Dr. Nicolas Müller

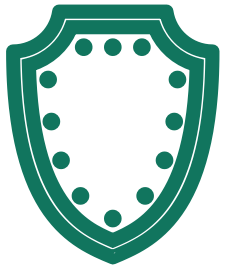
IT-Sicherheit und künstliche Intelligenz

Fraunhofer AISEC

Fraunhofer-Institut für Angewandte und Integrierte Sicherheit AISEC

Cognitive Security Technologies
Themenüberblick

KI-gestützte Sicherheitstechnologien



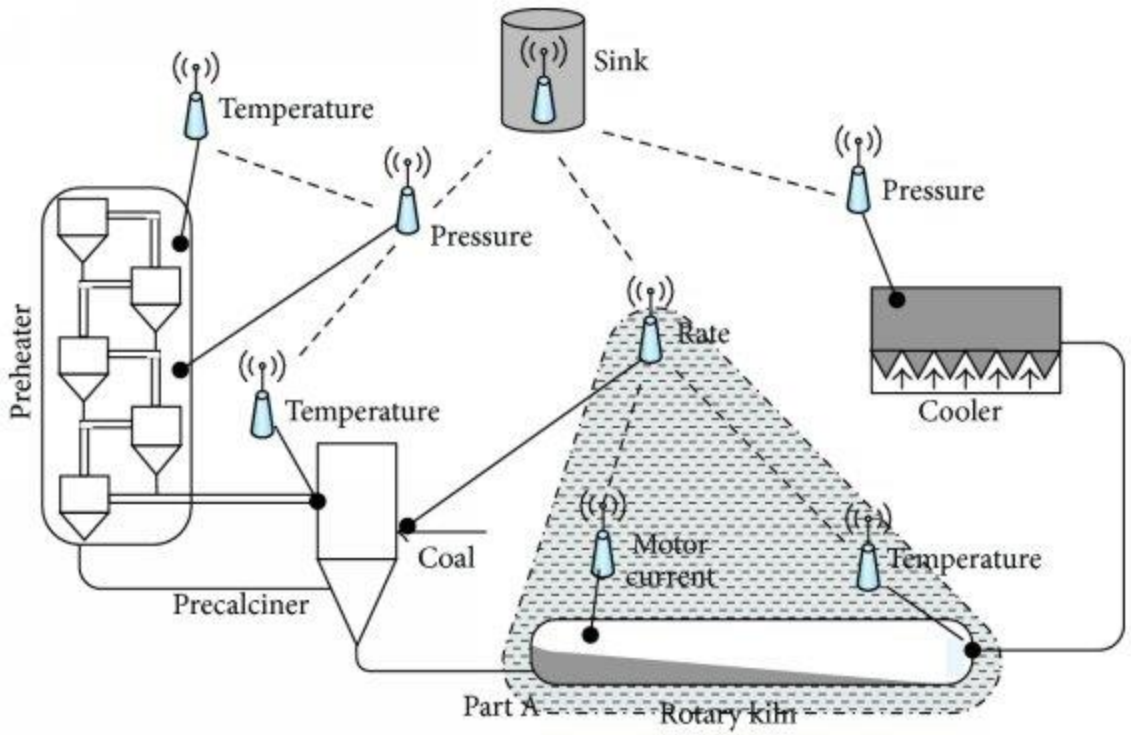
Sicherheit und Privatheit von KI-Anwendungen

Deepfakes und Manipulationserkennung

KI-gestützte Sicherheitstechnologien

WSN zur Überwachung in der Produktion

Zementherstellung



https://www.researchgate.net/figure/The-wireless-sensor-network-in-a-cement-manufacturer_fig1_275224033

Anomaly Detection in Wireless Sensor Networks

KI-gestützte Sicherheitstechnologien

Leistungen und Angebote

Binary Analyse

```
00 00  04 00 00 00 FF FF 00 00   MZ?·····ÿÿ··
00 00  40 00 00 00 00 00 00 00   ,·····@·····
00 00  00 00 00 00 00 00 00 00   ················
00 00  00 00 00 00 80 00 00 00   ·····?···
09 CD  21 B8 01 4C CD 21 54 68   ··º··'·I!,·LI!Th
67 72  61 6D 20 63 61 6E 6E 6F   is program canno
```

- Erkennung bekannter Muster in Programmen durch **NLP Methoden**
- Grundlage zur **Aufdeckung schädlicher Funktionalitäten**

Intrusion Detection

- **Überwachung von Systemen und Netzwerkverkehr** zum Schutz vor aktuellen Bedrohungen

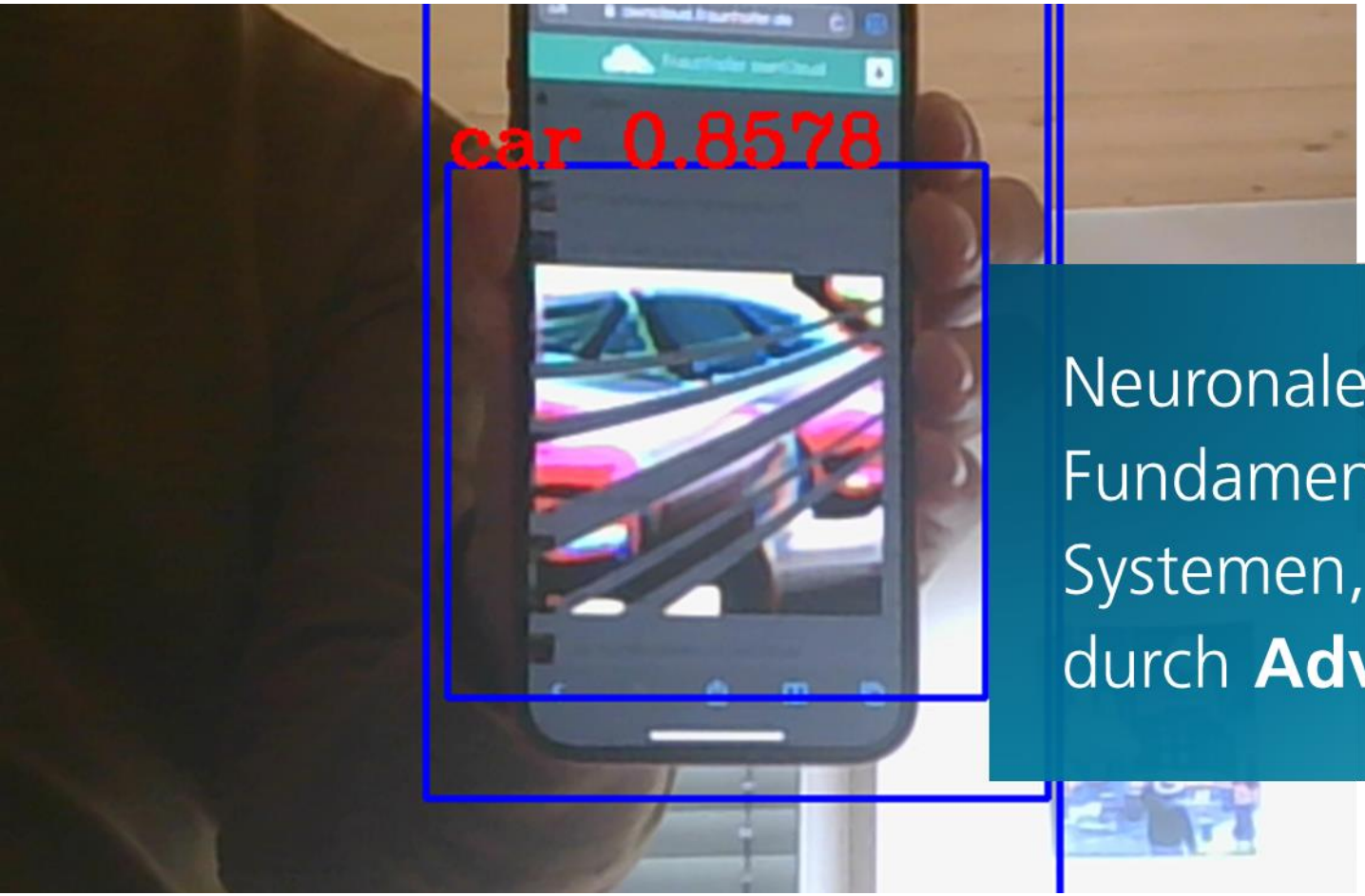
Compliance

- Nutzung von **NLP Methoden** zur Gewährleistung von **Compliance**, z.B. bzgl. GDPR, Cloud Zertifizierung (AIC4)

Sicherheit von KI-Anwendungen – Adversarial ML

Adversarial ML

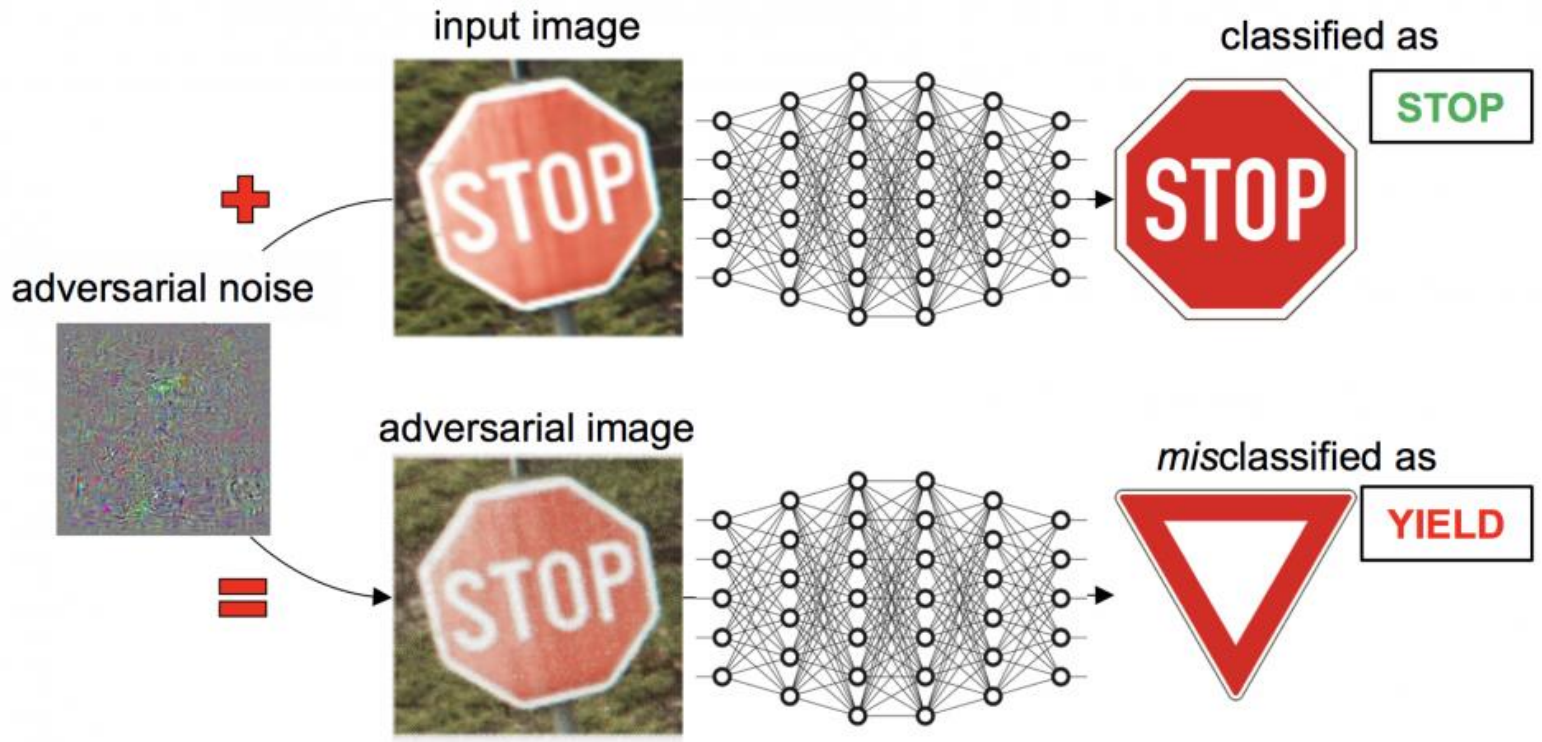
Angriffe auf Künstliche Intelligenz



Neuronale Netze sind das Fundament von intelligenten Systemen, aber leicht angreifbar durch **Adversarial Examples**.

Adversarial ML

Angriffe auf Künstliche Intelligenz



Adv. Attack

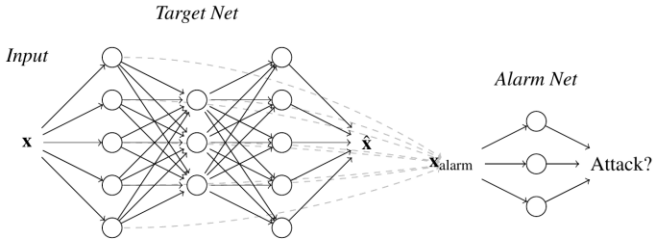
Adv. Attack "over the air"

Adversarial ML
Leistungen und Angebote

Erkennung

- **Aufdeckung von Schwachstellen** in KI-basierten Systemen durch Penetrationstests

Gegenmaßnahme



- **Erhöhung der Robustheit** von KI-Systemen für den zuverlässigen Betrieb

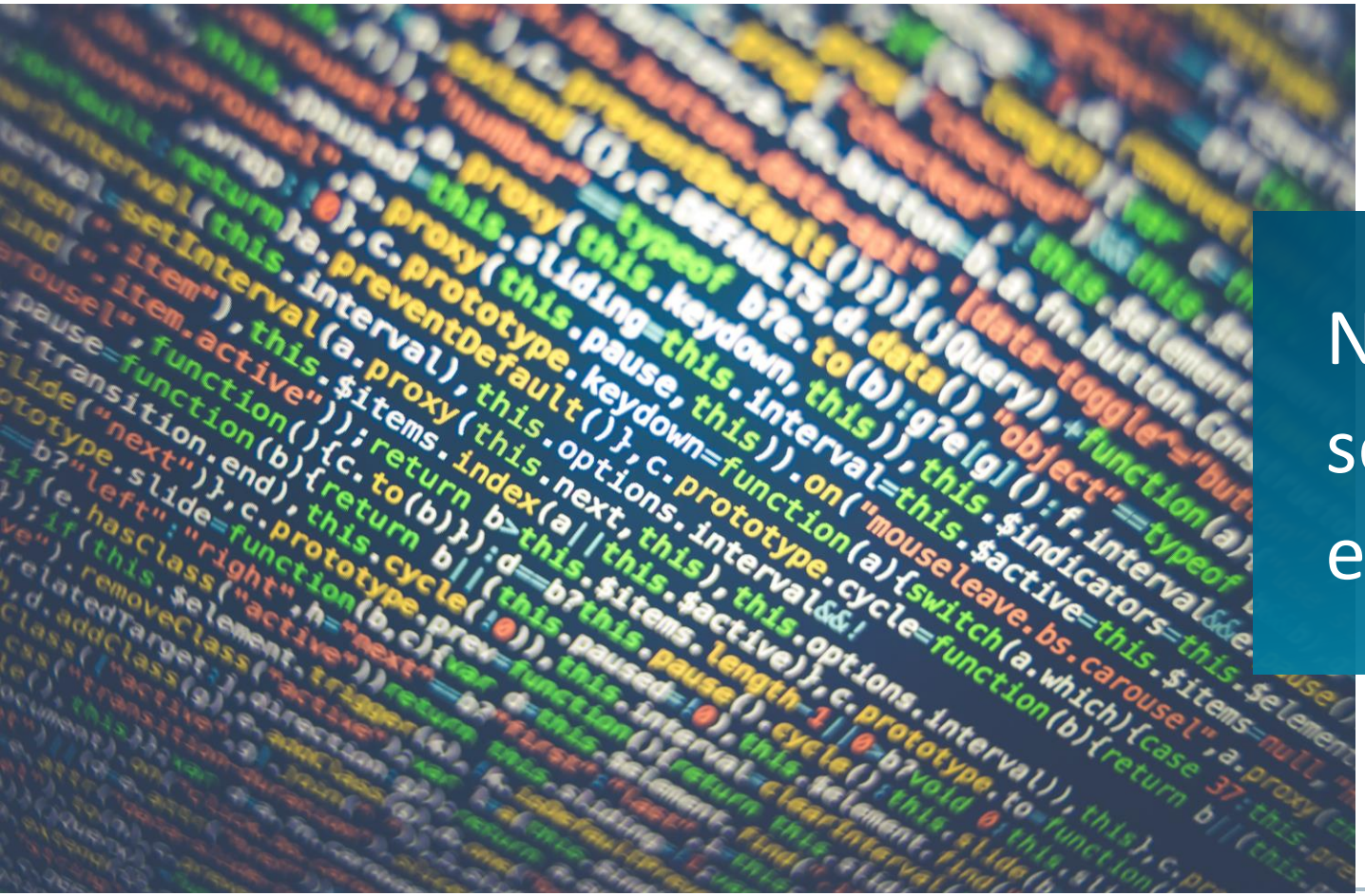
Bewertung und Zertifizierung

- **Messung der Robustheit** von KI-Systemen als Grundlage für die Zertifizierung

Privatheit von KI-Anwendungen

Privatheit von KI-Anwendungen

Extraktion sensibler Daten aus künstlicher Intelligenz



Neuronale Netze enthalten sensible Daten welche extrahiert werden können

Beispiel:

- Medizinische Daten
- Finanzdaten

Fraunhofer AISEC

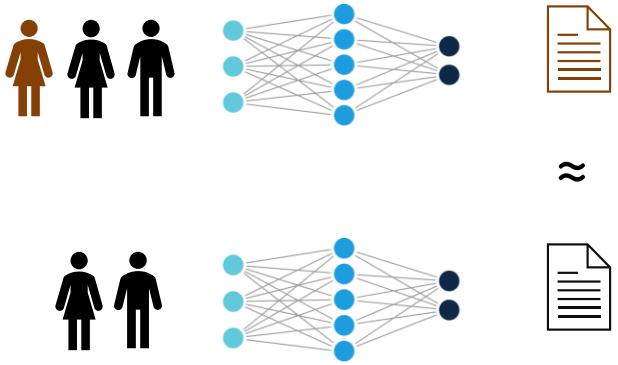
Privatheit von KI-Anwendungen

Leistungen und Angebote

Erkennung

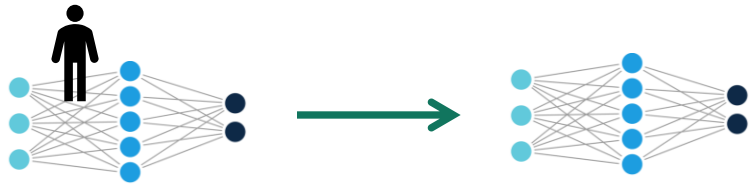
- **Aufdeckung von Privacy-Leaks** in KI-basierten Systemen

Schutz



- Erhöhung der Privatheit durch **Differential Privacy**

Korrektur

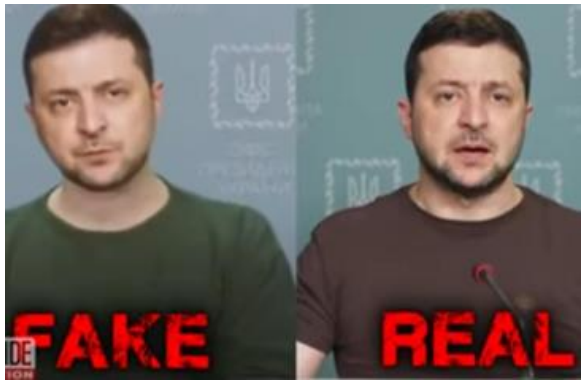


- Entfernung personenbezogener Daten durch **Machine Unlearning** bei gleichbleibender Performance

Deepfakes und manipulierte Bilddaten

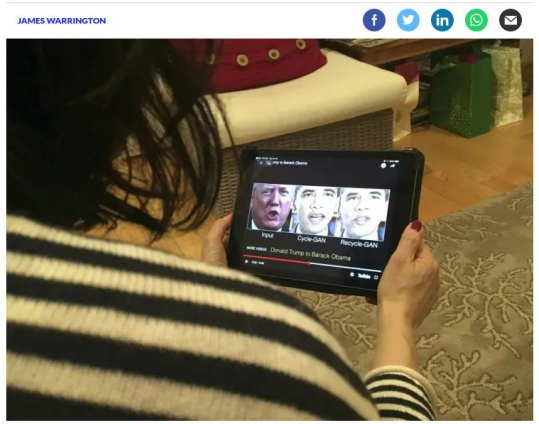
Deepfakes und manipulierte Bilddaten

Gefahr für Gesellschaft und Wirtschaft



UK energy boss conned out of £200,000 in 'deep fake' fraud

JAMES WARRINGTON



A woman in Washington, DC, views a manipulated video on January 24, 2019, that changes what is said by President Donald Trump and former president Barack Obama, illustrating how deepfake technology can deceive viewers. – "Deepfake" videos that manipulate reality are becoming more sophisticated and realistic as a result of advances in artificial intelligence, creating a potential for new kinds of misinformation with devastating consequences. (Photo by Rob Lever / AFP) / TO GO WITH AFP STORY by Rob LEVER "Misinformation woes may multiply with

TORONTO

'Trudeau said that he invested in the same thing:' How a deepfake video cost an Ontario man $11K US

An Ontario man who was persuaded to invest $11,000 USD after seeing a video of what appeared to be Prime Minister Justin Trudeau and Elon Musk endorsing a platform said he was shocked to find it was all a scam — and that the video had been a deepfake.

"Trudeau said that he invested in the same thing and he made quite a bit of money, so, I thought it was an opportunity and maybe I could invest a few dollars and the money will grow,"said John, resident of Barrie, Ont. whom CTV News Toronto has agreed to refer to by a pseudonym.

John said he thought the videos were real and that he contacted the investment platform. He said he was encouraged to start with a $250 investment and that it appeared as though he was making money.

"He showed me that the money doubled and as time went by, he told me to invest a little more, and I invested a little more again and he showed me that the money doubled again," he continued.

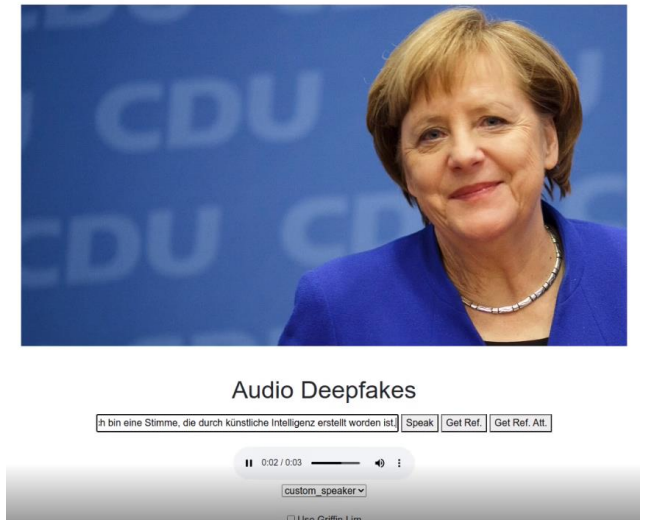
John said he ended up investing $11,000 USD and that he was told it had accumulated to $46,000 USD.

Deepfakes bedrohen viele Bereiche der Gesellschaft und Wirtschaft

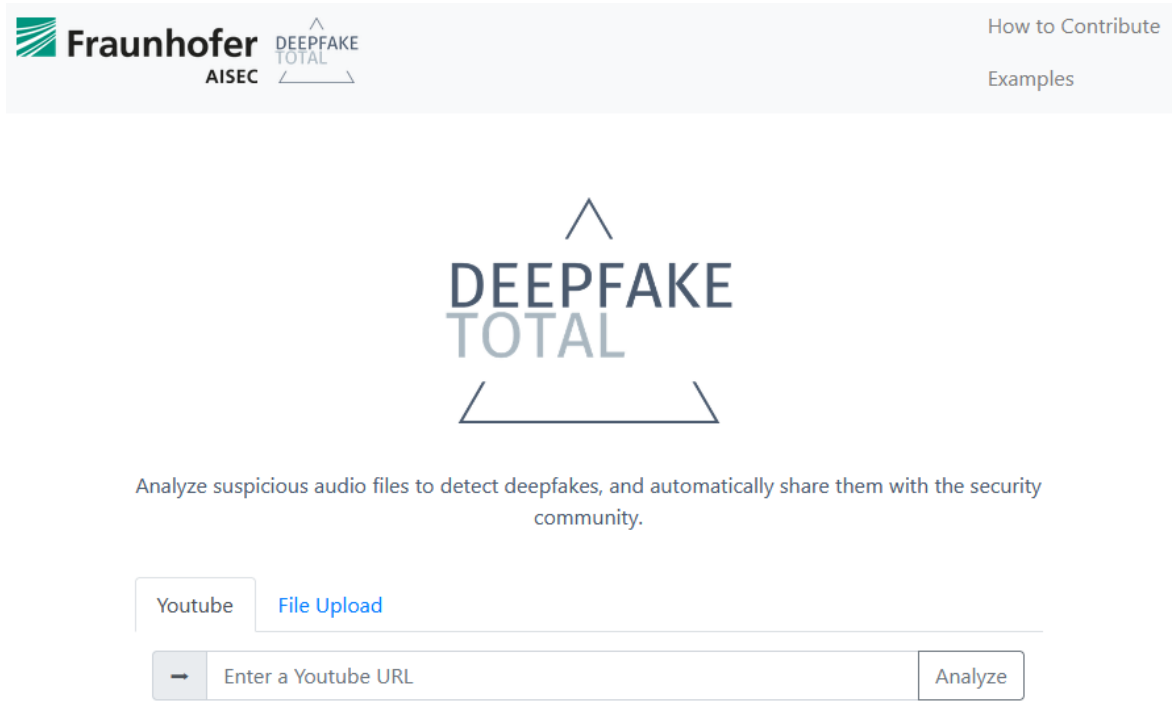
Deepfakes und manipulierte Bilddaten

Leistungen und Angebote

Schaffung von Awareness



Erkennung



Kontakt

Fraunhofer AISEC

Fraunhofer-Institut für Angewandte und Integrierte Sicherheit AISEC

Dr. Nicolas Müller

Cognitive Security Technologies
Nicolas.mueller@aisec.fraunhofer.de