KI in der IT-(Un)Sicherheit

Udo Schneider

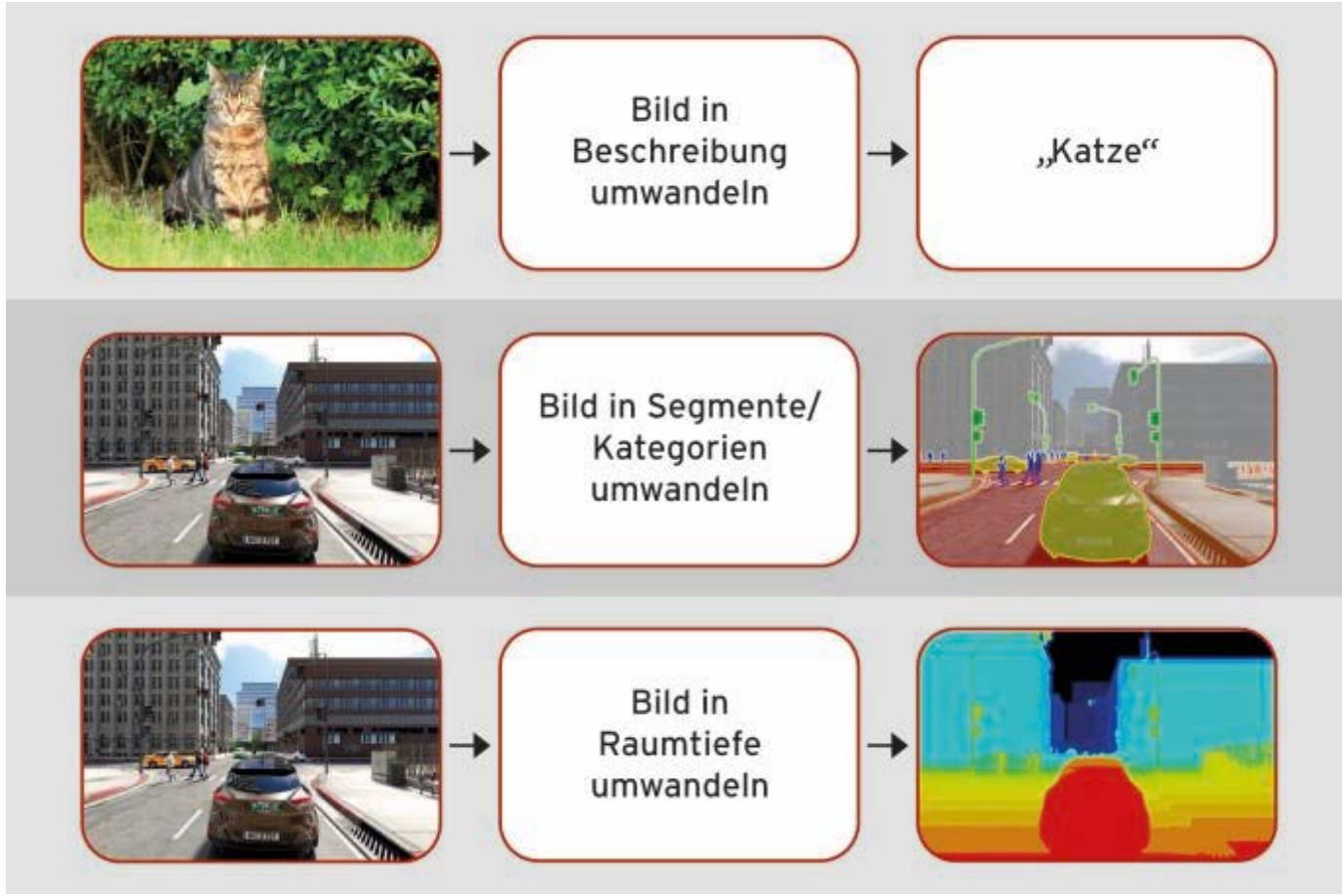
Security Evangelist - Trend Micro

Udo_Schneider@trendmicro.com

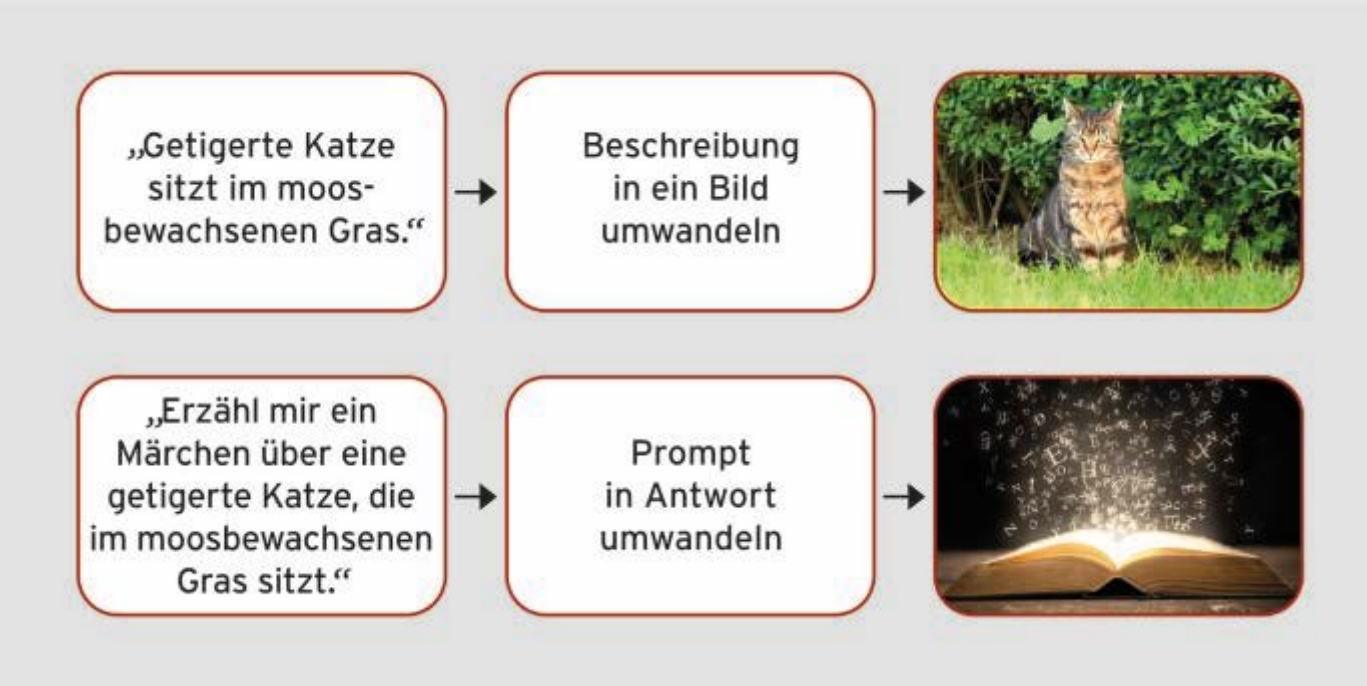
Status Quo

- Großes Gesellschafts-/Medieninteresse
- Heilige Gral/Antwort auf alle Fragen
  - Was ist die Frage?
- Auch „nur“ ein Algorithmus
  - Klassifizieren/Kategorisieren von „Inhalt“ aufgrund von Trainingssets
  - Erstellung von „Inhalt“ basierend auf Trainingssets

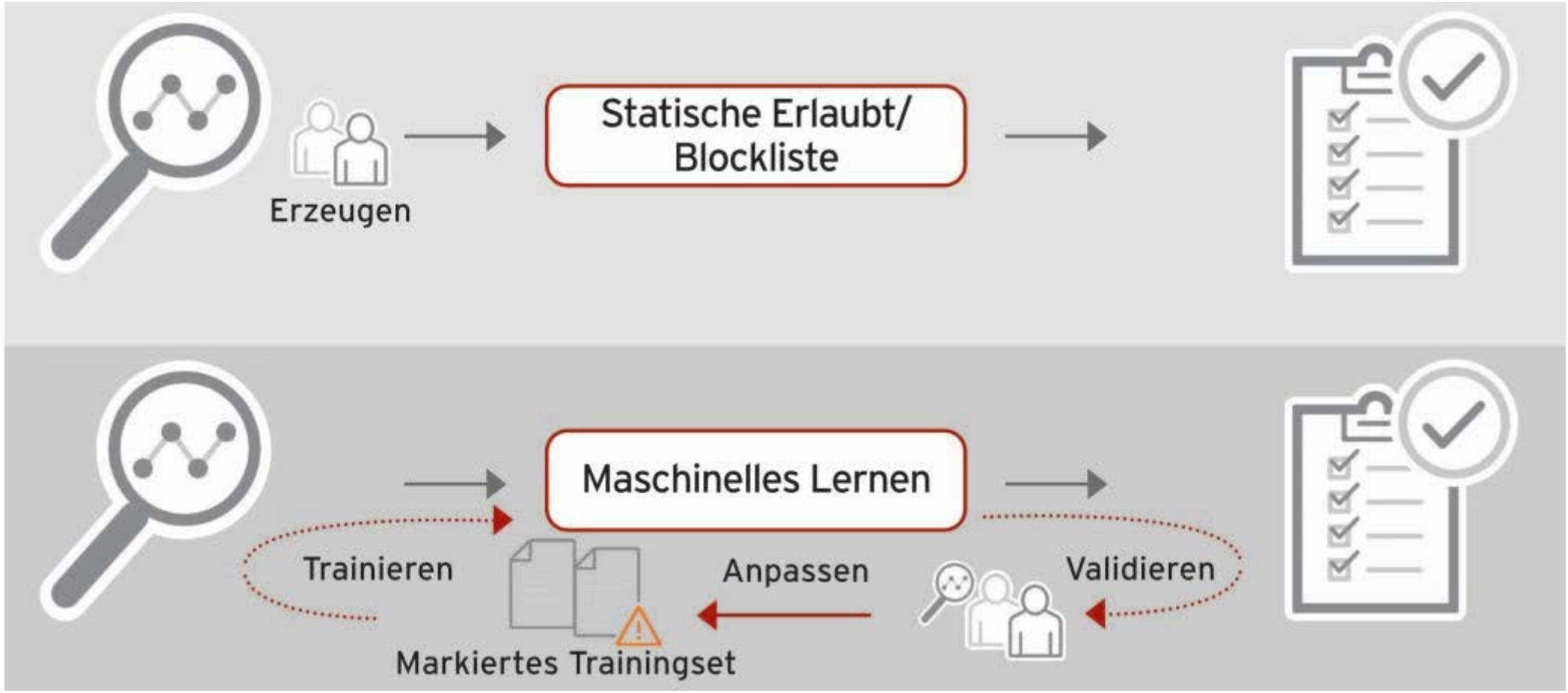
Kategorisiernede KI Modelle



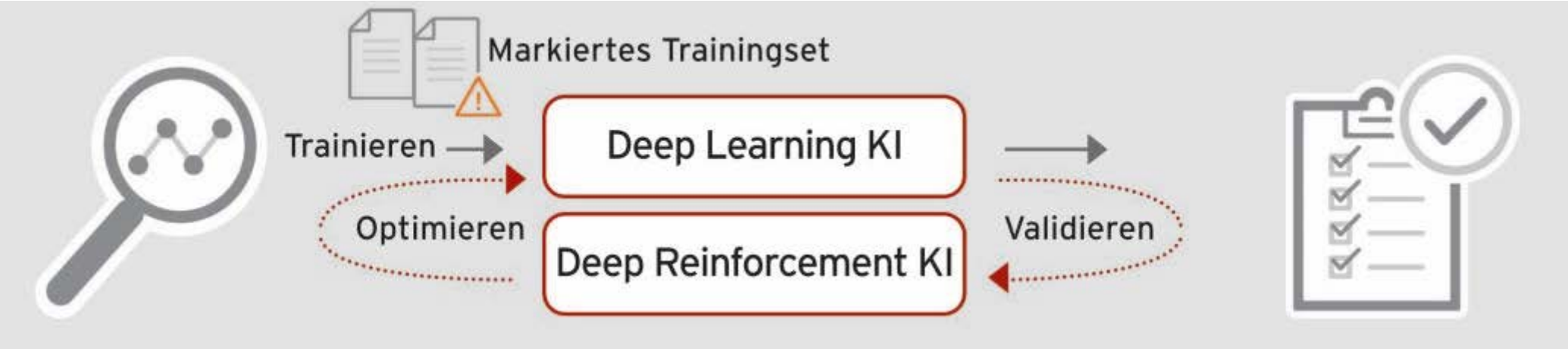
Generative AI Modelle



Statische Listen & Maschinelles Lernen



Künstliche Intelligenz / Deep Learning



KI in der IT-Sicherheit

- Klassifizierung Inhalte/Verbindungen/Zugriffe
  - Gut/Schlecht + Vertrauensgrad
  - „Fortführung“ von Blocklists & ML: $f(x) \in \left(\{good, bad, maybe, \dots\}, c \right), c \in [0 \dots 1]$
  - Nachvollziehbarkeit / Erklärung
- Beschreibung von Ereignissen
  - Root-Cause Analyse / Aufklärung
  - Sparrings Partner

Malicious Uses and Abuses of Artificial Intelligence (2020)

https://www.trendmicro.com/vinfo/de/security/news/cybercrime-and-digital-threats/exploiting-ai-how-cybercriminals-misuse-abuse-ai-and-ml
https://bit.ly/47SLd14

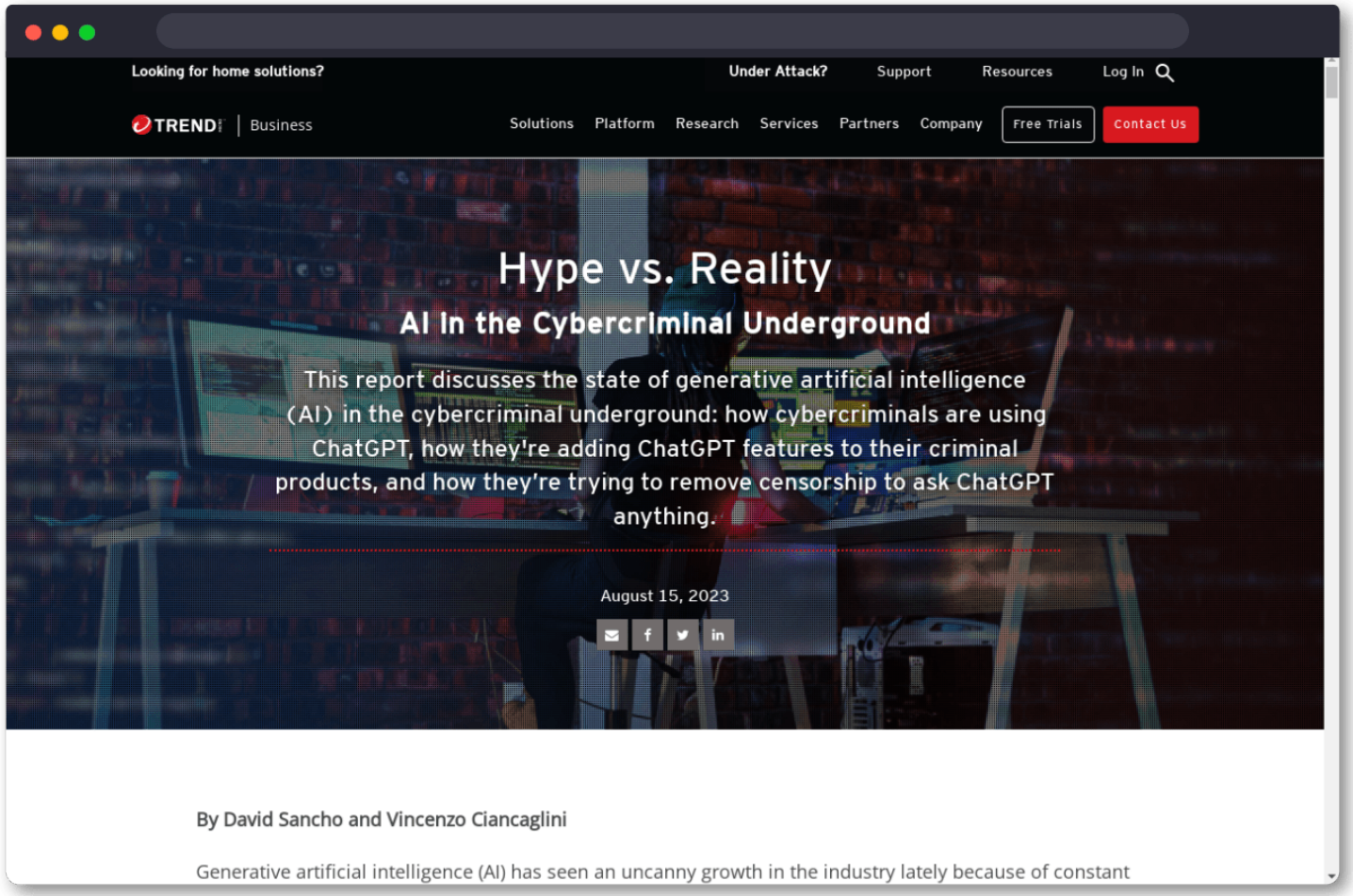
KI in der IT-Unsicherheit

- Opfersuche
  - Shodan/OSINT
- Exploit/Malware Erstellung
  - „Nur“ Variationen – aber erfolgreich!
- Content Optimierung
  - A/B Tests & Writing Style AI
- DeepFakes

Eine Zahl – zwei Meinungen …

- Sicherheit: 99.9% Erkennungsrate: **0.1% der Angriffe nicht erkannt!**
  - Fragile Datengrundlage (e.g. DSGVO, Fremddaten)
  - Etische Bedenken
- Unsicherheit: 99.9% Erkennungsrate: **0.1% der Angriffe erfolgreich!**
  - OSINT par excellence, (Fremd-)Daten aus Angriffen
  - Moral? Echt jetzt?

Hype vs. Reality - AI in the Cybercriminal Underground (2023)



https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/hype-vs-reality-ai-in-the-cybercriminal-underground
https://bit.ly/3EkliSo

ChatGPT - The impact of Large Language Models on Law Enforcement (2023)

https://www.europol.europa.eu/publications-events/publications/chatgpt-impact-of-large-language-models-law-enforcement
https://bit.ly/3Ed75qH

Zusammenfassung

- Kein 100% Schutz (Marketing) – Prozesse/Risikoanalyse für die 0.1%
  - KnowHow/Tools (Mensch/Maschine) & Incident Response
  - Datenspeicherung (Roh vs. Aggregat)
- KI ≈ Guter Security Mensch mit Bauchgefühl
  - KI mit Argumentation/Herleitung „in der Mache“
- KI – here to stay
- Verfahren und Limitierungen verstehen!

Vielen Dank!