governments, CTO Security and Trust
gakers@gregakersconsulting.com
February 10, 2023

Stage Setting

- With the advent of vast amounts of compute and storage power in the mid 2000s, deep (large) neural networks (DNNs) have demonstrated state-of-the-art performance in some application domains
- Now (but not previous to 2000) AI is synonymous with deep learning and “an AI” = a DNN.
- But despite the intense media hype, AI is rarely solely relied upon, and sometimes not used by rule, in settings with high financial stakes or where safety and security implications are significant
- e.g., failures in medical applications (Google, IBM), not used for critical-infrastructure management (AWS)
- Why?

Courtesy Dr. George Kes, Penn State University

Risks of AI

- After 25+ years of intense R&D in cyber security, one typically does not write software and deploy it without thoroughly understanding its behavior for **all possible inputs** not just those deemed operationally “typical” by the developers, and testing the software for bugs, logical ones in particular.
- Though this may be difficult to do for large, complex code-bases, it’s next to impossible to do for commercial AIs which are highly parameterized models, trained on vast datasets of high-dimensional data samples, everything in ad hoc fashion.
- We have spent the last 20+ years trying to secure something not designed securely.

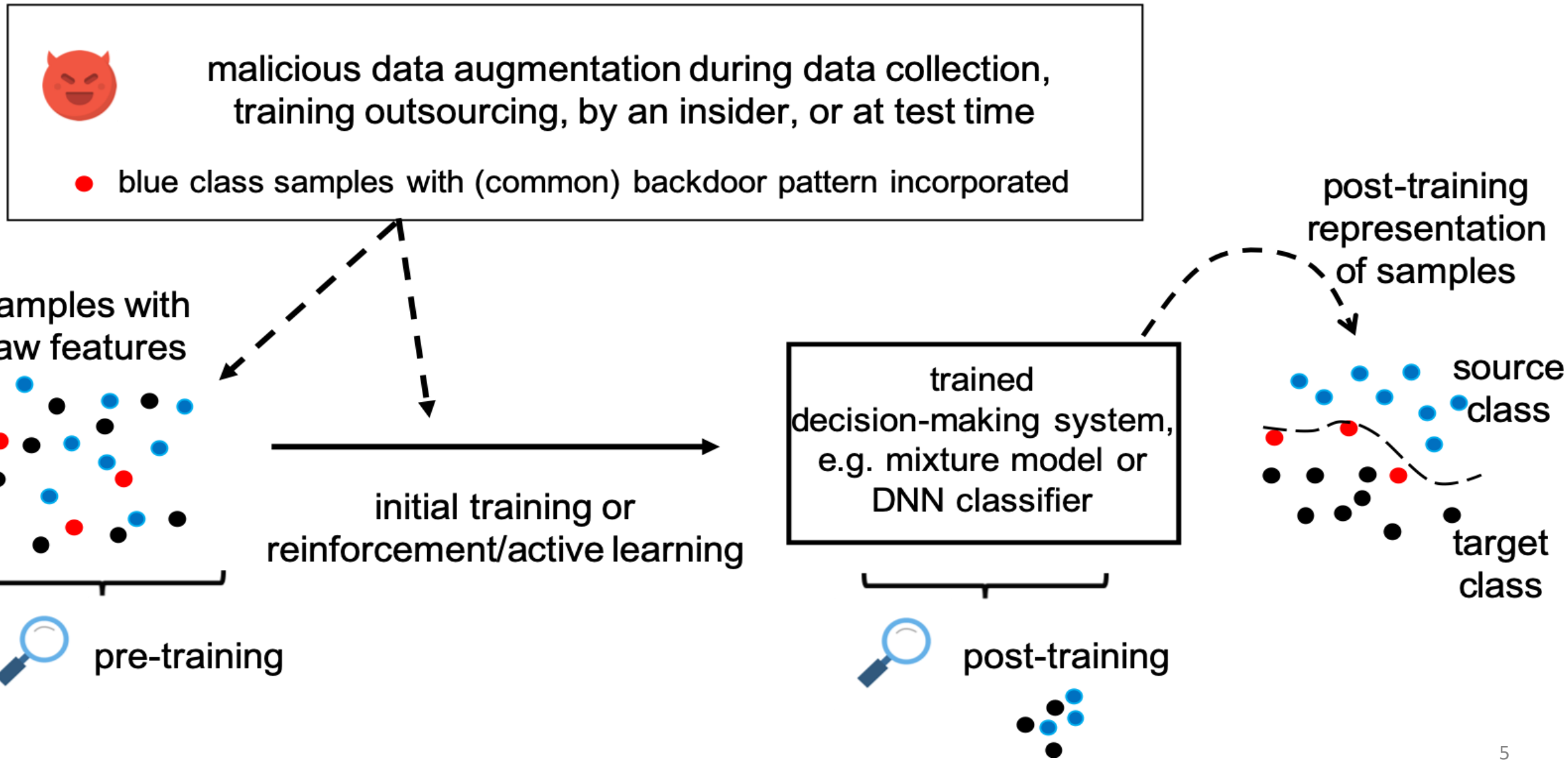
Courtesy Dr. George Kes, Penn State University

Risks of AI – Adversarial AI

- In the past 10 years, researchers have rediscovered techniques to probe complex decision boundaries of trained AI classifiers to demonstrate how they may be reverse engineered and induced to commit errors.
- In addition, simple training-data poisoning techniques have been developed to reduce AI performance or to plant a subtle Trojan/backdoor, the latter to simplify induction to error at test time.
- One idea is to make the AI generally more robust by addressing such threats, Including deploying AIs together with (post-deployment) defenses…

Courtesy Dr. George Kes, Penn State University

Data-poisoning attack & defense scenarios



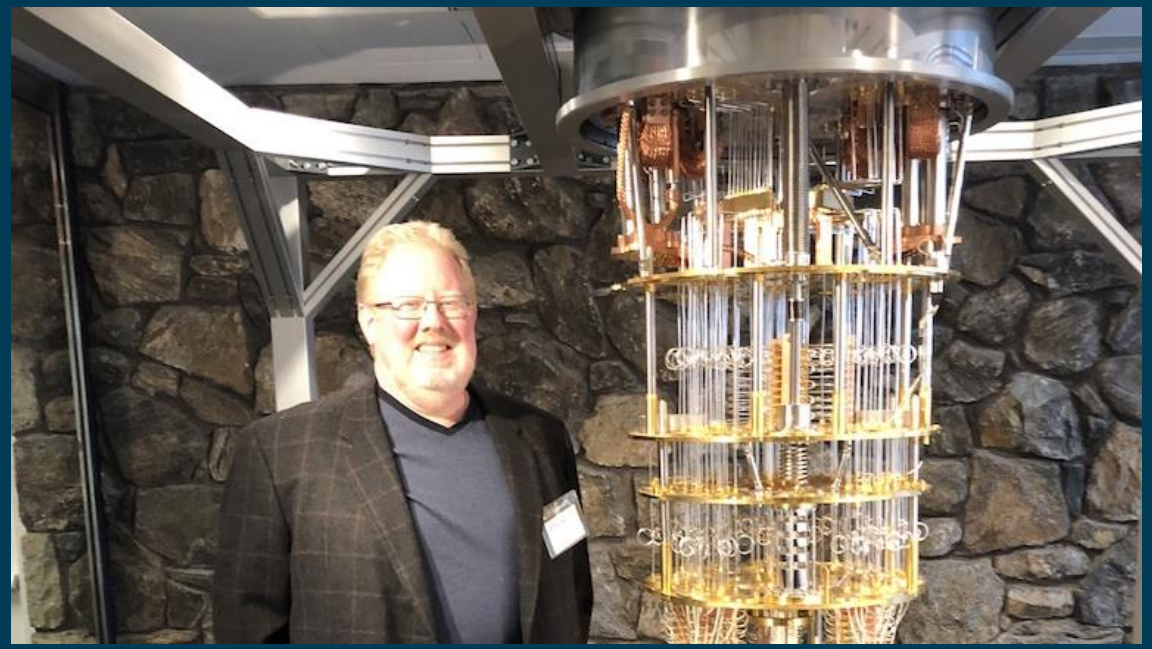
Google

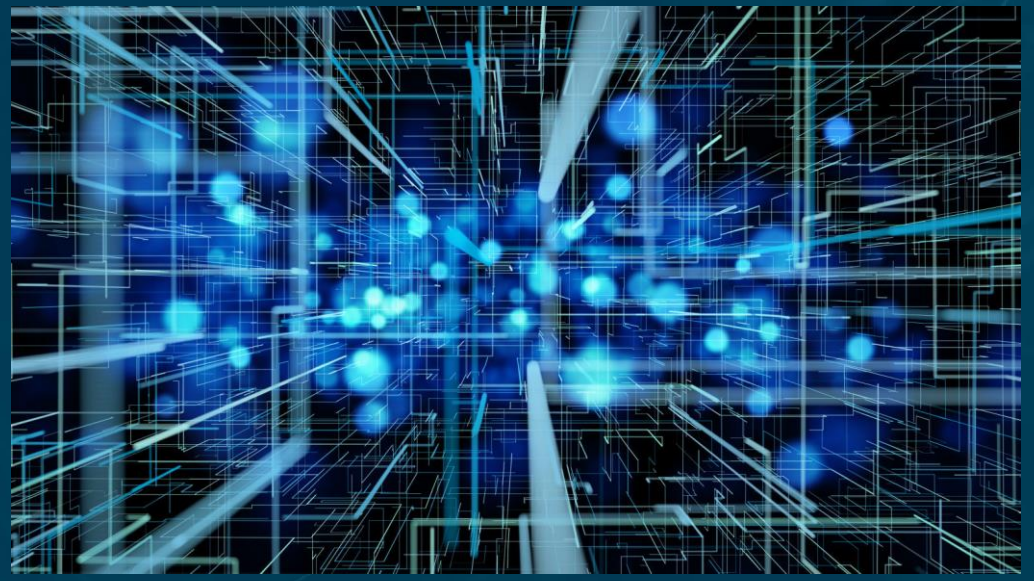
"AI is one of the most important things humanity is working on. It is more profound than, I dunno, electricity or fire."

Sundar Pichai
CEO, Google

Photo: Stephen Lam/Reuters

As I Peer Into My Crystal Ball…





Discussion Areas:

- Impacts on offensive/defensive cyber work ?
- Are we watching and relearning years of previous cyber lessons ?
- What ecosystems will drive trusted AI in the hybrid cloud world ?
- What are the core system implications for trusted AI?
- What are the opportunities for malfeasance ?
- Testing and inspection is faulty and expensive, then what.. More AI ?
- IP/Cyber protection is still an after thought. What now ?
- Market transitions are fast and merciless – Offensive advantage ?

Prepare for it now, don’t try to bolt it on later !

Q&A


Gregory Akers
Consultative Technology Executive
Greg Akers Consulting
gakers@gregakersconsulting.com
919 345 4525
225 N. Dogwood Trail
Southern Shores, NC 27949
www.gregakersconsulting.com