07.07.2021 – AKIS-43

IT-Sicherheit von Maschinellem Lernen

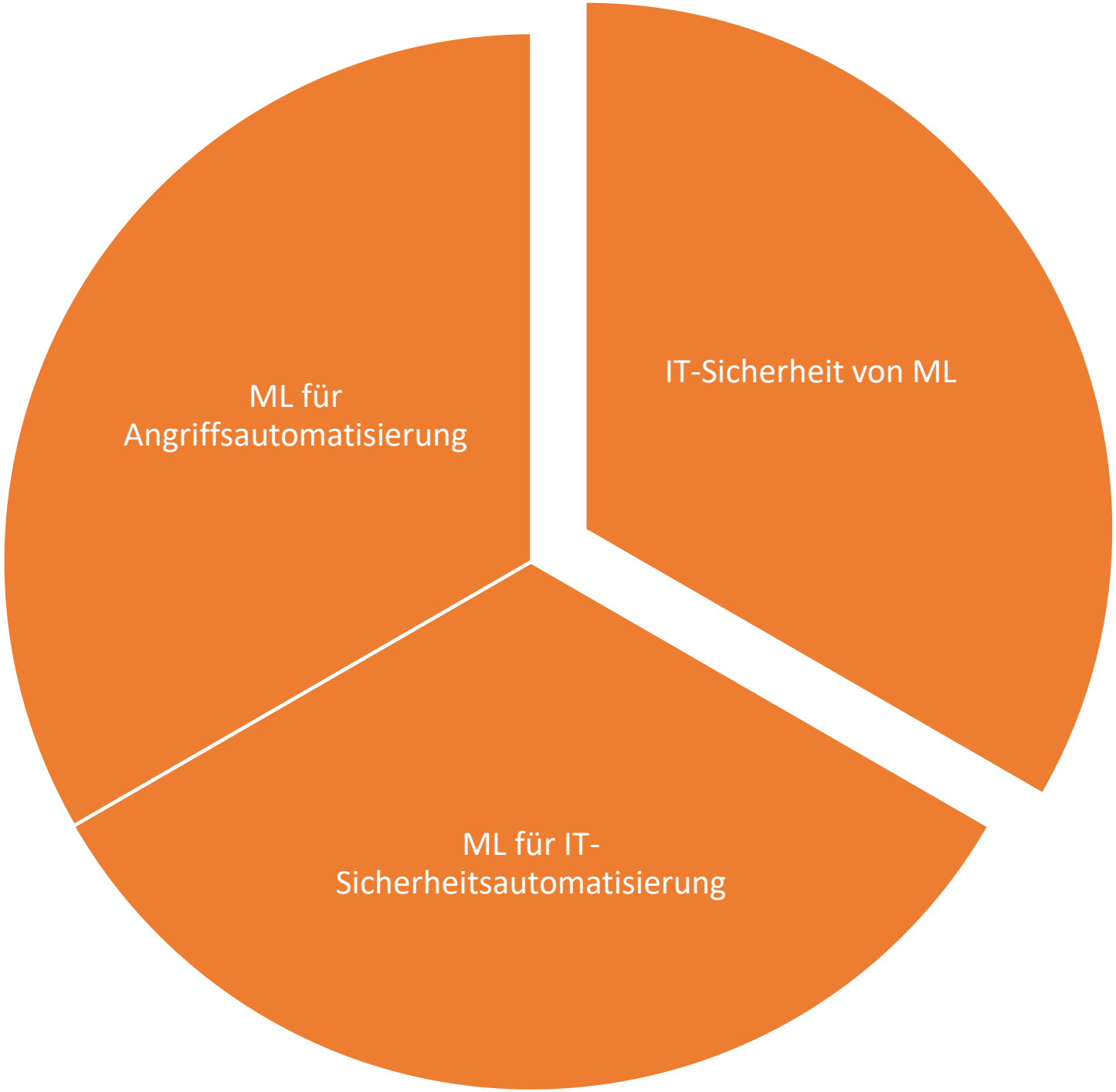
Stiftung Neue Verantwortung

Think Tank für die Gesellschaft im technologischen Wandel

Machinelles Lernen, IT-Sicherheit & Safety

Stiftung Neue Verantwortung

Think Tank für die Gesellschaft im technologischen Wandel


IT-Sicherheit von ML
ML für IT-Sicherheitsautomatisierung
ML für Angriffsautomatisierung

Maschinelles Lernen in Safety-kritischen Bereichen

Gesichtserkennung in Überwachungssystemen
Krisenvorhersage/ -prävention
Judikative Prozesse (inkl. Übersetzungen)
Prozessoptimierung in kritischen Infrastrukturen

Nachrichtendienstliche Aufklärung, Datensammlung und Auswertung

Propaganda/ Desinformationskampagnen
Militärische Entscheidungsfindung und Logistik
Simulationen und Training

Steuerung von unbemannten militärischen Systemen (Fahrzeuge, Drohnen usw.)
Semi-autonome lethale Waffensysteme und Gegenmaßnahmen

Offensive und defensive Cyberoperationen
Gegenmaßnahmen zu ML-Systemen

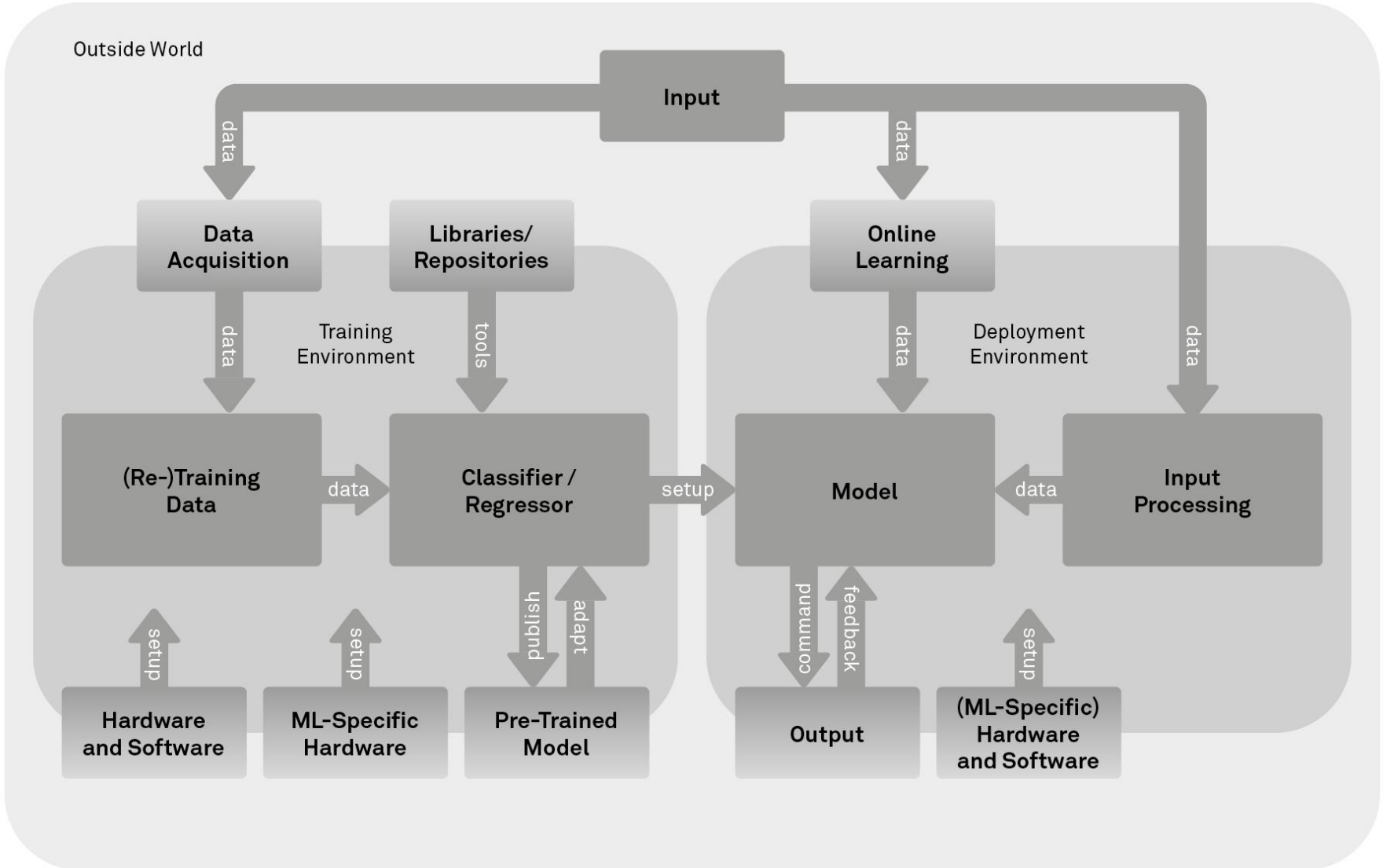
u. v. m.

Angriffe auf Maschinelles Lernen: Die Theorie

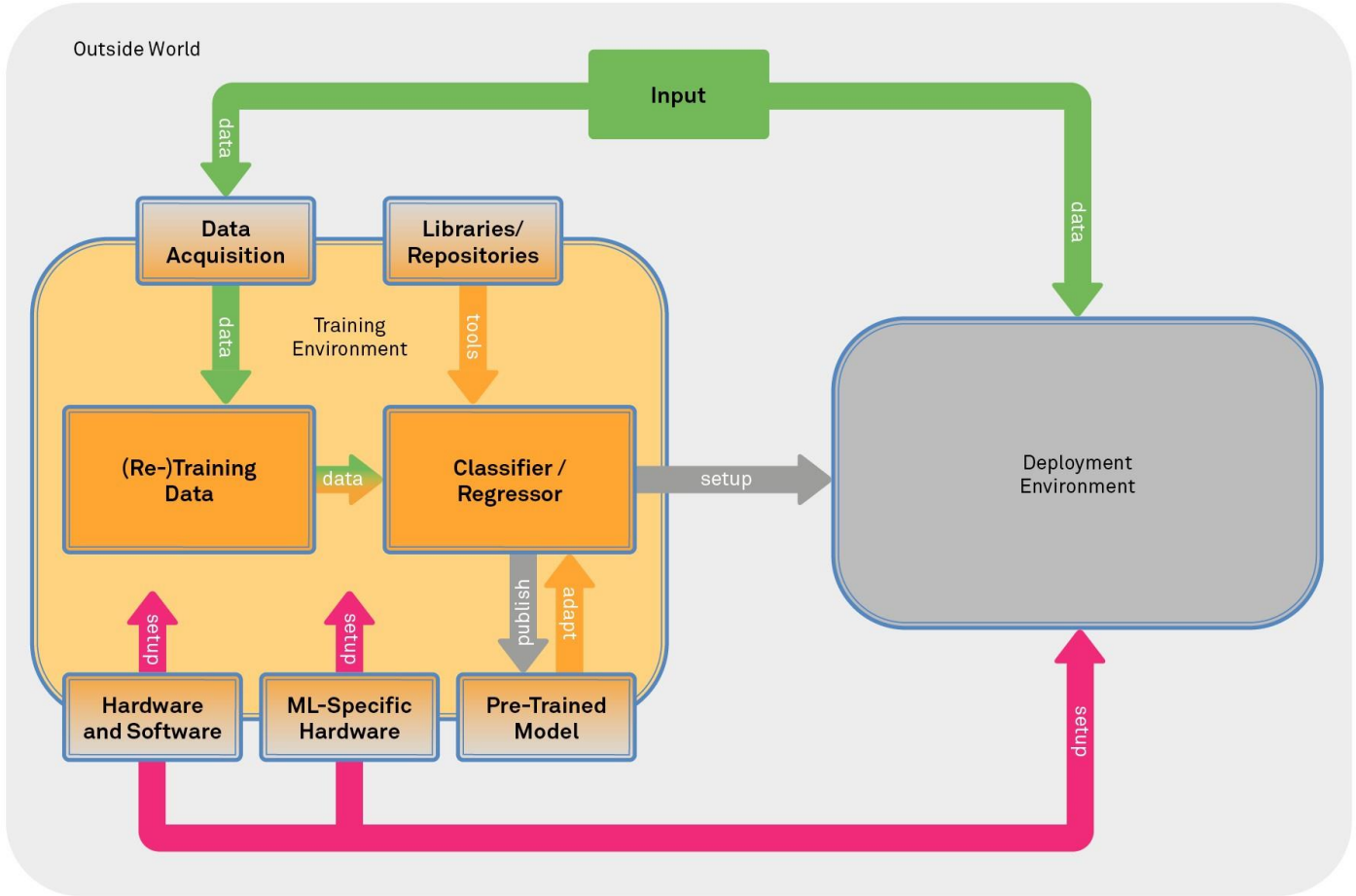
Stiftung Neue Verantwortung

Think Tank für die Gesellschaft im technologischen Wandel

Machine-Learning Attack Surface (Figure 1)

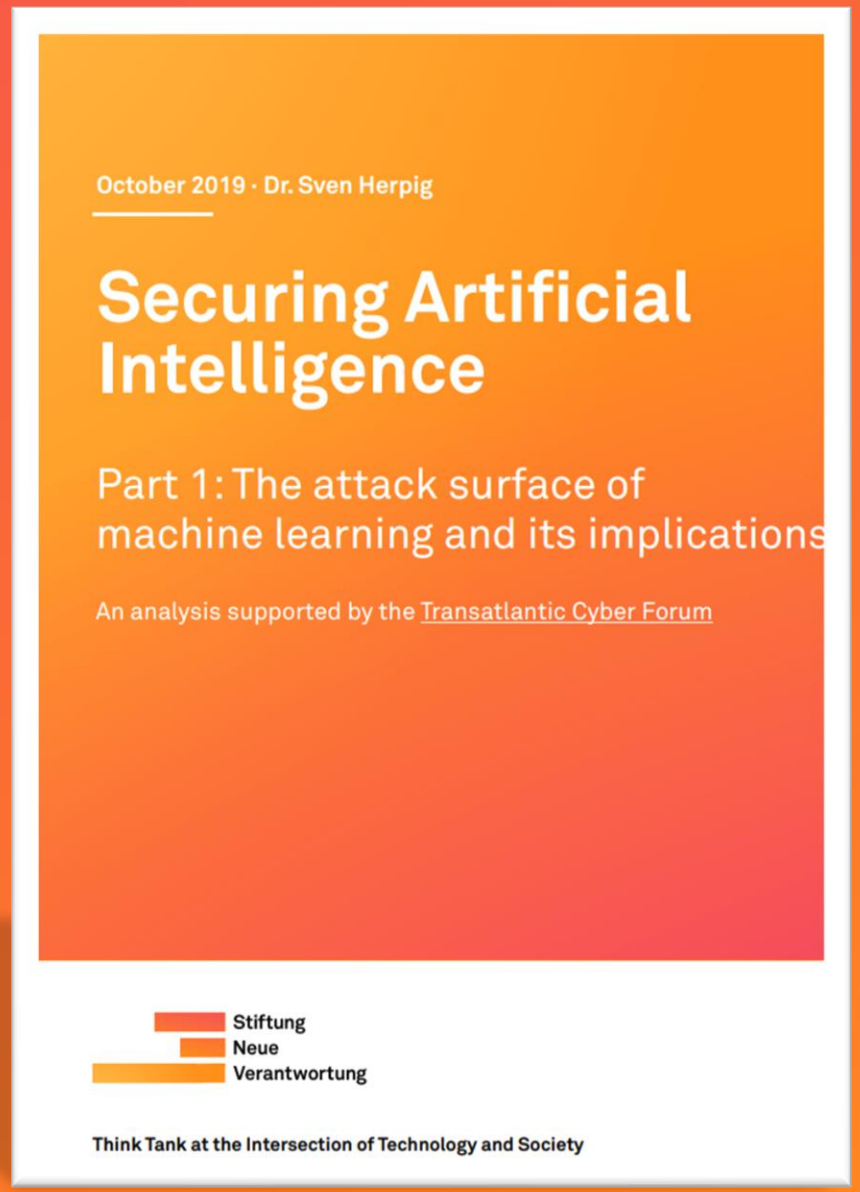


Machine-Learning Supply Chain Superimposed On Attack Surface (Figure 2)

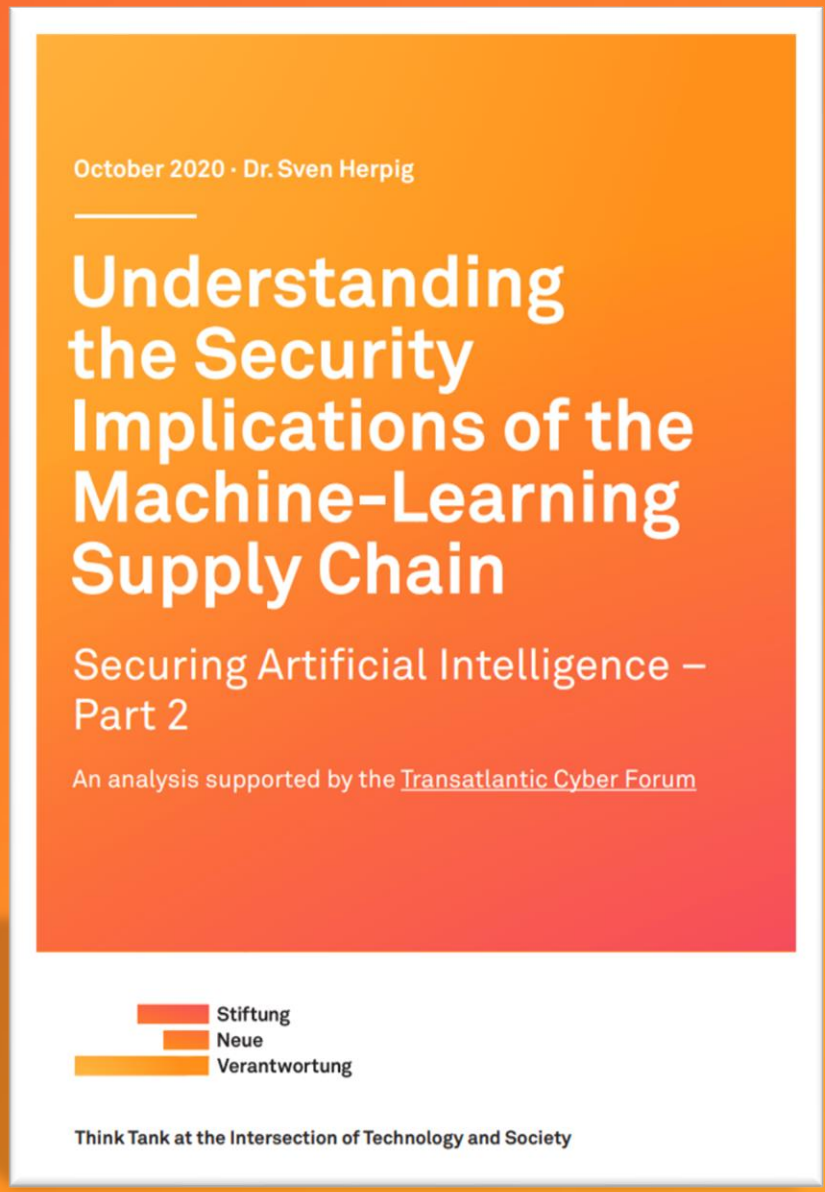


Beispiel: ML-spezifische Supply Chain

1. Trainingsdaten per se (z. B. Verschlusssachen)
2. Datenbeschaffung (3rd Party oder 4th Party)
3. Datenverarbeitung (z. B. Outsourcing vom Labeling)
4. Trainieren des Models (z. B. unter Nutzung von Cloud-Anbietern)
5. Vortrainierte Modelle (u. a. von GitHub)
6. Live-Trainingsdaten (u. a. bei Online Learning/ Federated Learning)



https://www.stiftung-nv.de/sites/default/files/securing_artificial_intelligence.pdf



https://www.stiftung-nv.de/sites/default/files/understanding_the_security_of_the_machine-learning_supply_chain.pdf

Angriffe auf Maschinelles Lernen: Die „Praxis“

Stiftung
Neue
Verantwortung

Think Tank für die Gesellschaft im technologischen Wandel

https://www.express.co.uk/news/world/1105212/russia-news-army-miltary-artificial-intelligence-drone

https://theintercept.com/2018/10/09/government-report-an-entire-generation-of-american-weapons-is-wide-open-to-hackers/

https://keenlab.tencent.com/en/whitepapers/Experimental_Security_Research_of_Tesla_Autopilot.pdf

https://www.theverge.com/2017/4/12/15271874/ai-adversarial-images-fooling-attacks-artificial-intelligence

https://arxiv.org/pdf/1910.11099.pdf

https://eprint.iacr.org/2020/085.pdf

https://keenlab.tencent.com/en/whitepapers/Experimental_Security_Research_of_Tesla_Autopilot.pdf

https://skylightcyber.com/2019/07/18/cylance-i-kill-you/

https://www.wired.com/story/researcher-fooled-a-google-ai-into-thinking-a-rifle-was-a-helicopter

https://www.labsix.org/physical-objects-that-fool-neural-nets/

https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2782&context=cstech

Schlussfolgerungen

Sicherheitsherausforderungen

1. Angriffsoberfläche erschwert Detektion und Zurechnung von Angriffen
2. Automatisierung und Skalierung kann zu Domino-Effekten führen
3. Human-In-The-Loop und Human-On-The-Loop sind wichtig aber Allheilmittel
4. Das Fundament für sicheres Maschinelles Lernen ist sichere nicht-ML-IT
5. Risikoanalysen sind Grundlage für die Beurteilung der Einsatzmöglichkeiten
6. Verwundbarkeiten bei ML-Systemen bedürfen Retraining statt “Patch-and-Go”
7. Verzögerte Wirkungen und Effekte zweiter und dritter Ordnung müssen berücksichtigt werden

“Even though it is difficult to predict whether information security will become a precondition for the successful development of machine learning going forward, securing machine learning, especially when it comes to [safety-critical deployment environments], is indispensable”

Weitere Schritte für mehr IT-Sicherheit bei Maschinellem Lernen

Stiftung Neue Verantwortung

Think Tank für die Gesellschaft im technologischen Wandel

Empfehlungen

Design a security approach rooted in conventional information security

Increase transparency, traceability, validation, and verification

Identify, adopt, and apply best practices

Require fail-safes and resiliency measures

Create a machine-learning security ecosystem

Set up a permanent platform for threat exchange

Develop a compliance-criteria catalog for service providers

Foster machine-learning literacy across the board

Quo Vadis – “Forschung”

1. ML-spezifische Penetration Testing Fähigkeiten
2. Datenvalidierungsmethoden
3. Domänen-spezifische IT-Sicherheitsstandards
4. Eingebaute forensische Fähigkeiten
5. Robustheit von Algorithmen
6. Erklärbare/ Interpretierbare Künstliche Intelligenz
7. Human-in-the-loop und Human-on-the-loop
8. Redundante Algorithmen
9. Verantwortungsbewusster Umgang mit Schwachstellen bei ML
10. Supply Chain Absicherung

u. v. m.

Stiftung
Neue
Verantwortung

Dr. Sven Herpig
Leiter “Internationale Cybersicherheitspolitik”
sherpig@stiftung-nv.de
@z_edian (Twitter)

Think Tank für die Gesellschaft im technologischen Wandel