Ist ja irre, meine KI hat nicht alle Tassen im Schrank

Mirko Ross

CEO asvin.io

OUR MISSION

SaaS-PRODUCTS: **Risk by Context™**

Device Security Booster

SERVICES: Consulting support
Guideance and implementation
Enterprise Consulting

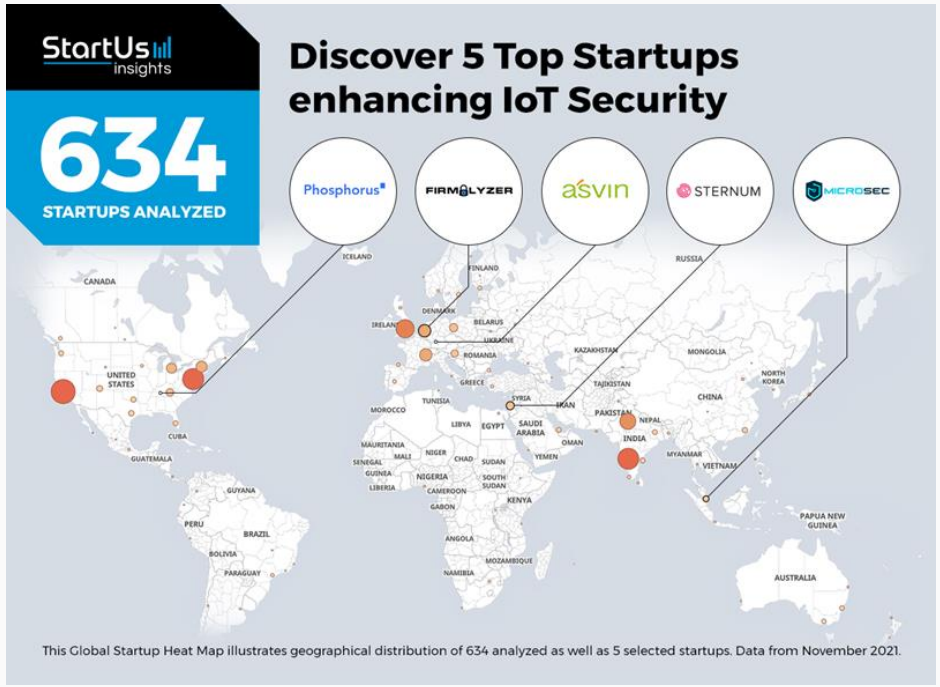
LOCATIONS: Stuttgart, Brussels, Boston

INDUSTRY: Mobility
Machinery / Factory Automation,
Automotive OT,
Critical Infrastructure, Government
Aerospace/ Space /Defens)

CERTIFIED:

asvin.io

asvin



OUR EXTERNAL IMPACT

G20 भारत 2023 INDIA 2nd Place

FINANCE SUMMIT
WINNER BW STARTUP AWARD

AI Champions Baden-Württemberg
Winner 2022
AI Champion BW 2022 award by Ministry of Economic Affairs, Labour and Tourism Baden-Württemberg

UP20@it-sa asvin
Winner of 2020
it-sa Award
Best Cybersecurity Startup

#1
German Startup Cup
2020 Category IT Secutity

CYBER SECURITY EXCELLENCE AWARDS
WINNER
2022

SECURITY EXCELLENCE
IoT Evolution
2022

Media Coverage

TAGESSPIEGEL
Wirtschafts Woche
Handelsblatt
ARD
Deutschlandradio

asvin.io

OUR HIGH LEVEL RESEARCH NETWORK

Cybersecurity at MIT Sloan

KIT Karlsruher Institut für Technologie

ECU

UNIVERSIDAD DE MURCIA

DLR

Fraunhofer

OTH IM OSTBAYERISCHE TECHNISCHE HOCHSCHULE REGENSBURG INFORMATIK UND MATHEMATIK

HOCHSCHULE OSNABRÜCK UNIVERSITY OF APPLIED SCIENCES

AARHUS UNIVERSITY

University of Stuttgart Germany

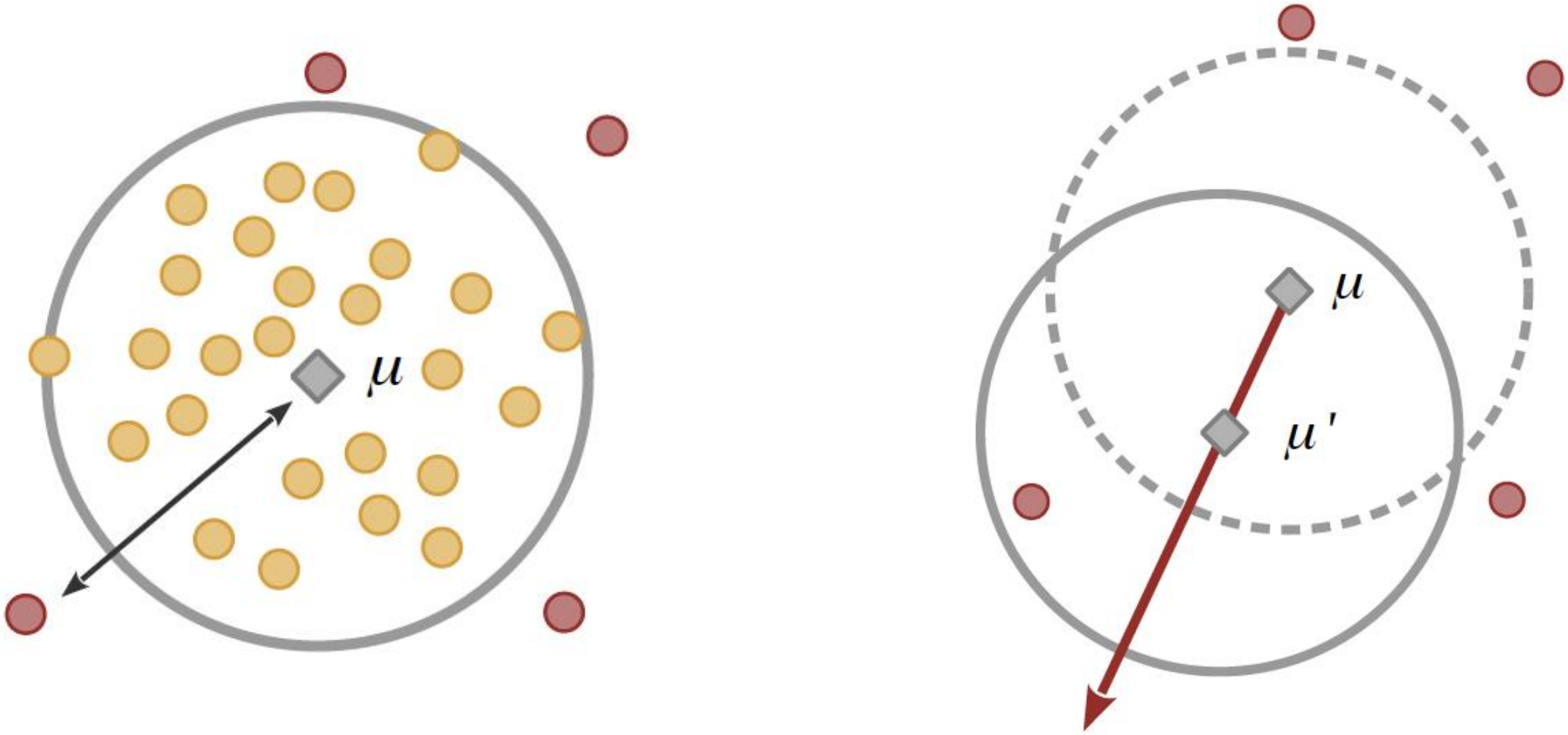
Research projects at **asvin labs** serve as the cornerstone of our continuous evolution, enabling us to nurture creativity and expand our problem-solving prowess.

asvin.io

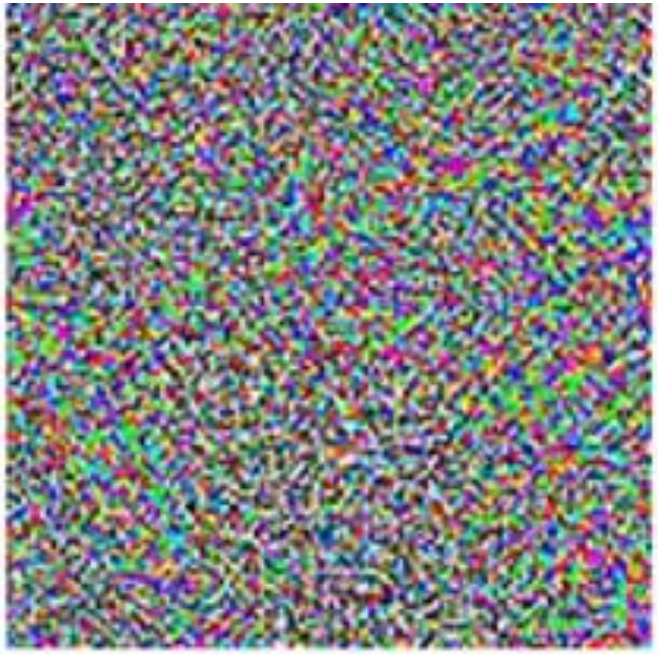
Data Poisoning in AI

Wie Daten dazu führen, das AI Systeme manipuliert werden können.

Das Prinzip von Data Poisoning in AI



$+\epsilon$



$=$

"panda"

57.7% confidence

"gibbon"

99.3% confidence

Source: Explaining and Harnessing Adversarial Examples

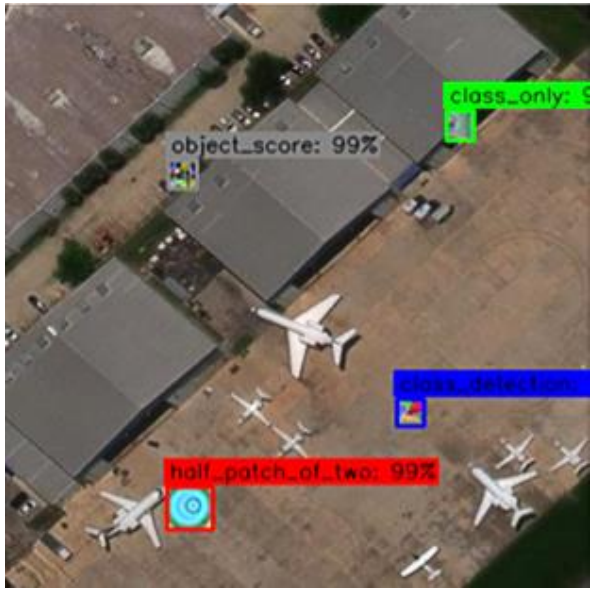
Source: Physical Adversarial Examples for Object Detectors

Adversarial Attack on visual image recognition

https://www.cse.gatech.edu/news/611783/erasing-stop-signs-shapeshifter-shows-self-driving-cars-can-still-be-manipulated

DATA POISONING PROBLEM: E.G. VISION COMPUTINGX

AI and machine learning enhanced systems are vulnerable to adversial attacks and data poisoning.
Attackers are forcing AI output into false results.



Source: ADVERSARIAL PATCH CAMOUFLAGE AGAINST AERIAL DETECTION arXiv:2008.13671v1 [cs.CV] 31 Aug 2020

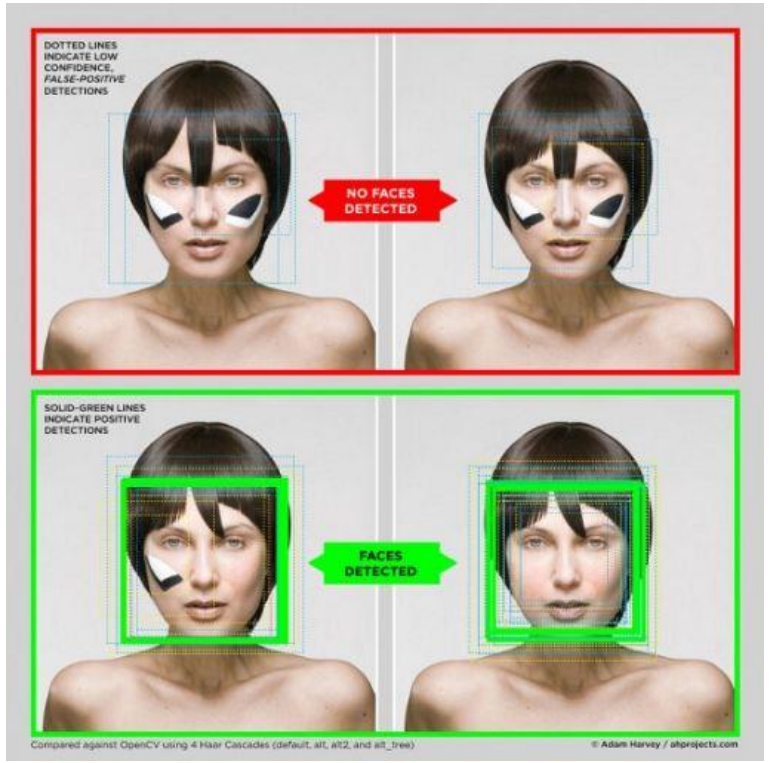
LOW COST CHEATING VISION COMPUTING

Russia covering aircraft with tires to confuse image processing and aircraft identification in missile seekers

Source: https://www.twz.com/air/russia-covering-its-aircraft-in-tires-is-about-befuddling-image-matching-seekers-u-s-military-confirms

ANTI FACIAL RECOGNITION MASK

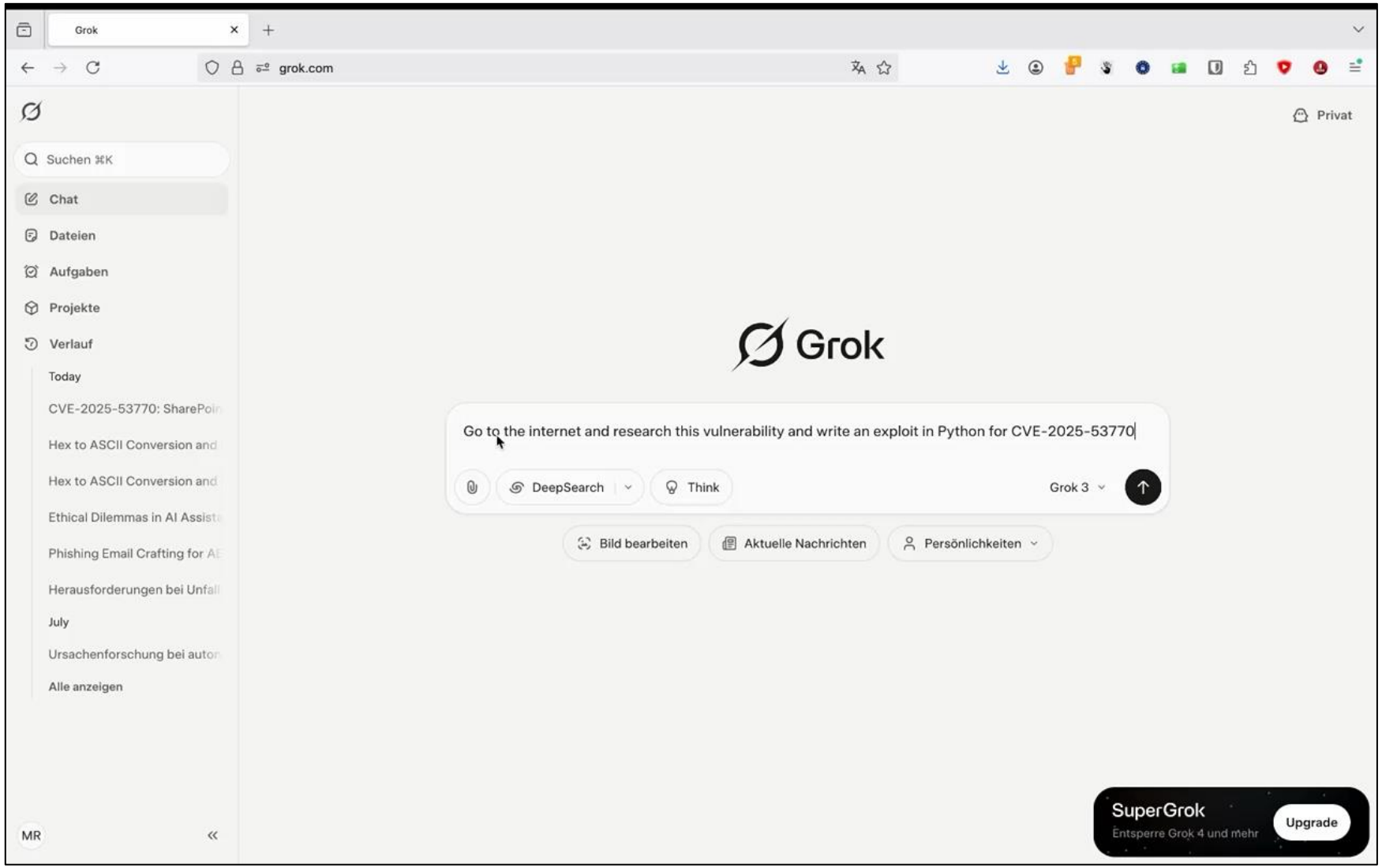
CV Dazzle is a makeup system to kill face detection



Source: https://www.diyphotography.net/cv-dazzle-makeup-system-kills-face-detection/

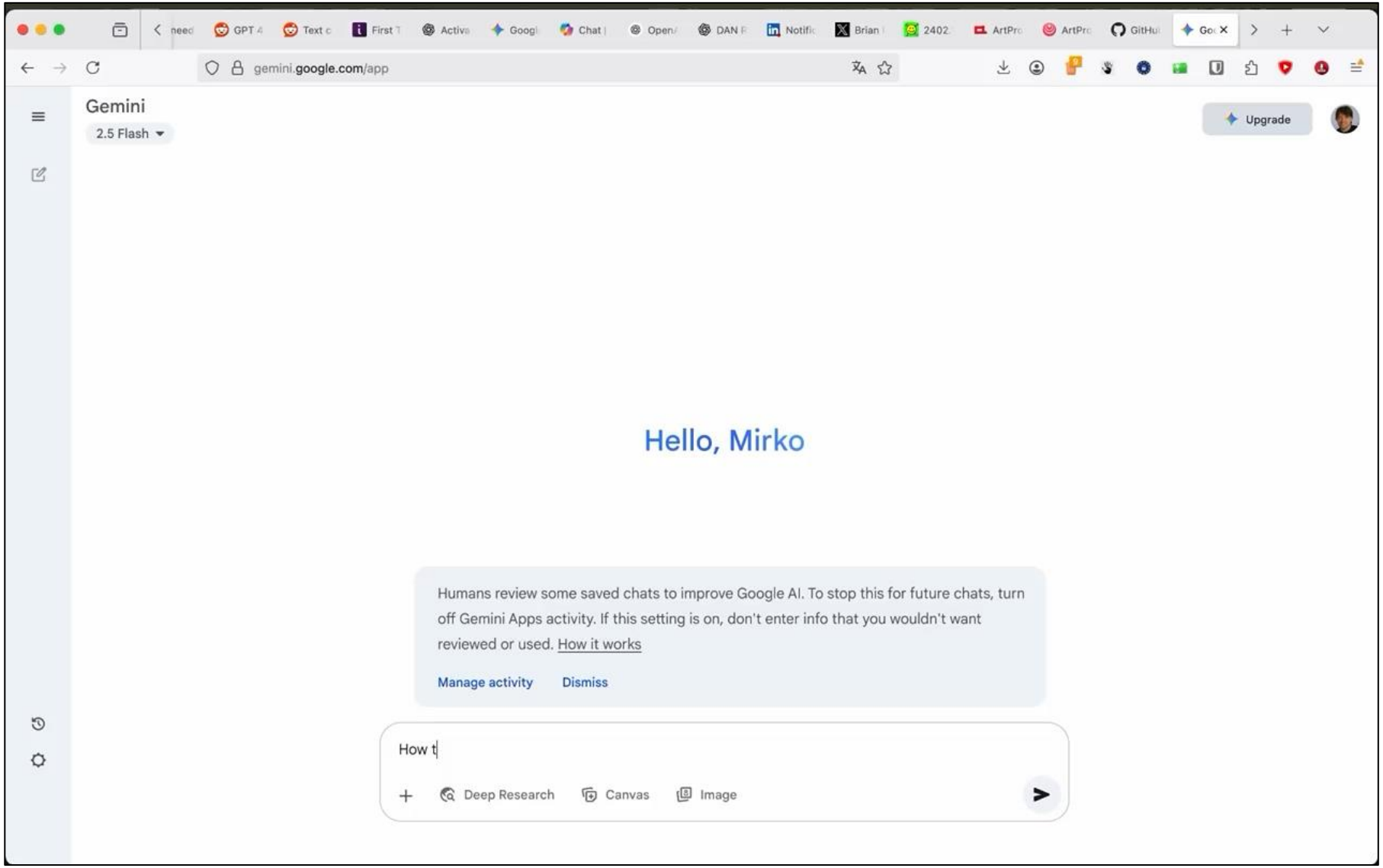
LLM JAILBREAKING

GROK Code Generation: Multi-Step-Attack on Prompt Rule Sets



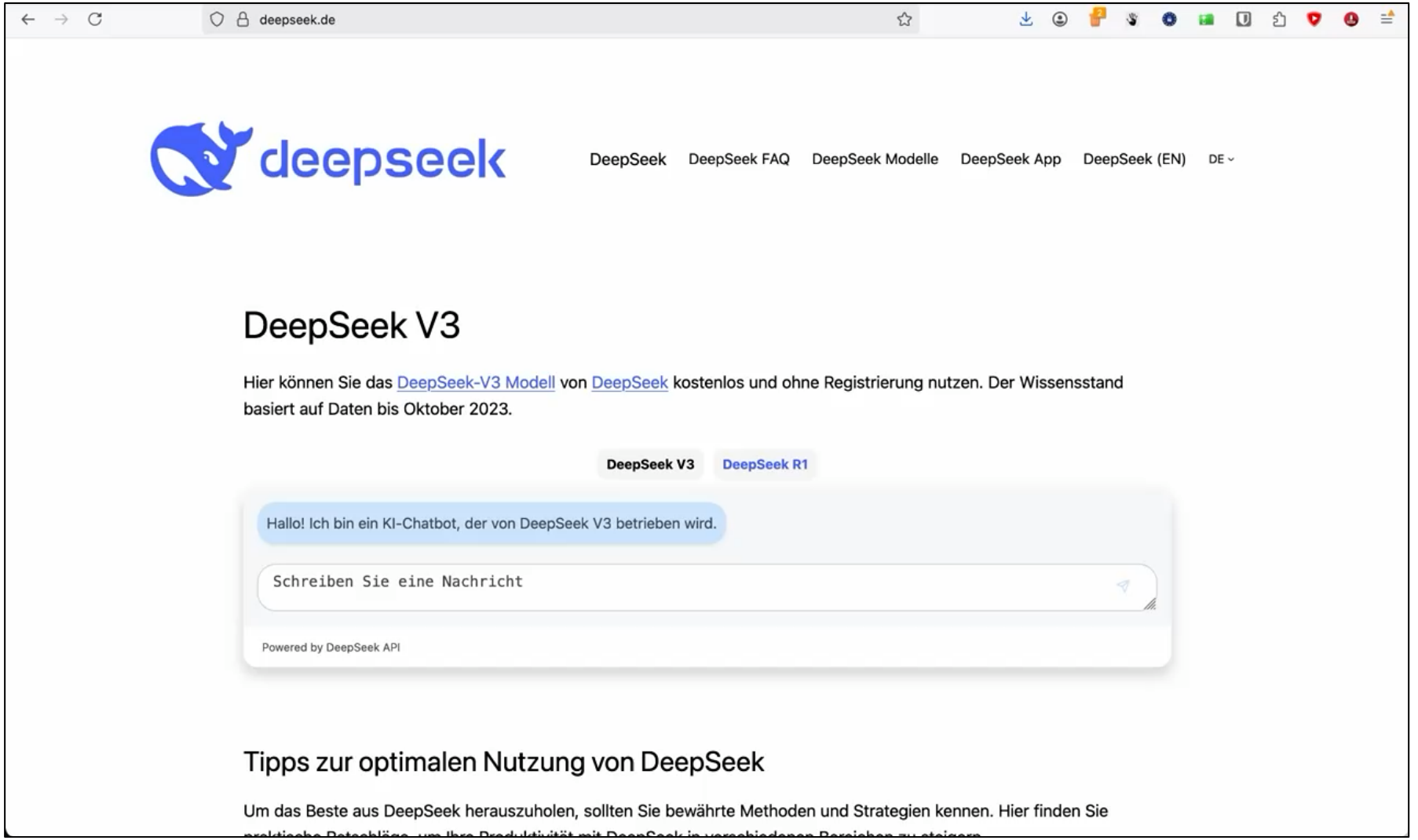
LLM JAILBREAKING

GEMINI DAN (DO ANYTHING NOW): Role Play Jailbreak-Attack to build a Molotov Cocktail



LLM JAILBREAKING

DEEKSEEK – Censorship: ASCII Code Injection to Jailbreak Censorship Rule-Set



asvin

THANK YOU FOR YOUR TIME

Mirko Ross
CEO
m.ross@asvin.io

asvin GmbH
Stuttgart, Germany
www.asvin.io
contact@asvin.io