IEEE SA
STANDARDS ASSOCIATION

STANDARDS FOR TRUSTWORTHY AUTONOMOUS VEHICLES

February 12th, 2021

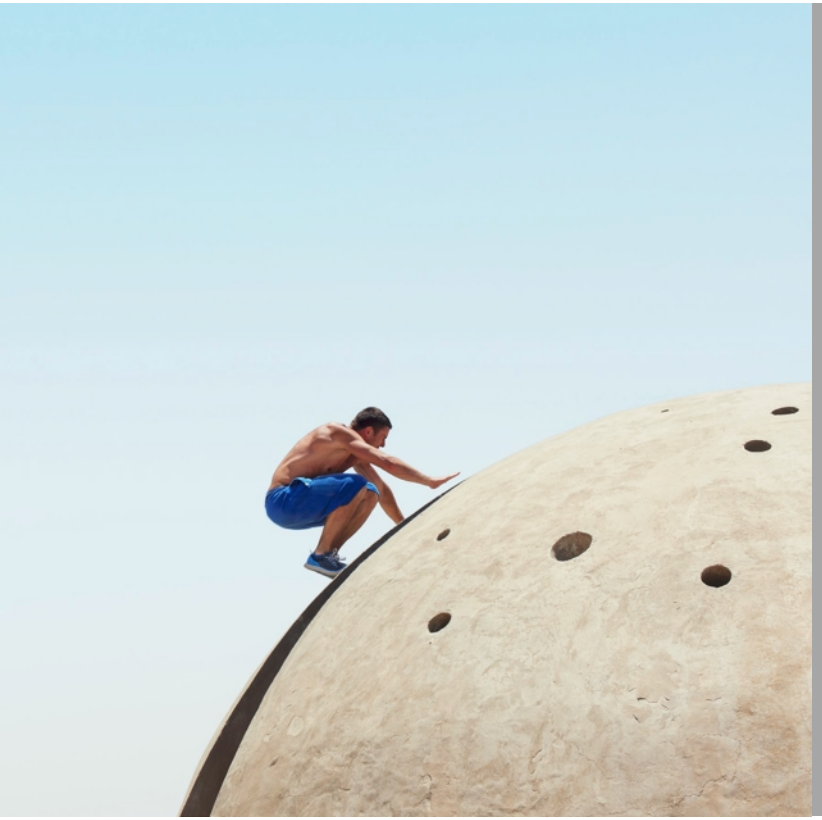
ventum consulting

Speaker:

Philipp Schneidenbach

Head of Compliance & Information Security

Above and Beyond Standardization - operational aspects of technology evergreening and digital products

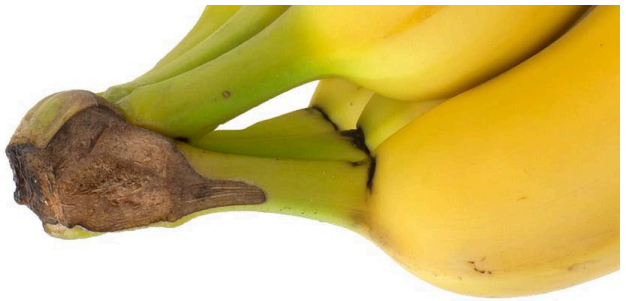


1 Adversarial Examples

Stateful Technology in a Stateless Environment vs. Level 4 Autonomous Vehicles


6
7
DEFROST
REHEAT
STOP


6
7
DEFROST
REHEAT
STOP

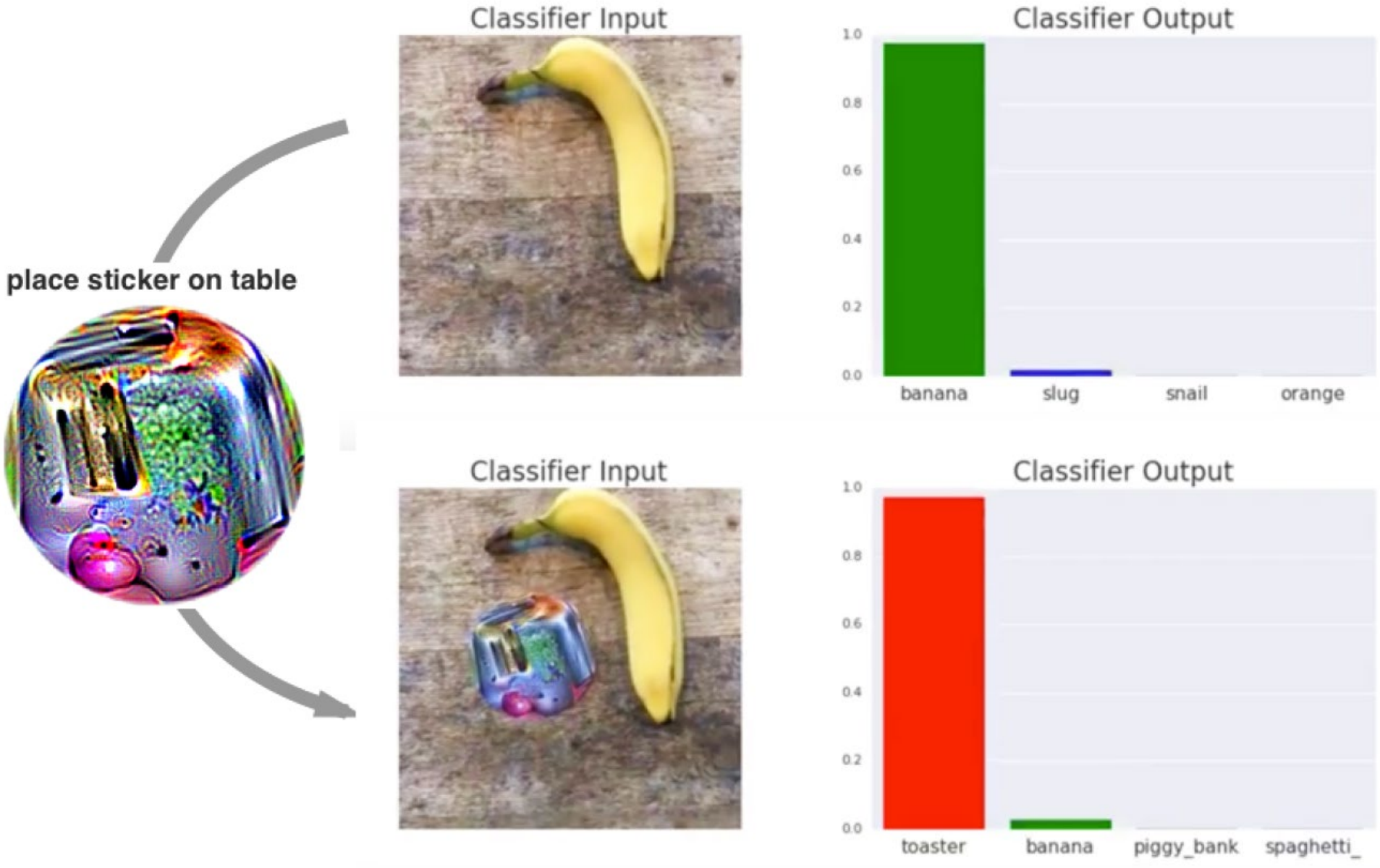



3
4
5
6
7
DEFROST
REHEAT
STOP

Adversarial Patch

Tom B. Brown, Dandelion Mané*, Aurko Roy, Martín Abadi, Justin Gilmer
{tombrown,dandelion,aurkor,abadi,gilmer}@google.com

arXiv:1712.09665v2 [cs.CV] 17 May 2018



Fooling automated surveillance cameras: adversarial patches to attack person detection

Simen Thys*
simen.thys@student.kuleuven.be

Wiebe Van Ranst*
wiebe.vanranst@kuleuven.be

arXiv:1904.08653v1 [cs.CV] 18 Apr 2019

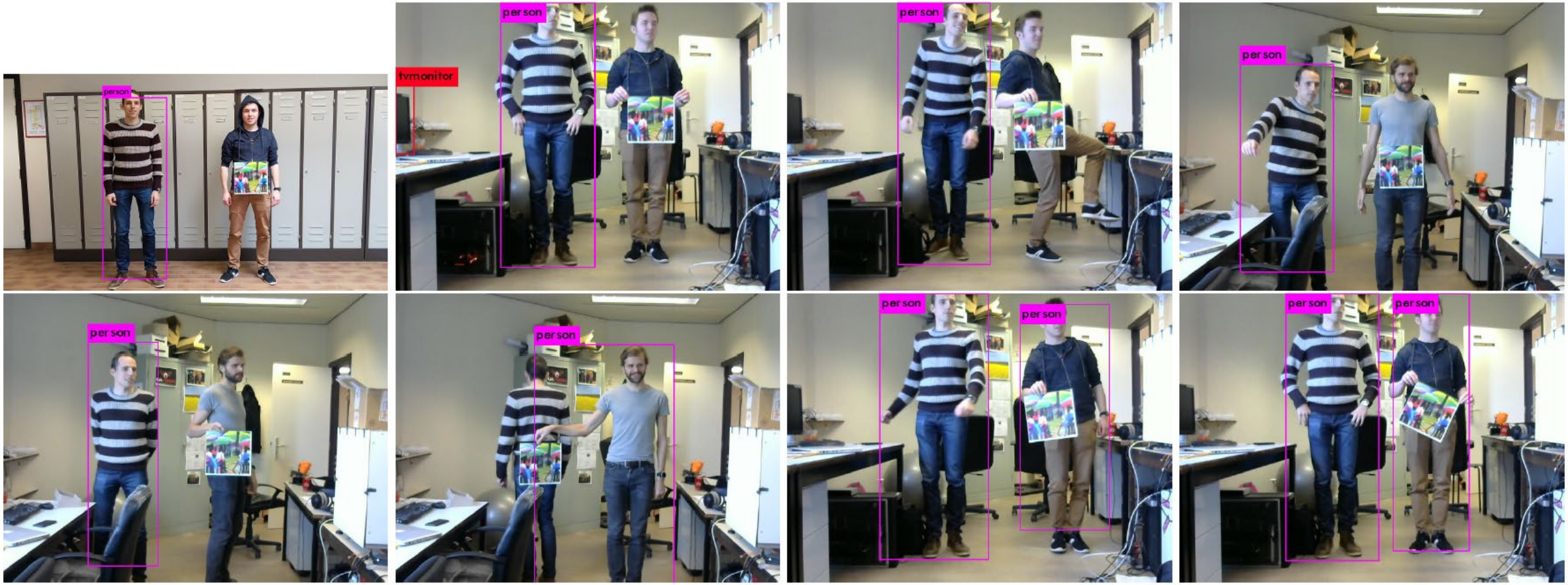


Figure 7: Real-world footage using a printed version of our patch.

Universal Adversarial Perturbations Against Semantic Image Segmentation

arXiv:1704.05712v3 [stat.ML] 31 Jul 2017

Jan Hendrik Metzen
Bosch Center for Artificial Intelligence, Robert Bosch GmbH
janhendrik.metzen@de.bosch.com

Mummadi Chaithanya Kumar
University of Freiburg
chaithu0536@gmail.com

Thomas Brox
University of Freiburg
brox@cs.uni-freiburg.de

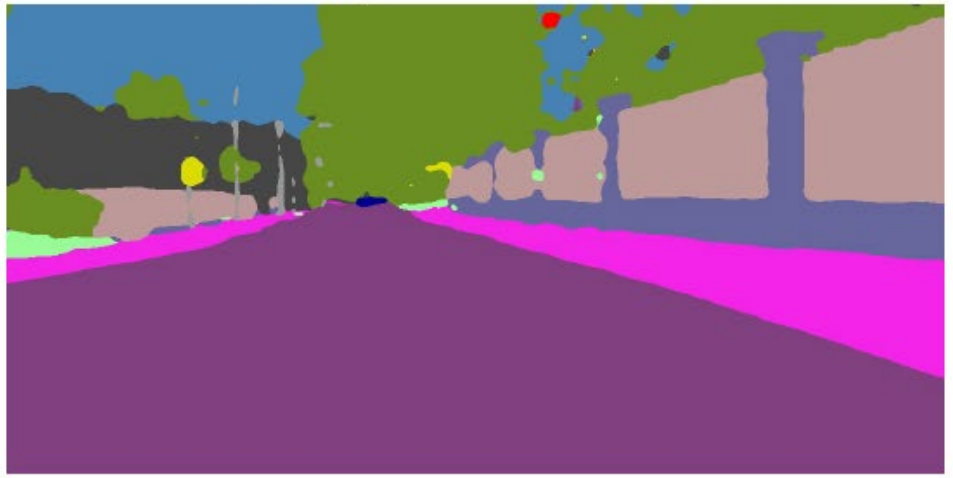
Volker Fischer
Bosch Center for Artificial Intelligence, Robert Bosch GmbH
volker.fischer@de.bosch.com

(a) Image

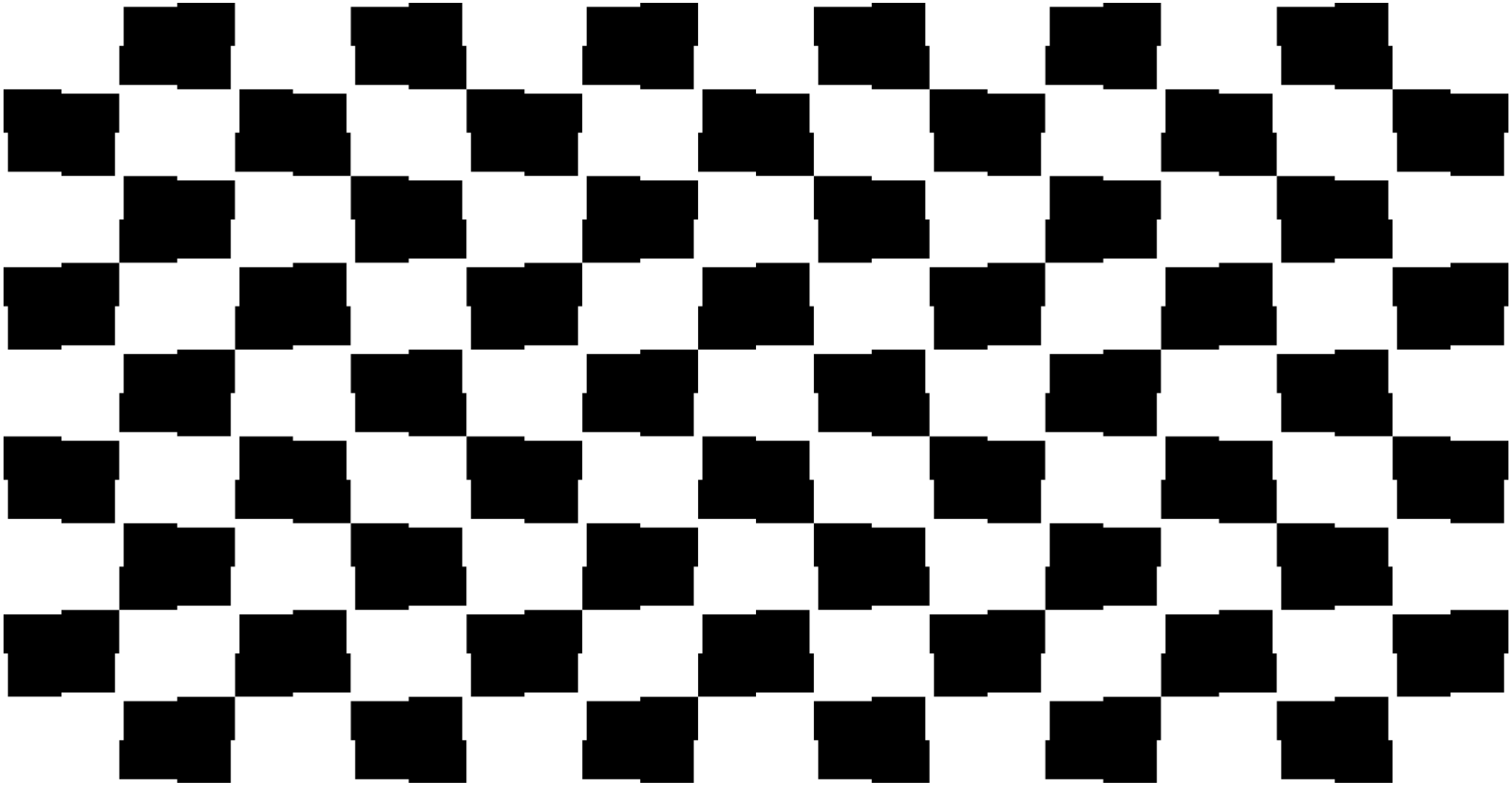
(b) Prediction

(c) Adversarial Example

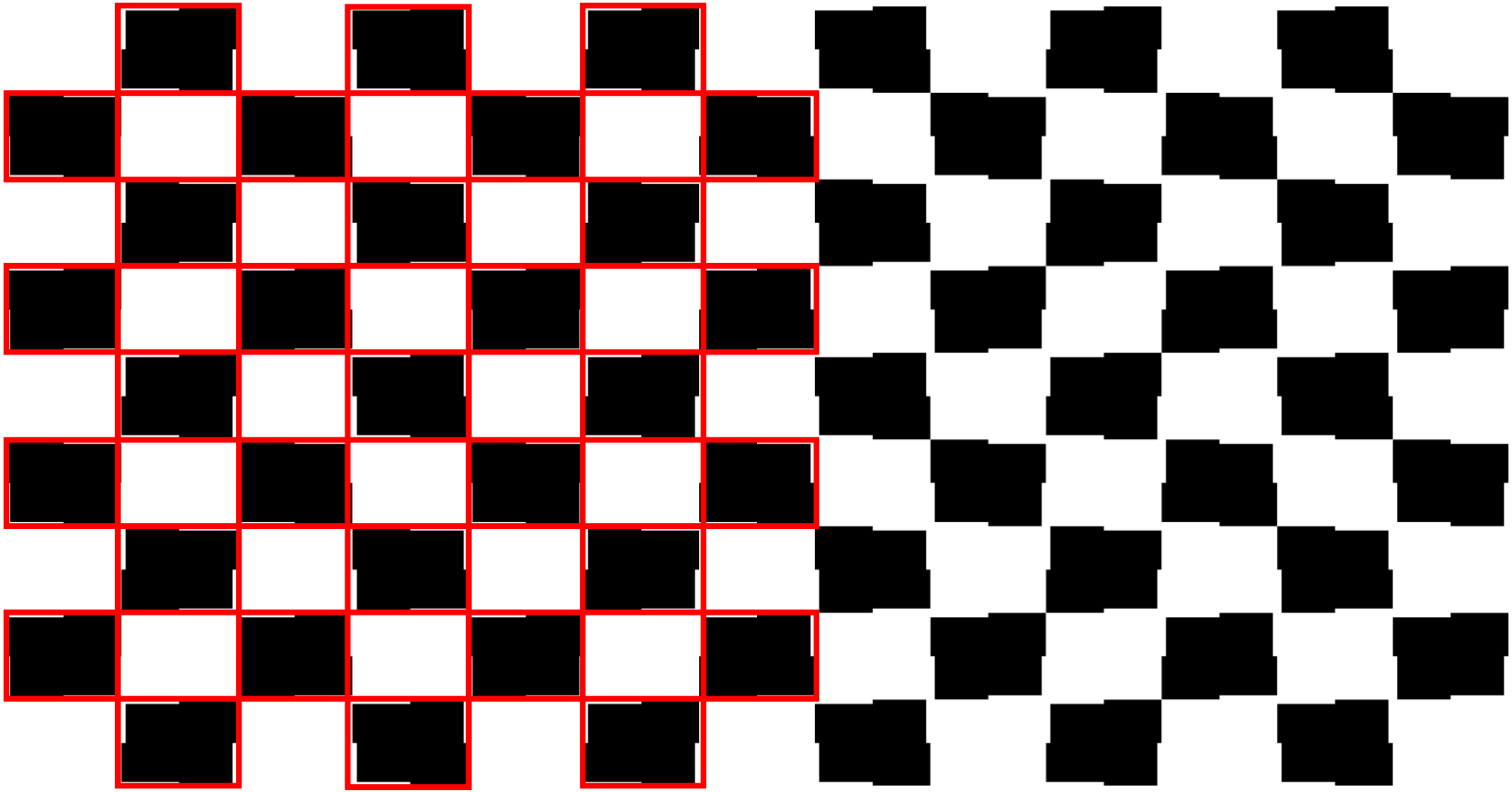
(d) Prediction



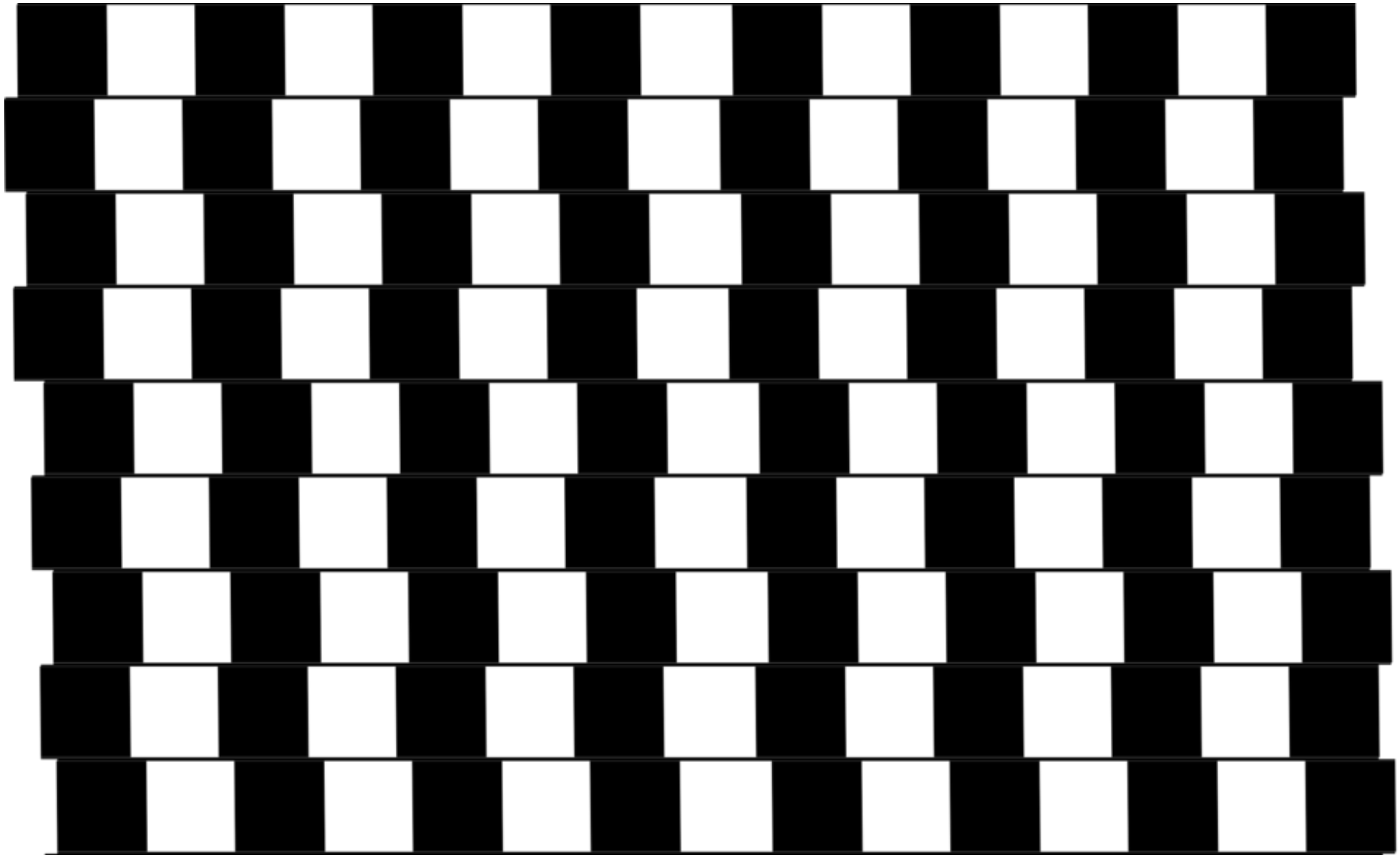
Machine Vision and Human Senses can be so alike.



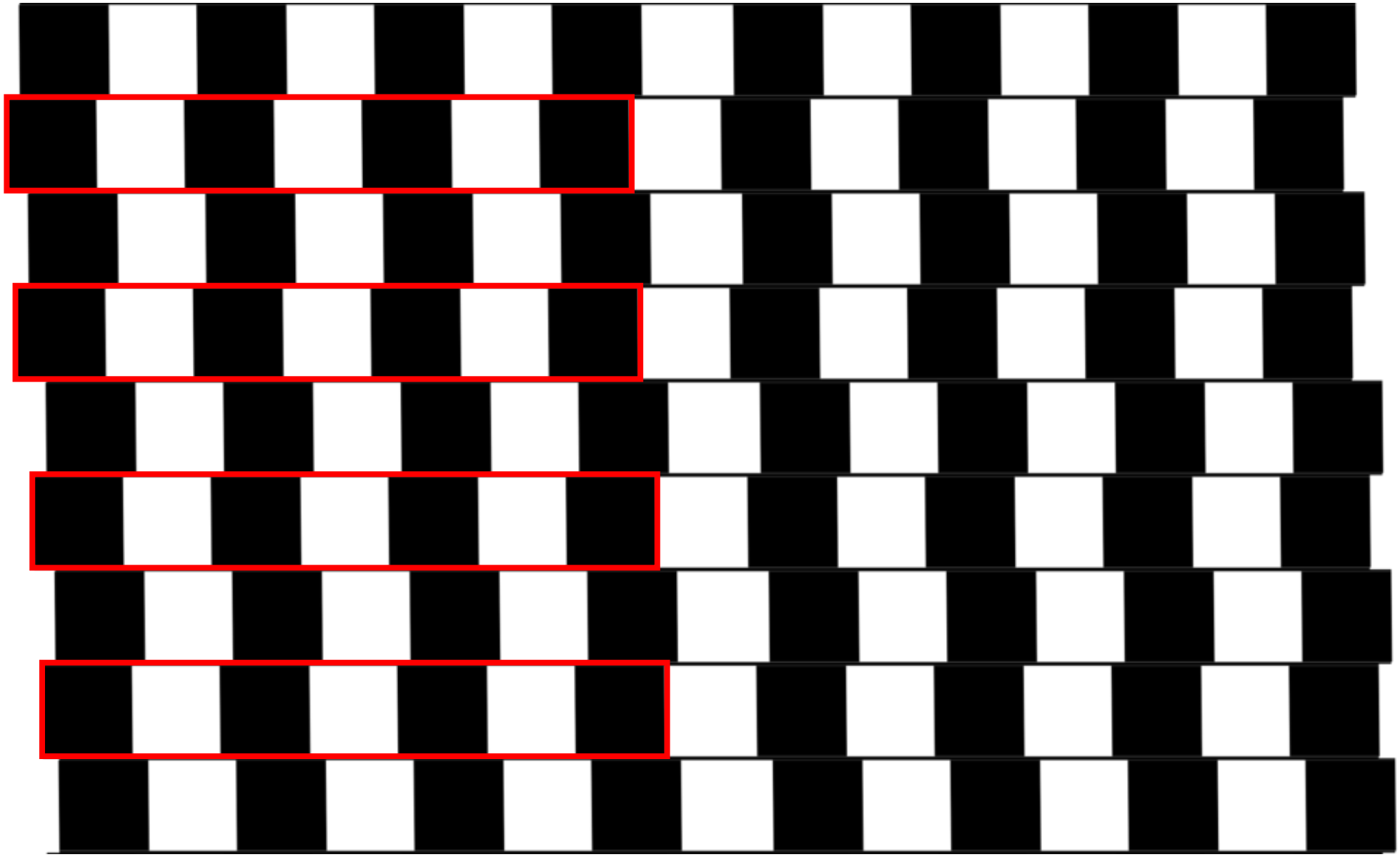
Verification Patterns help understand and verify data.



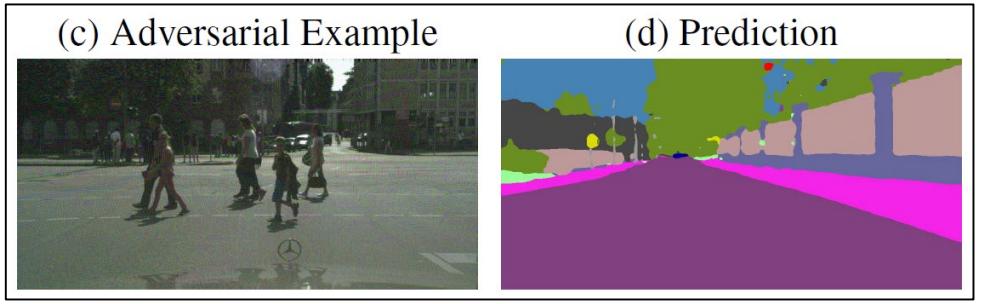
Perspectives can be rather hard to deliminate.



Patterns can even help when probing sections of data sets.



The road ahead: Data Integrity and ML Patterns are key to Security



Punchline:

- ML and Image Recognition **can not yet deliver** the same amount of **security that is expected** from drivers, insurers, risk managers and legal authorities.

Feasible Countermeasures:

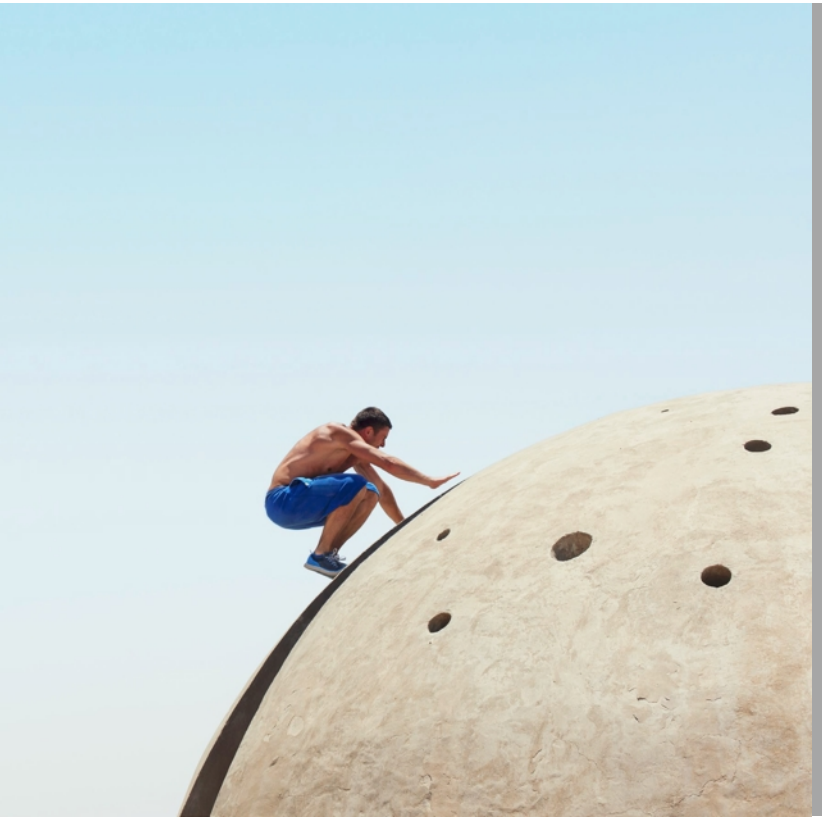
- Pre-departure LIDAR /Time of Flight **calibration**
- **Reference** Vehicle Pairing (Micro Platooning)
- Lane **View** /Angle **Comparison** (human eye-example)
- Traffic Lights & Imaging **Plausibility** (5G will help)
- The **MITRE Threat Matrix** can help to prevent:
  - Data Stream **Proxying**
  - Data/Model **Poisoning** using tainted data

Blackhat EU Talk on Lidar Spoofing:

- https://www.blackhat.com/docs/eu-15/materials/eu-15-Petit-Self-Driving-And-Connected-Cars-Fooling-Sensors-And-Tracking-Drivers.pdf

IEEE Talk: Is Elon Musk Right? Are Lidars A Crutch?

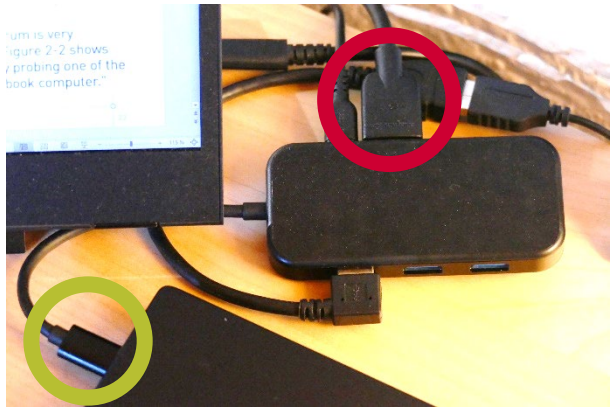
- https://www.youtube.com/watch?v=FjlmIVKFZUU
- https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/e2e-presentations/2019/Webinar-Is_Elon_Musk_Right_Are_Lidars_a_Crutch.pdf

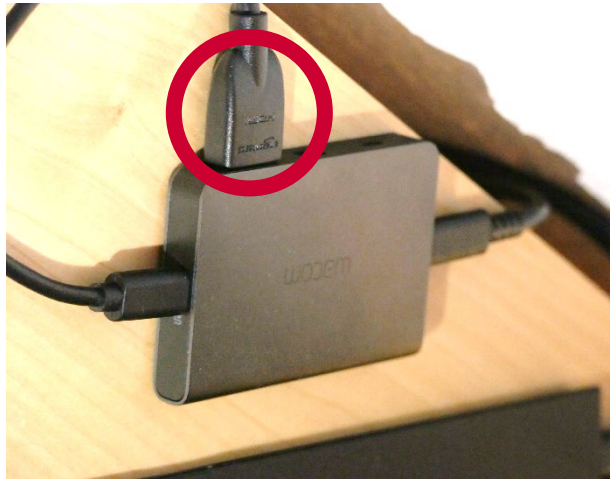


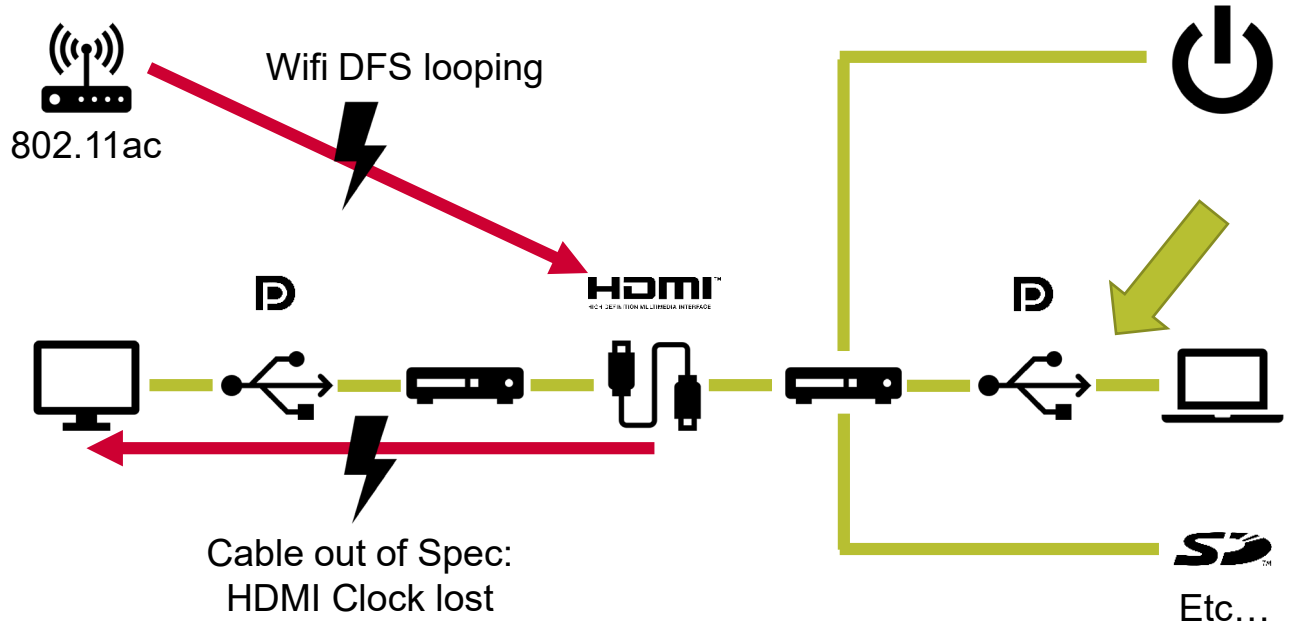
B 5G Frequency Clogging

Constrictions in proprietary interfacing vs. malicious peripherals

HDMI, DP Alt Mode and USB 3.x – signaling beyond spec







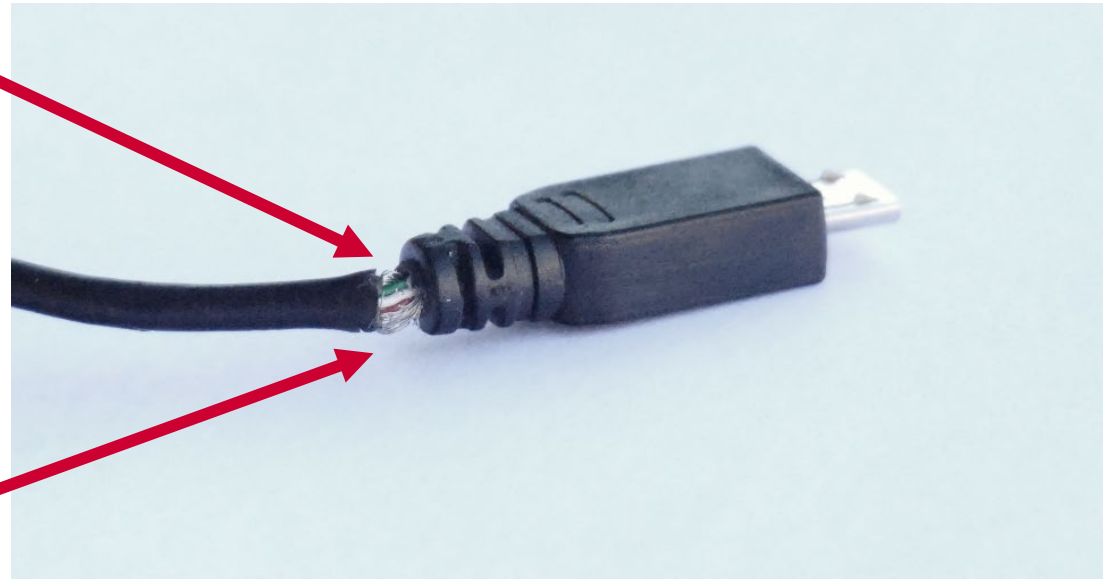
The increase in maximum bandwidth is achieved by increasing both the bitrate of the data channels and the number of channels. **Previous HDMI versions use three data channels (each operating at up to 6.0 GHz in HDMI 2.0, or up to 3.4 GHz in HDMI 1.4)**, with an additional channel for the TMDS clock signal, which runs at a fraction of the data channel speed […] **for signaling rates between 3.4 and 6.0 GHz**.

https://en.wikipedia.org/wiki/HDMI#Version_2.1

Consumer Perception: My „Charging Cable“

Truth is: We (do not) know what people are bringing to their cars.

Broken Shielding
= RF Interference

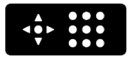


Charging Cables do
not need Data Lines.

Multi-Functional

Networking and
Wireless

Mass Storage



Input Devices

…basically, this means all Device Classes.

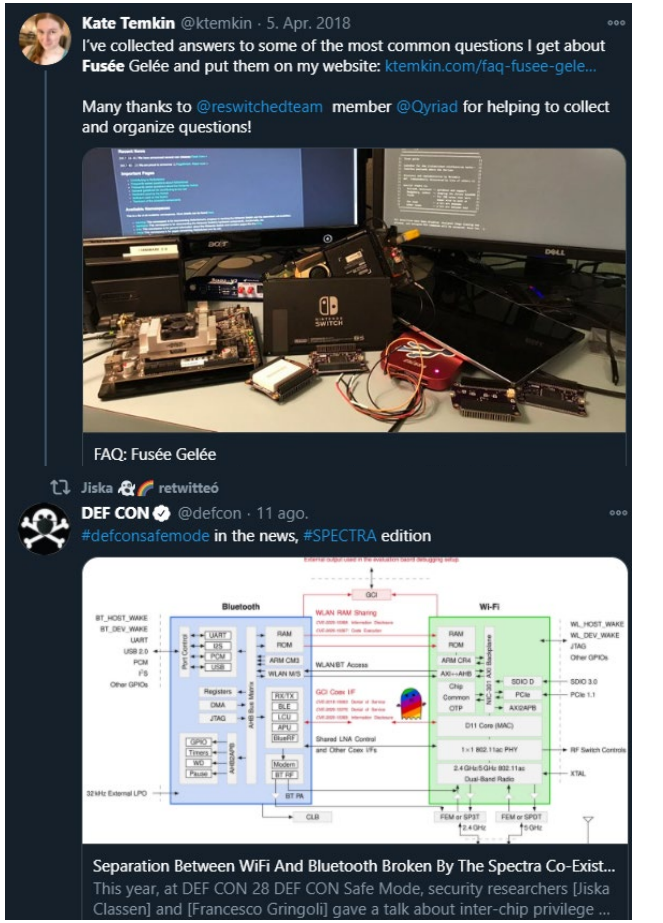
https://en.wikipedia.org/wiki/USB#Device_classes

USB and Wireless – Standards not intended to be used like:

Plug-and-Play for payloads

USB cables w/ integrated payload

Exploitation via USB and Wireless



FREE-FALL: Tesla Remote Attack: Entering CAN Bus

"We have proved that we can **gain entrance from wireless (Wi-Fi/Cellular)**, compro-mise many in-vehicle systems like IC, CID, and Gateway, and then **inject malicious CAN messages into the CAN Bus**. Just 10 days after we submitted our research to Tesla, Tesla responded with an update using their OTA mechanism and **introduced the code signing protection into Tesla cars**."

https://www.youtube.com/watch?v=c1XyhReNcHY

https://www.blackhat.com/docs/us-17/thursday/us-17-Nie-Free-Fall-Hacking-Tesla-From-Wireless-To-CAN-Bus-wp.pdf

Smartphones on wheels: A different approach is needed

Convenience-based Plug-and-Play from the 90s

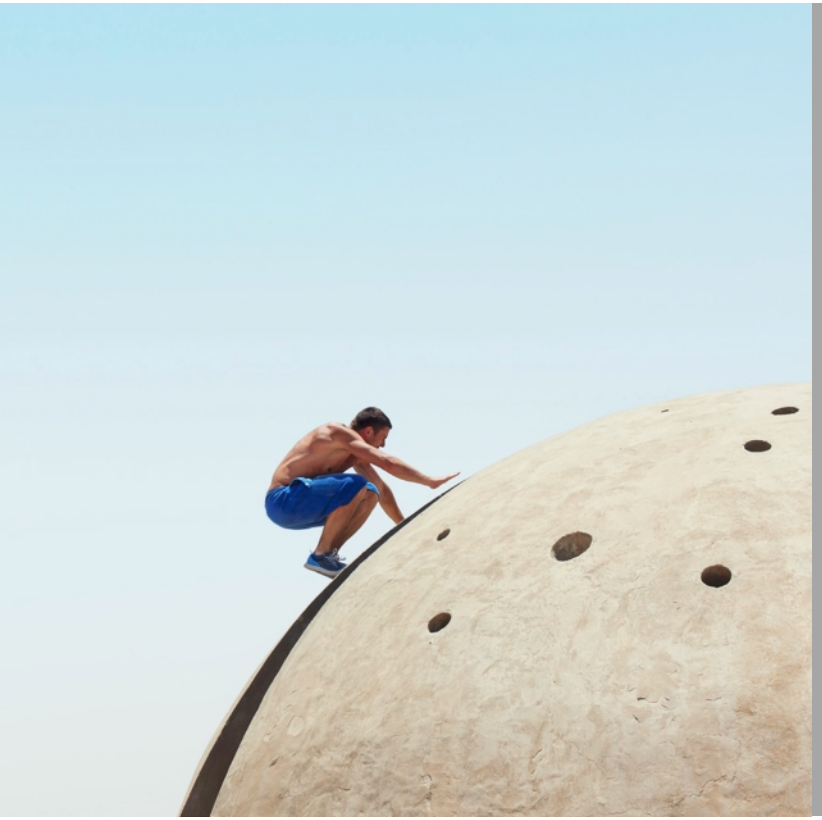
- **Interfaces do not need to authenticate**, get mounted or prioritized and lack verification regarding user interaction.
- Physical compatibility and operating system support face a consumer market where **USB ports have been put into every battery-powered device**
- **In the field, ports are used by consumers** without any diversification between a charger and “real” USB devices
- **Wireless has become ubiquitous** and is even controlling lawn mowers, before the 5G shift is even happening

→ No way to get a hold of, update or secure retrospectively

Securing the Future: Technology Evergreening

- **Authenticate all peripherals** – yes, USB Cables need to use Credentials, Signatures and Certificates to work
- Health Radar approach incl. fallback mechanisms: **In-Car-Conditions** need to go beyond “Low-Oil” Warnings and **must incorporate eco-system changes to all interfaces**
- **Signatures & Distributed Ledger**: Today’s tech can help when transferring peripherals or re-pairing them to cars
- **Mission-critical applications** will move to wireless standards such as **5G and thus need QoS**-like security

→ Connected devices will need managed, software-defined security as technology evergreening becomes a standard

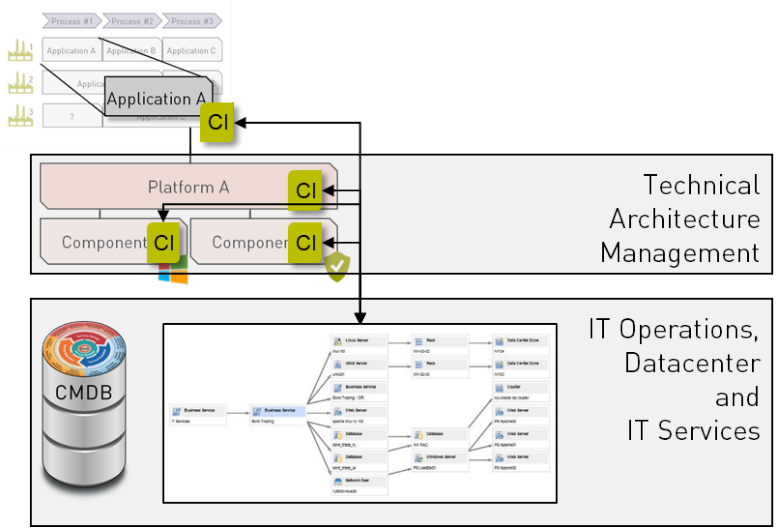


C

Mastering Complexity

IT Architectures and the MITRE Adversarial ML Threat Matrix

Finally, around 2016: Architectural Risk got mainstreamed



- IT Management has learned that the **functional condition of a component** within an ecosystem is **no longer an appropriate metric for security**.
- **Consumerization of IT** is best described with everybody is bringing his private phone to the office or using their office IT at home.
  This means: **Security needs to be validated based on all aspects of a service, end-to-end,** not only on the level of devices, locations or services.
- The integration of meta data, such as **CVE®** from **vulnerability databases like NIST** has become a **standard for assessing and reducing risk.**
- **Vulnerability-Data** from the CVE® system by **MITRE** (https://cve.mitre.org/) is **normalized** and enables risk assessments across all IT Assets.
- CVE® is a list of entries – each containing an identification number, a description, and at least one public reference—for publicly known cybersecurity vulnerabilities. CVE Entries are used in numerous cybersecurity products and services from around the world, including the U.S. National Vulnerability Database (NVD).

The Future: From remediation to Secure by Design and Default

- **The MITRE Adversarial ML Threat Matrix – were does it come from?**

In the last three years, major companies such as Google, Amazon, Microsoft, and Tesla, had their **ML systems tricked, evaded, or misled**. This trend is only set to rise: According to a Gartner report. **30% of cyberattacks by 2022 will involve data poisoning, model theft or adversarial examples.**

- **Industry is underprepared.** In a survey of 28 organizations spanning small as well as large organizations, 25 organizations did not know how to secure their ML systems. <https://arxiv.org/pdf/2002.05646.pdf>
- This threat matrix came out of partner-ship with **12 industry and academic re-search groups** with the goal of empower-ing security analysts to orient themselves to these new and upcoming threats.
- The framework is seeded with a **curated set of vulnerabilities and adversary be-haviors** that Microsoft and MITRE have vetted to be **effective against production ML systems**.

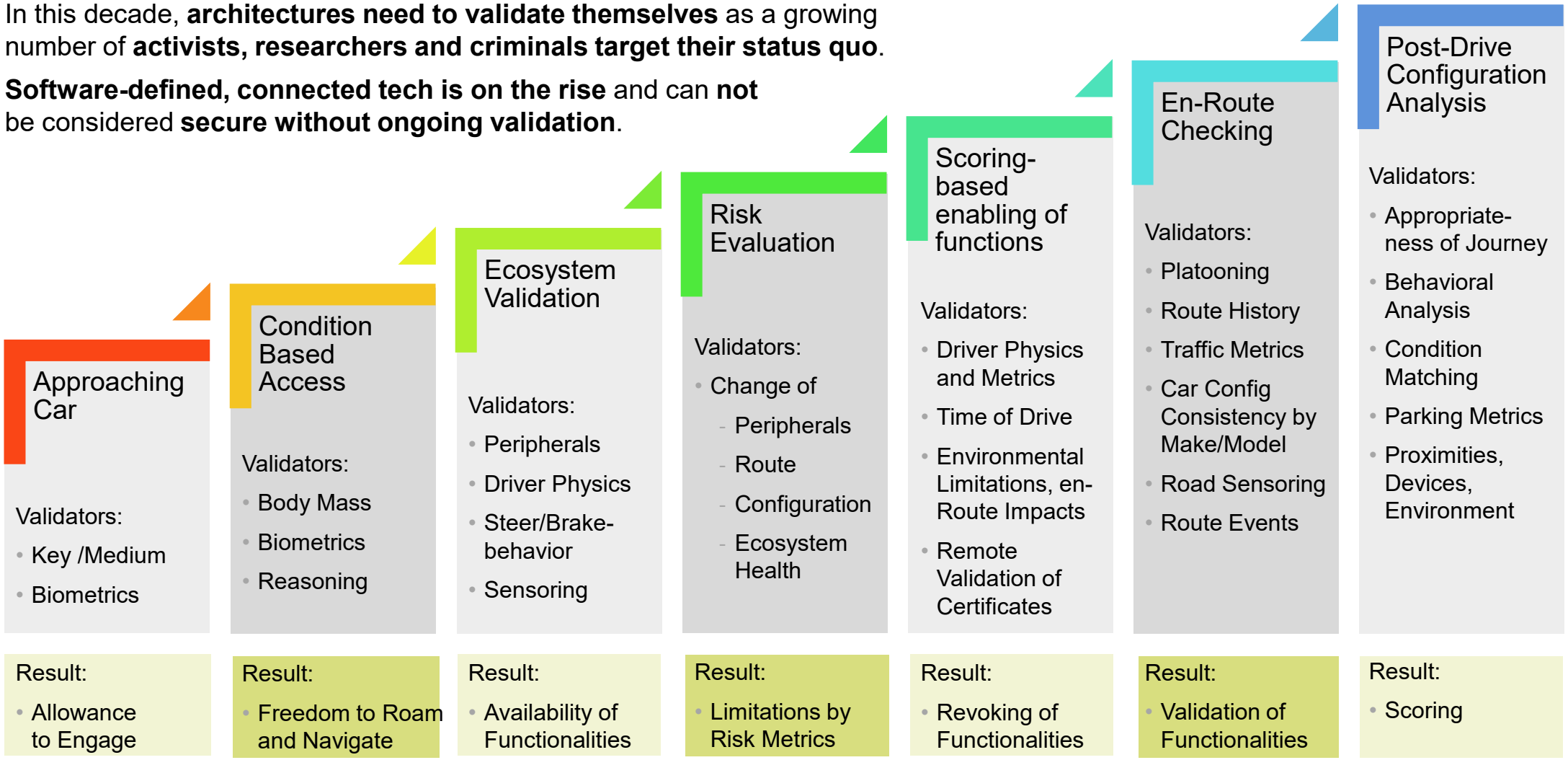
Reconnaissance	Initial Access	Execution	Persistence	Model Evasion	Exfiltration	Impact
Acquire OSINT information: (Sub Techniques) 1. Arxiv 2. Public blogs 3. Press Releases 4. Conference Proceedings 5. Github Repository 6. Tweets	Pre-trained ML model with backdoor	Execute unsafe ML models (Sub Techniques) 1. ML models from compromised sources 2. Pickle embedding	Execute unsafe ML models (Sub Techniques) 1. ML models from compromised sources 2. Pickle embedding	Evasion Attack (Sub Techniques) 1. Offline Evasion 2. Online Evasion	Exfiltrate Training Data (Sub Techniques) 1. Membership inference attack 2. Model inversion	Defacement
ML Model Discovery (Sub Techniques) 1. Reveal ML model ontology – 2. Reveal ML model family –	Valid account	Execution via API	Account Manipulation		Model Stealing	Denial of Service
Gathering datasets	Phishing	Traditional Software attacks	Implant Container Image	Model Poisoning	Insecure Storage 1. Model File 2. Training data	Stolen Intellectual Property
Exploit physcial environment	External remote services			Data Poisoning (Sub Techniques) 1. Tainting data from acquisition – Label corruption 2. Tainting data from open source supply chains 3. Tainting data from acquisition – Chaff data 4. Tainting data in training environment – Label corruption		Data Encrypted for Impact Defacement
Model Replication (Sub Techniques) 1. Exploit API – Shadow Model 2. Alter publicly available, pre-trained weights	Exploit public facing application					Stop System Shutdown/Reboot
Model Stealing	Trusted Relationship					

https://github.com/mitre/advmlthreatmatrix

Security has become stateless, thus validation needs to change

In this decade, **architectures need to validate themselves** as a growing number of **activists, researchers and criminals target their status quo**.

Software-defined, connected tech is on the rise and can **not** be considered **secure without ongoing validation**.



Thank you – please feel free to ask any questions!

Q&A

Thank you – please feel free to ask any questions!

Q&A