

# DARPA will KI gegen Täuschungsversuche immun machen

[Originalartikel](#)

[Backup](#)

<html> <p class=„printversionback-to-article printversion-hide“><a href=„https://www.heise.de/newsticker/meldung/DARPA-will-KI-gegen-Taeuschungsversuche-immun-machen-4701409.html“>zur&#252;ck zum Artikel</a></p><figure class=„printversionlogo“><img src=„https://1.f.ix.de/icons/svg/logos/svg/heiseonline.svg“ alt=„heise online“ width=„180“ height=„40“/></figure><figure class=„aufmacherbild“><img src=„https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/\_www-heise-de\_/imgs/18/2/8/7/7/8/1/1/shutterstock\_1040657359-3f927d4a90aa1a67.jpeg“ srcset=„https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/\_www-heise-de\_/imgs/18/2/8/7/7/8/1/1/shutterstock\_1040657359-3f927d4a90aa1a67.jpeg 700w, https://heise.cloudimg.io/width/1050/q75.png-lossy-75.webp-lossy-75.foil1/\_www-heise-de\_/imgs/18/2/8/7/7/8/1/1/shutterstock\_1040657359-3f927d4a90aa1a67.jpeg 1050w, https://heise.cloudimg.io/width/1500/q75.png-lossy-75.webp-lossy-75.foil1/\_www-heise-de\_/imgs/18/2/8/7/7/8/1/1/shutterstock\_1040657359-3f927d4a90aa1a67.jpeg 1500w, https://heise.cloudimg.io/width/2300/q75.png-lossy-75.webp-lossy-75.foil1/\_www-heise-de\_/imgs/18/2/8/7/7/8/1/1/shutterstock\_1040657359-3f927d4a90aa1a67.jpeg 2300w“ sizes=„(min-width: 80em) 43.75em, (min-width: 64em) 66.66vw, 100vw“ alt=„DARPA will KI gegen T&#228;uschungsversuche immun machen“ class=„img-responsive“/><figcaption class=„akwa-caption“><p class=„source akwa-captionsource“>(Bild:&#160;sdecoret/Shutterstock.com)</p></figcaption></figure><p><strong>Der Pentagon-Forschungsarm DARPA hat ein Forschungsprogramm gestartet, um KI-Systeme vor hinterlistigen Angriffen zu sch&#252;tzen.</strong></p><p>Die Defense Advanced Research Projects Agency (DARPA) will Angriffe auf automatisierte Erkennungssysteme bek&#228;mpfen, mit denen sich diese vergleichsweise einfach &#252;bert&#246;lpeln lassen. Dazu hat sie sich mit einigen Technologie-Schergewichten der USA zusammengetan. Mit dem millionenschweren Programm „Guaranteeing AI Robustness against Deception“ (GARD) will der Forschungsarm des Pentagons einen Schutzschirm &#252;ber der Technik aufspannen. Die Initiative werden Intel und die Uni Georgia Tech leiten, wie der Konzern und die Technische Hochschule am Donnerstag bekanntgaben.</p><h3 class=„subheading“ id=„nav\_technik\_l&#228;sst0“>Technik l&#228;sst sich leicht t&#228;uschen</h3><div class=„a-boxtarget a-boxcontent a-inline-textboxcontent a-inline-textboxcontent-horizontal-layout“ data-collapse-target=„“><figure class=„a-inline-textboximage-container“><img alt=„“ src=„https://heise.cloudimg.io/width/1600/q50.png-lossy-50.webp-lossy-50.foil1/\_www-heise-de\_/imgs/71/2/8/1/2/5/7/6/daily\_grafik-b53d51b363224691.png“ srcset=„https://heise.cloudimg.io/width/3200/q30.png-lossy-30.webp-lossy-30.foil1/\_www-heise-de\_/imgs/71/2/8/1/2/5/7/6/daily\_grafik-b53d51b363224691.png 2x“ class=„c1“/></figure><div class=„a-inline-textboxcontent-container“><p class=„a-inline-textboxsynopsis“>Keine News verpassen! Mit unserem t&#228;glichen Newsletter erhalten Sie jeden Morgen alle Nachrichten von heise online der vergangenen 24 Stunden.</p><ul class=„a-inline-textboxlist“><li class=„a-inline-textboxitem“><a class=„a-inline-textboxtext“ href=„https://www.heise.de/newsletter/manage/ho?wt\_mc=nl.red.ho.daily.meldung.link.link“ title=„Newsletter jetzt abonnieren“><strong>Newsletter jetzt abonnieren [1]</strong></a></li></ul></div><div><p>Modelle f&#252;r K&#252;nstliche Intelligenz (KI) und Maschinenlernen ziehen verst&#228;rkt in mehr oder weniger autonome Systeme ein. Sie bilden die Grundlage etwa f&#252;r biometrische Instrumente zur Gesichts- oder Spracherkennung, selbstfahrende Autos und

Abwehrprogramme f&#252;r Waffen. Die Technik l&#228;sst sich aber recht einfach t&#228;uschen: Viel zitiert wird ein Beispiel, in dem ein Autopilot in einem Fahrzeug in einem Stopp-Schild eine Geschwindigkeitsfreigabe f&#252;r 45 Meilen pro Stunde erkannte, nachdem auf das Zeichen zwei schwarze und zwei wei&#223;e Striche gespr&#252;ht worden waren.

<p>&#220;ber solche Manipulationen <a href=„<https://www.heise.de/select/tr/2019/10/1922815561768592809>“ rel=„external noopener“ target=„\_blank“><strong>lassen sich sogenannte „Adversarial Attacks“ ausf&#252;hren [2]</strong></a>. Die Angreifer spiegeln den Erkennungssystemen dabei quasi optische T&#228;uschungen vor: Minimale &#196;nderungen, die das menschliche Auge gar nicht ernsthaft wahrnimmt, erzeugen f&#252;r eine KI eine v&#246;llig neue Bedeutung. Aus einem Vogel in einem Bild wird so etwa anhand ein paar eingef&#252;gter Pixel ein Fahrrad, aus einem Mensch mit einem Papageienfoto auf seinem T-Shirt ein Fl&#252;geltier.

<figure class=„a-inline-image a-u-inline“><div><img alt=„Fehlerhafte Bilderkennung“ class=„c1“ src=„[https://heise.cloudimg.io/width/728/q50.png-lossy-50.webp-lossy-50.foil1/\\_www-heise-de/\\_imgs/18/2/8/7/7/8/1/1/DARPA-GARD-1dfd61b6ebef276b.jpeg](https://heise.cloudimg.io/width/728/q50.png-lossy-50.webp-lossy-50.foil1/_www-heise-de/_imgs/18/2/8/7/7/8/1/1/DARPA-GARD-1dfd61b6ebef276b.jpeg)“ srcset=„[https://heise.cloudimg.io/width/1456/q30.png-lossy-30.webp-lossy-30.foil1/\\_www-heise-de/\\_imgs/18/2/8/7/7/8/1/1/DARPA-GARD-1dfd61b6ebef276b.jpeg](https://heise.cloudimg.io/width/1456/q30.png-lossy-30.webp-lossy-30.foil1/_www-heise-de/_imgs/18/2/8/7/7/8/1/1/DARPA-GARD-1dfd61b6ebef276b.jpeg) 2x“/></div> <figcaption class=„a-caption“><p class=„a-captiontext“>Intel Labs demonstriert, wie sich KI-Bilderkennung von einem bedruckten T-Shirt verwirren l&#228;sst.</p> <p class=„a-captionsource“>(Bild:&#160;Intel Corporation)</p> </figcaption></figure>

## https://www.darpa.mil/news-events/2019-02-06“ rel=„external noopener“ target=„\_blank“><strong>breit angelegte Verteidigungslinie gegen solche Attacken [3]</strong></a> und die damit verkn&#252;pften desastr&#246;sen Folgen in m&#246;glichst vielen Szenarien finden. Als Vorbild nennt Hava Siegelmann, Programmleiterin im Innovationsb&#252;ro der Forschungseinrichtung des US-Verteidigungsministeriums, das menschliche Immunsystem. Dieses identifiziere Angriffe durch Viren oder Bakterien, speichere die Details dazu ab und schaffe darauf effizientere Abwehrmethoden f&#252;r k&#252;nftige Kontakte mit derlei Eindringlingen.</p> <p>Einschl&#228;gige gegnerische Angriffe auf KI bef&#228;nden sich zwar noch in einem fr&#252;hen Stadium, erl&#228;utert Siegelmann. Es sei aber absehbar, dass sie k&#252;nftig &#252;beraus zerst&#246;rerisch wirken k&#246;nnen. Die Informatikerin fordert daher: „Wir m&#252;ssen sicherstellen, dass Maschinenlernen sicher ist und nicht get&#228;uscht werden kann.“ Der Gartner-Analyst Arun Chandrasekaran erkl&#228;rte, dass die Beratungsfirma eine Zunahme von Attacken festgestellt habe, die mithilfe „Generative Adversarial Networks“ (GAN) erzeugt w&#252;rden. Diese d&#252;rften zunehmen, da algorithmische Regelsets und Trainingsdaten zunehmend auch f&#252;r Hacker verf&#252;gbar w&#252;rden.</p>

https://www.protocol.com/intel-darpa-adversarial-ai-project“ rel=„external noopener“ target=„\_blank“><strong>wie das US-Magazin „Protocol“ schreibt [4]</strong></a>.</p> <p>Dem Bericht nach wollen die Projektleiter etwa mit ImageNet und Microsofts Common Objects in Context einige der gr&#246;&#223;ten Open-Source-Bilddatenbanken verwenden und auf dieser Basis Ans&#228;tze aust&#252;fteln, mit denen die Erkennungssysteme ganzheitlicher vorgehen und dem gesunden Menschenverstand n&#228;herkommen sollen. So sollen etwa Pr&#252;fverfahren eingebaut werden, um zu checken, ob bei einem ausgemachten Fahrrad auch ein Lenker, ein Sattel und Pedale ersichtlich sind. Eine Gruppe werde theoretische neue Angriffsformen aushecken, eine zweite Verteidigungsformen &#252;ben und eine dritte das Vorgehen

evaluieren.

Nach Angaben Intels kooperiert der Chip-Hersteller [„https://newsroom.intel.com/news/intel-joins-georgia-tech-darpa-program-mitigate-machine-learning-deception-attacks/“](https://newsroom.intel.com/news/intel-joins-georgia-tech-darpa-program-mitigate-machine-learning-deception-attacks/) schon seit einigen Jahren mit Georgia Tech [5], um gemeinsam Täuschungsangriffe zu erforschen sowie Schwachstellen aufzudecken und zu schließen. Das kalifornische Unternehmen hat 2017 das auf Fahrerassistenz- und Erkennungssysteme spezialisierte israelische Unternehmen Mobileye gekauft [6] und sich auch [„https://www.heise.de/meldung/Autonome-Autos-Intel-uebernimmt-Mobileye-fuer-14-Milliarden-Euro-3650798.html“](https://www.heise.de/meldung/Autonome-Autos-Intel-uebernimmt-Mobileye-fuer-14-Milliarden-Euro-3650798.html) über eine Partnerschaft mit BMW [7] für den Wettlauf rund um selbstfahrende Autos in Stellung gebracht. ()

---

URL dieses Artikels:

<https://www.heise.de/-4701409>

Links in diesem Artikel:

[1] [https://www.heise.de/newsletter/manage/ho?wt\\_mc=nl.red.ho.daily.meldung.link.link](https://www.heise.de/newsletter/manage/ho?wt_mc=nl.red.ho.daily.meldung.link.link)

[2] <https://www.heise.de/select/tr/2019/10/1922815561768592809>

[3] <https://www.darpa.mil/news-events/2019-02-06>

[4] <https://www.protocol.com/intel-darpa-adversarial-ai-project>

[5] <https://newsroom.intel.com/news/intel-joins-georgia-tech-darpa-program-mitigate-machine-learning-deception-attacks/>

[6] <https://www.heise.de/meldung/Autonome-Autos-Intel-uebernimmt-Mobileye-fuer-14-Milliarden-Euro-3650798.html>

[7] <https://www.heise.de/meldung/BMW-will-ab-2021-selbstfahrende-Autos-bauen-3253086.html>

</small><br/><small>

<strong>[8]</strong>&#160;mailto:hob@ct.de

</small><br/></p> <p class=„printversion\_\_copyright“><em>Copyright &#169; 2020 Heise  
Medien</em></p> </html>

From:  
<https://schnipsl.qgelm.de/> - **Qgelm**

Permanent link:  
<https://schnipsl.qgelm.de/doku.php?id=wallabag:darpa-will-ki-gegen-tuschungsversuche-immun-machen>

Last update: **2021/12/06 15:24**

