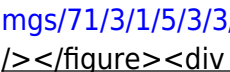


Wie synthetische Daten dem Datenschutz helfen

[Originalartikel](#)

[Backup](#)

```
<html> <header class=„article-header“><h1 class=„articleheading“>Wie synthetische Daten dem
Datenschutz helfen</h1><div class=„publish-info“> Eva Wolfangel</div></header><figure
class=„aufmacherbild“><img
src=„https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de/_imgs/
18/3/1/5/4/1/2/9/shutterstock_232738360-fd22c3a0c6d8e9c8.jpeg“
srcset=„https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de/_im
gs/18/3/1/5/4/1/2/9/shutterstock_232738360-fd22c3a0c6d8e9c8.jpeg 700w,
https://heise.cloudimg.io/width/1050/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de/_imgs/18/3/
1/5/4/1/2/9/shutterstock_232738360-fd22c3a0c6d8e9c8.jpeg 1050w,
https://heise.cloudimg.io/width/1500/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de/_imgs/18/3/
1/5/4/1/2/9/shutterstock_232738360-fd22c3a0c6d8e9c8.jpeg 1500w,
https://heise.cloudimg.io/width/2056/q75.png-lossy-75.webp-lossy-75.foil1/_www-heise-de/_imgs/18/3/
1/5/4/1/2/9/shutterstock_232738360-fd22c3a0c6d8e9c8.jpeg 2056w“ alt=„“ class=„img-responsive“
referrerpolicy=„no-referrer“ /><figcaption class=„akwa-
caption“>(Bild: &#160;rvlsoft/Shutterstock.com)</figcaption></figure><p><strong>Synthetische
Daten k&#246;nnten auch eine Rolle dabei spielen, KI datenschutzfreundlicher zu trainieren. Dazu
bedienen sich Experten eines Tricks.</strong></p><p>Bekanntlich brauchen maschinelle
Lernsysteme sehr viele Daten, um zu lernen. Dabei hat sich aber immer wieder gezeigt, dass sich
selbst anonymisierte Datens&#228;tze de-anonymisieren lassen. Das hei&#223;t, es ist
m&#246;glich, in anonymisierten Datens&#228;tzen R&#252;ckschl&#252;sse auf Einzelne zu
ziehen, denn oft gibt es weitere Zusammenh&#228;nge, die k&#252;nstliche Intelligenz
findet.</p><p>„Synthetische Daten helfen uns, die DSGVO umzusetzen“, sagt Alexandra Ebert vom
Start-up MostlyAI. Nvidias KI-Gesichter erschienen ihr als Beweis, dass diese Methode helfen
k&#246;nnte, Datens&#228;tze zu anonymisieren &#8211; oder eben nicht im klassischen Sinn zu
anonymisieren, sondern auf deren Basis neue Datens&#228;tze mit den gleichen Eigenschaften zu
konstruieren, die keine R&#252;ckschl&#252;sse auf die urspr&#252;nglichen Daten zulie&#223;en
und damit nicht de-anonymisierbar seien.</p><p>„Bei Nvidia ist es nicht vorgekommen, dass von
einem Individuum auch nur ein Element &#252;bernommen wurde“, sagt Ebert, „der Algorithmus hat
die Regeln verstanden, wie menschliche Gesichter aussehen.“</p><p>Ein Kunde von MostlyAI ist die
&#196;rtebank, und f&#252;r Banken sind Daten interessant, die gleichzeitig aus
Datenschutzgr&#252;nden sehr heikel sind: Der Zusammenhang zwischen Verm&#246;gen,
Lebensstil und Verhalten. Wer mehr &#252;ber diese Zusammenh&#228;nge wei&#223;, kann seine
Produkte besser an die Kunden anpassen. „Diese Daten d&#252;rften aber nicht verwendet werden“,
sagt Ebert.</p><p>Als Notl&#246;sung nutzen Entwickler bisher entweder ihre eigenen Daten oder
arbeiteten mit sogenannten „Personas“, groben Ann&#228;herungen an verschiedene Typen von
Kunden auf der Basis von Hypothesen. Doch diese &#252;bereinfachten Personen halfen nicht
wirklich weiter, weil sie viel zu pauschal waren. Damit lie&#223;en sich die Bed&#252;rfnisse
einzelner Kunden nicht genau genug abbilden.</p><header class=„a-boxheader“ data-collapse-
trigger=„“>TR 6/2021</header><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/2484/q50.png-lossy-50.webp-lossy-50.foil1/_www-heise-de/_imgs
/71/3/1/5/3/3/2/4/tr0621-f9e91da1b649c0fe.png“
srcset=„https://heise.cloudimg.io/width/4968/q30.png-lossy-30.webp-lossy-30.foil1/_www-heise-de/_i
```

 2x" class="c1" referrerpolicy="no-referrer" /></figure><div class="a-inline-textboxcontent-container"><p class="a-inline-textboxsynopsis">Dieser Beitrag stammt aus Ausgabe 6/2021 der Technology Review. Das Heft ist ab dem 19.08.2021 im Handel sowie direkt im heise shop erhältlich. Highlights aus dem Heft:</p><ul class="a-inline-textboxlist"><li class="a-inline-textboxitem">Hey KI: Wie Forscher Künstliche Intelligenz gerecht machen wollen [1]<li class="a-inline-textboxitem">Bakterien: Die nächste Seuche ist schon da [2]<li class="a-inline-textboxitem">Wachmacher: Mehr Daten für besseren Kaffee [3]<li class="a-inline-textboxitem">KI-Texterstellung: Zwietracht-Maschine GPT-3 [4]<li class="a-inline-textboxitem">Mangelware Chips: Flaschenhals des Fortschritts [5]<li class="a-inline-textboxitem">Fokus: Smart Cities [6]<li class="a-inline-textboxitem">TR bestellen [7]</div></div><p>MostlyAI hat ein System des maschinellen Lernens entwickelt, das aus den Kundendaten der Ärtebank lernt, eine eigene Datenbank aus komplett künstlichen Daten zu erstellen – die aber die gleichen Eigenschaften haben. „Sie sind komplett repräsentativ und granular“, schwärmt Ebert. Die Software sei direkt bei der Bank installiert und lerne dort aus deren Kundendaten. Das Start-up bekomme die Daten nicht zu sehen. Am Ende stünde ein komplett separater Datensatz ohne eine zurückverfolgbare Beziehung zwischen den echten und den synthetischen Menschen.</p><p>„Bei solchen Behauptungen bin ich sehr skeptisch“, sagt Aaron Roth von der University of Pennsylvania: „Es ist noch nicht möglich, synthetische Datensätze mit strikten Datenschutzgarantien zu erstellen.“ So gebe es eine ganze Reihe an Forschungspapieren, die entsprechende Angriffe nachweisen, in denen sehr wohl Rückschlüsse gezogen werden konnte auf die Personen aus dem ursprünglichen Datensatz. Sogenannte „privacy attacks“ werden auch dadurch geförderet, dass KI-Systeme in manchen Fällen dazu neigen, auswendig zu lernen – und damit im schlimmsten Fall eine Kopie des Datensatzes erstellen.</p><p>Generell ist Roth aber optimistisch, dass synthetische Daten helfen können, KI datenschutzfreundlich lernen zu lassen. „Prinzipiell sollte es möglich sein, synthetische Datensätze zu erstellen, von denen man die gleiche Antwort bekommt wie von einem realen Datensatz.“ Es gebe vielversprechende Forschungen dazu – aber eben auch bekannte Rückschläge.</p><figure class="branding">

768743d2733393170782725323076696577426f783d27302532303025323036393625323033393127
25334525334372656374253230783d273027253230793d27302725323077696474683d27363936272
532306865696768743d273339312725323066696c6c3d27253233663266326632272533452533432f
726563742533452533432f737667253345/" class="c2" width="1200" referrerpolicy="no-referrer"
> [8]</figure><p>()
</p><hr /><p>URL dieses Artikels:
<small>

<https://www.heise.de/-6166906>

</small></p><p>Links in diesem Artikel:
<small>

[1] https://www.heise.de/select/tr/2021/6/212011200069
0184712

</small>
<small>

[2] https://www.heise.de/select/tr/2021/6/212011311555
6140178

</small>
<small>

[3] https://www.heise.de/select/tr/2021/6/212100709433
3071493

</small>
<small>

[4] https://www.heise.de/select/tr/2021/6/212011206087
2995354

</small>
<small>

[5] https://www.heise.de/select/tr/2021/6/212011304319
7977131

</small>
<small>

[6] https://www.heise.de/select/tr/2021/6/2120112345330889874

</small>
<small>

[7] https://shop.heise.de/technology-review-06-2021/Print?wt_mc=intern.shop.shop.tr_2106.dos.textlink.textlink

</small>
<small>

[8] https://www.heise.de/tr/

</small>
<small>

[9] mailto:bsc@heise.de

</small>
</p><p class=„printversion__copyright“>Copyright © 2021 Heise Medien</p> </html>

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

Permanent link:
<https://schnipsl.qgelm.de/doku.php?id=wallabag:wb2wie-synthetische-daten-dem-datenschutz-helfen>

Last update: 2025/06/27 11:17

