

Dr. Nicolas Müller

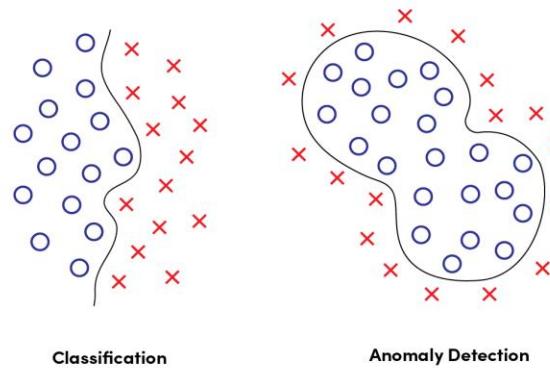
Audio- und Video-Deepfakes – Erzeugen, Erkennen, Verstehen

Bucerius Law School
09-10.01.2026

Cognitive Security Technologies (CST)

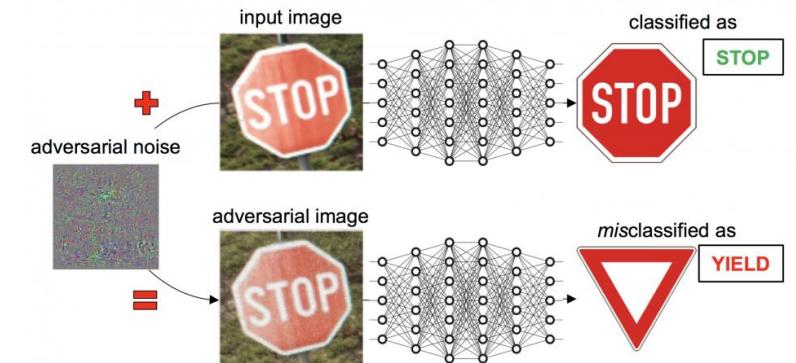
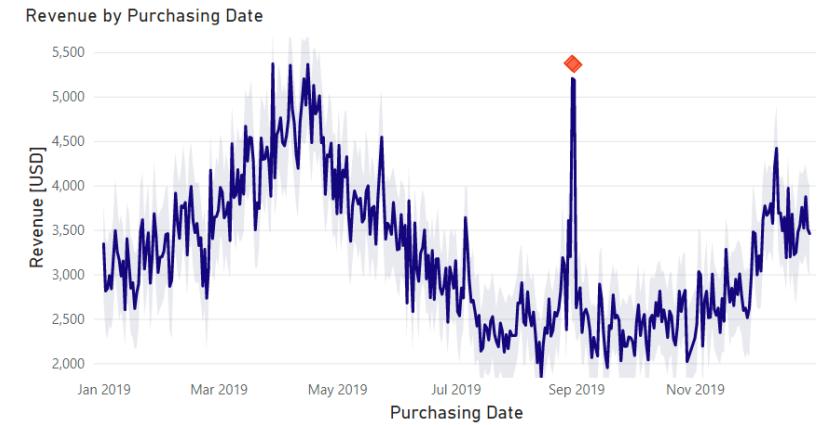
Fraunhofer:

- Auftragsforschung
- Prototypenbau
- Forschungsprojekte (EU/Bund)



Abteilung CST

- Anomalieerkennung
- Adversarial Machine Learning
- Large Language Models
- Deepfakes



Cognitive Security Technologies (CST)

Theory:

- These slides

Practical: lets build DFD

- Students: Data collection via resemble TTS and running models from HF locally
- Students: Setup code for DFD via Google Colabs. Maybe Nicolas create MLAAD_mini (10%) s.t. MLAAD + MAILABS usable
- Pentest resemble DFD
- Pretest deepfake-total

Übersicht: Deepfakes @ Fraunhofer AISEC

1) Forschung

- Grundlagenforschung
- Research Management: Reviewer und Organisator bei internationalen Konferenzen

2) Softwarelösungen

- Forschung in die Anwendung bringen: Deepfake-Total.com
- Awareness schaffen: Interaktive Demonstratoren zur "Schulung des Auges / Ohrs"

3) Öffentlichkeitsarbeit

- Handelsblatt, Financial Times, ZDF, BR3-Nano, Staatsministerium BaWü, NTV, Tagesspiegel...
- Konferenzbesuche, Vorträge, Seminare, Schulungen (Industrie und TU-München)

Does Audio Deepfake Detection Generalize?

Nicolas M. Müller¹, Pavel Czempin², Franziska Dieckmann²,
Adam Froglyar³, Konstantin Böttiger¹

¹Fraunhofer AISEC ²Technical University Munich ³why do birds GmbH
nicolas.mueller@aisec.fraunhofer.de

Abstract

Current text-to-speech algorithms produce realistic fakes of human voices, making deepfake detection a much-needed area of research. While researchers have presented various deep learning models for audio spoof detection, it is often unclear exactly why these architectures are successful: Preprocessing steps, hyperparameter settings, and the degree of fine-tuning are not consistent across related work. Which factors contribute to success, and which are accidental?

In this work, we address this problem: We systematize audio spoofing detection by re-implementing and uniformly evaluating twelve architectures from related work. We identify overarching features for successful audio deepfake detection such

features professional speakers and has been recorded in a studio environment, using a semi-anechoic chamber. What can we expect from audio spoof detection trained on this dataset? Is it capable of detecting realistic, unseen, ‘in-the-wild’ audio spoofs like those encountered on social media?

To answer these questions, this paper presents the following contributions:

- We reimplement twelve of the most popular architectures from related work and evaluate them according to a common standard. We systematically exchange components to attribute performance reported in related work to either model architecture, feature extraction, or data preprocessing techniques. In this way, we identify fun-



Agenda des heutigen Vortrags

- 1) Einleitende Beispiele
- 2) Deepfake Erstellung
 - Wie einfach ist die Erstellung von Deepfakes?
 - Die Technik im Detail
- 3) Was tun?
 - Medienkompetenz
 - KI-gestützte Deepfake Erkennung
 - Content Credentials
- 4) Deepfake-Technologie für den guten Zweck
- 5) Fazit



Deutschland braucht Neuwahlen

Jetzt und sofort!

Achtung: Künstliche Inkompetenz

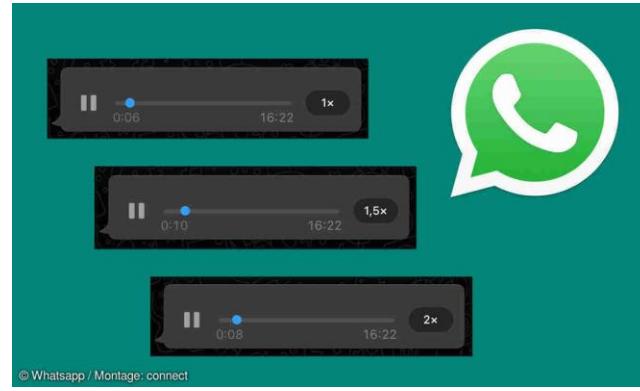


<https://www.youtube.com/watch?v=oxXpB9pSET0>



<https://www.youtube.com/watch?v=F4G6GNFz0O8>





© Whatsapp / Montage: connect

Forbes

FORBES > INNOVATION > CONSUMER TECH

A Voice Deepfake Was Used To Scam A CEO Out Of \$243,000

BIZTECH NEWS

Audio deepfake scams: Criminals are using AI to sound like family and people are falling for it

Betrug,
Enkelkindertrick 2.0

Telefon-Betrug

KI-Betrüger kopieren Stimme von Polizeichefin

Das Opfer war die Mutter der Top-Beamtin



Foto: Getty Images



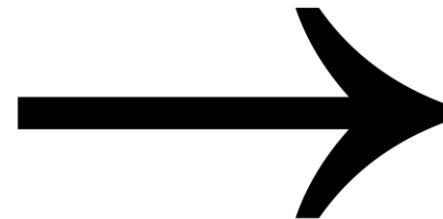
Robin
Mühlbach

04.01.2026 - 09:20 Uhr

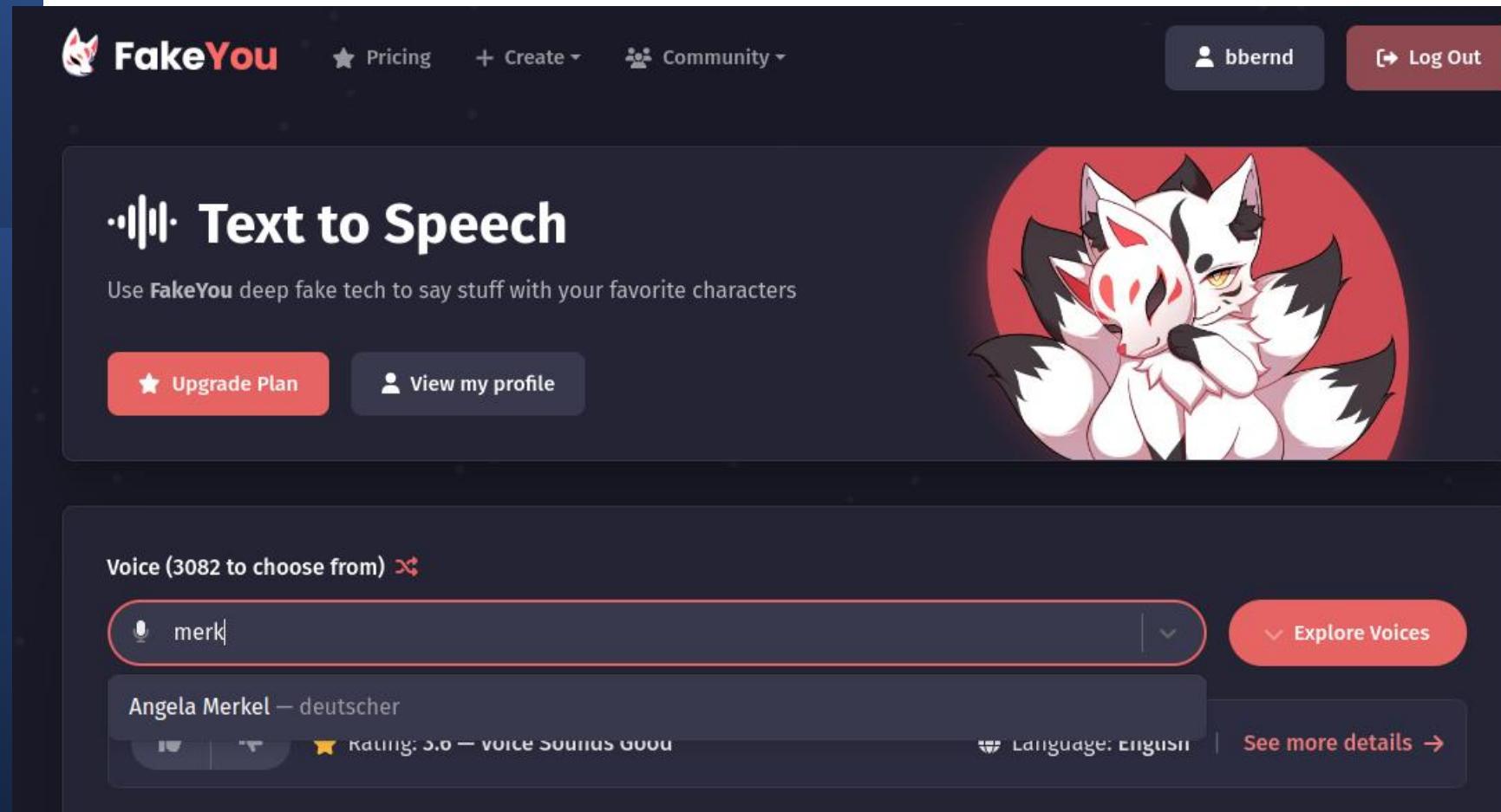
Die Stimme einer Polizeichefin aus Baden-Württemberg wurde mit KI-Technologie geklont, um einen perfiden Betrug zu begehen.



Immer einfacher zu erstellen



Voice-Cloning "On Demand"



The image shows a screenshot of the FakeYou website. At the top, there is a navigation bar with the logo "FakeYou", "Pricing", "Create", "Community", and user account information "bbernd" and "Log Out". Below the navigation bar, the main heading is "Text to Speech" with the sub-instruction "Use **FakeYou** deep fake tech to say stuff with your favorite characters". There are two buttons: "Upgrade Plan" (red) and "View my profile" (dark grey). To the right of the text area is a large, stylized illustration of two foxes (one white, one black) hugging. Below the main heading, there is a search bar with the placeholder "Voice (3082 to choose from)" and a microphone icon. The search term "merk" is typed into the bar. To the right of the search bar is a red button labeled "Explore Voices". Below the search bar, a specific voice entry is shown: "Angela Merkel – deutscher". This entry includes a small portrait of her, a rating of "Rating: 5.0 – voice sounds good", and a language indicator "Language: English". There is also a link "See more details →".

VoiceLab

Your creative AI toolkit. Design entirely new synthetic voices from scratch. Clone your own voice or a voice you have a permission and rights to. Only you have access to the voices you create.



Voice Design

Design entirely new voices by adjusting their parameters. Every voice you create is randomly generated and is entirely unique even if the same settings are applied.



Instant Voice Cloning

Clone a voice from a clean sample recording. Samples should contain 1 speaker and be over 1 minute long and not contain background noise.

Kommerzielle Anbieter

||Eleven
Labs

60

Click to upload a file or drag and drop
Audio files, up to 10MB each

Samples 3 / 25

Samples to Upload (3)

- recording3.mp3 7.4 MiB ▶ trash
- recording2.mp3 7.4 MiB ▶ trash
- recording1.mp3 7.4 MiB ▶ trash

i Sample quality is more important than quantity. Noisy samples may give bad results. Providing more than 5 minutes of audio in total brings little improvement.

Labels 2 / 5

accent : British trash source : Reading trash +

Description

Me dramatically narrating a book. Taking long pauses for additional effect.

I hereby confirm that I have all necessary rights or consents to upload and clone these voice samples and that I will not use the platform-generated content for any illegal, fraudulent, or harmful purpose. I reaffirm my obligation to abide by ElevenLabs' [Terms of Service](#) and [Privacy Policy](#).

Cancel Add Voice

Version 1



Version 2



Hands-on: Resemble.ai TTS

The screenshot shows the Resemble.ai platform interface. On the left, there's a sidebar with navigation links: MANAGEMENT (Projects), CONVERSATIONAL AI (Agents, Knowledge Base, Phone Numbers, Call Logs). The main area displays a conversation window. At the top of the window, it says "Model: Chatterbox" and "Ember" (with a microphone icon) "or [create a voice](#)". To the right is a "Voice Settings" button. The conversation text is: "Hi there! Welcome to the customer help desk! Let me briefly look up your account. It looks like you have a balance of \$25.47. I am more than happy to help you with this! Would you like to continue with your refund or, uh, is there anything else I can do for you?". Below the text is a circular button with a plus sign (+) and a file icon, and a status bar showing "262 / 2000".

Funktionsweise

Mustererkennung:
Text ~ Sprache

Herzliche
Willkommen

Schön Sie zu sehen



...



Herausforderungen

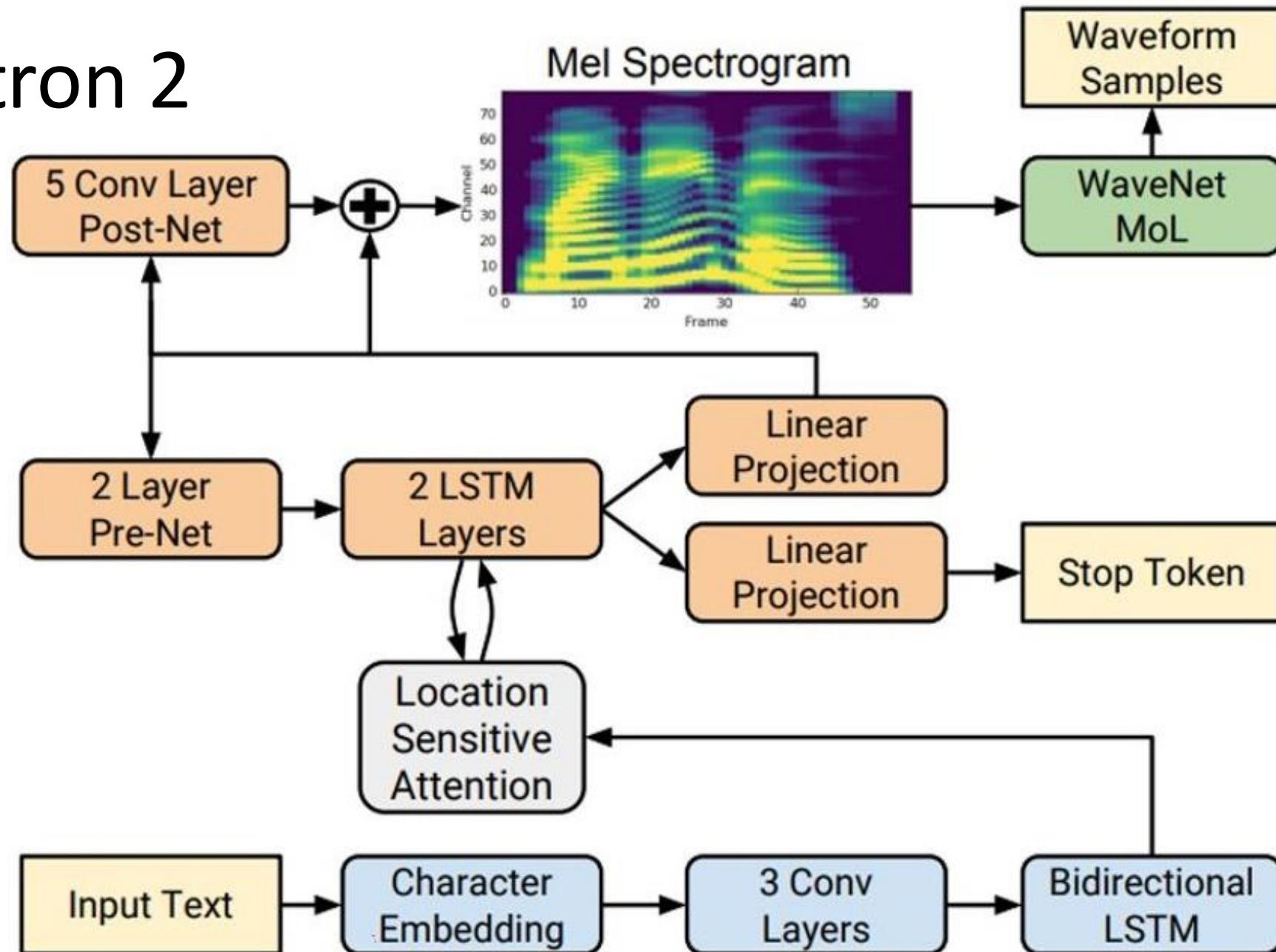
Information im Text:

- Was wird gesagt (Inhalt)

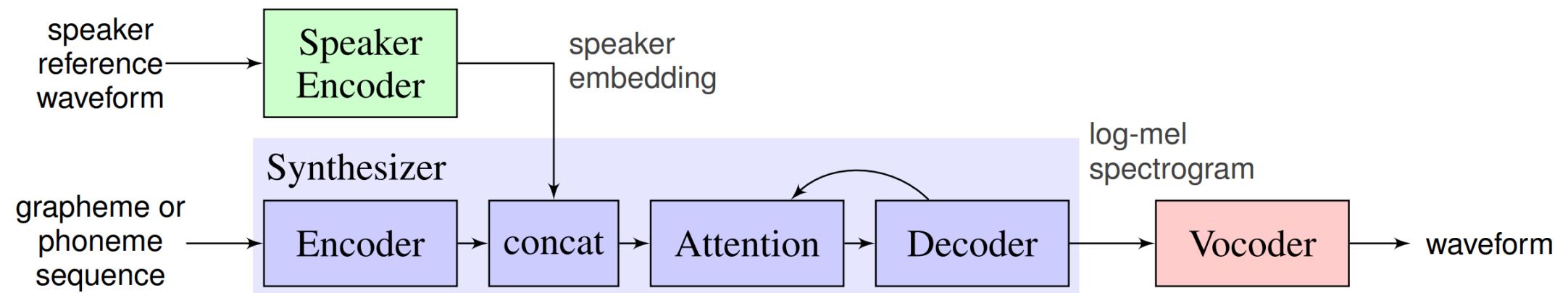
Information in der Stimme:

- Was wird gesagt (Inhalt)
- Wer sagt es (Sprecher)
- Wie wird es gesagt (Betonung)
- Wo wird es gesagt (Kanalinformation)

Tacotron 2

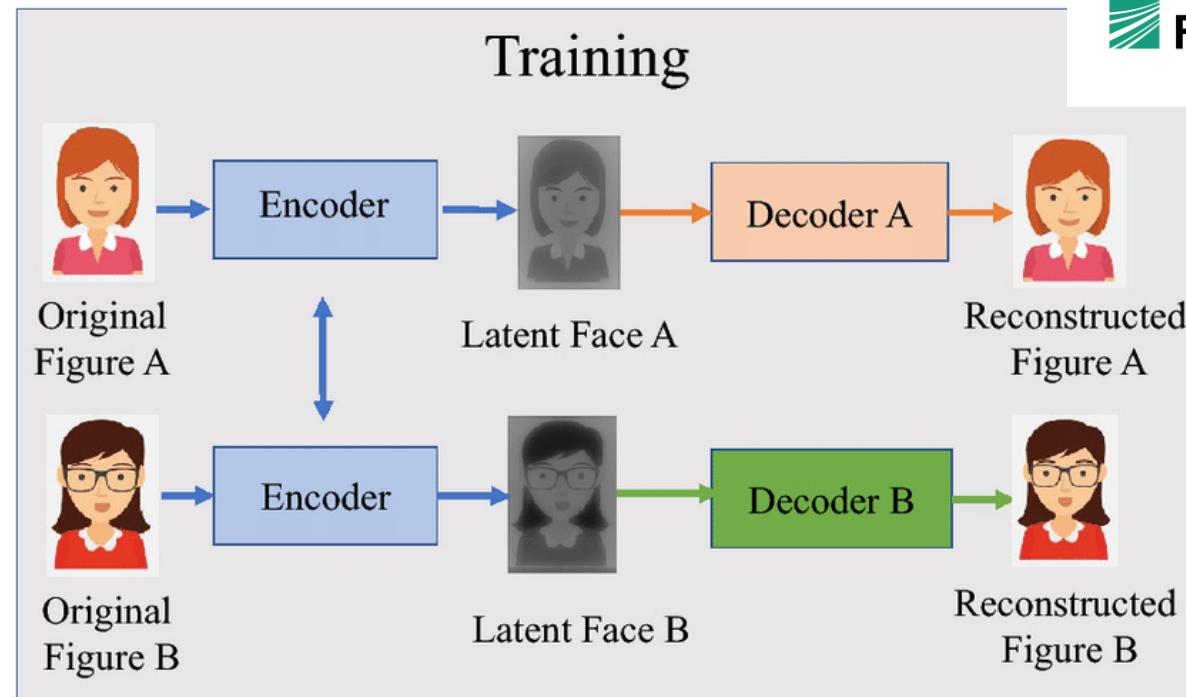
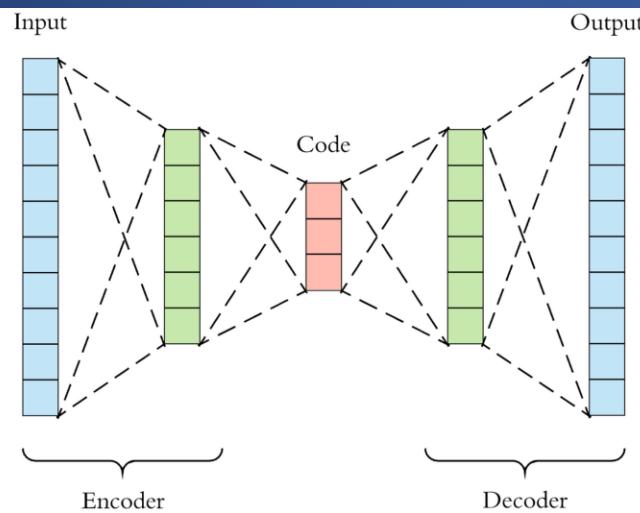


Multispeaker TTS

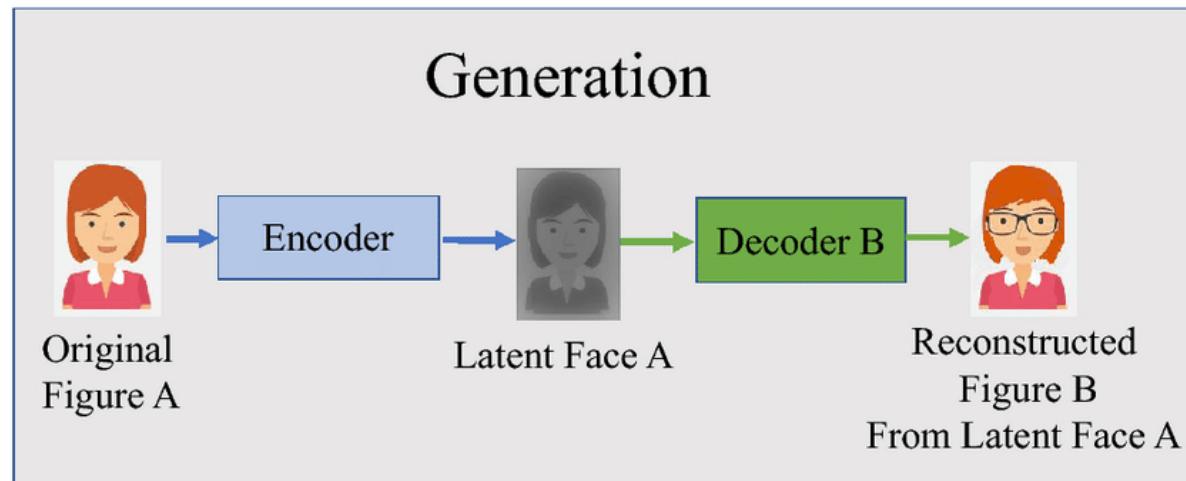


<https://arxiv.org/abs/1806.04558>

Faceswap



(a) Training Phase



(b) Generation Phase



Wie reagieren?

- 1) Medienkompetenz
- 2) Deepfake Erkennung
- 3) Signaturverfahren



Medienkompetenz (Media literacy)





ChatGPT 5.2



Google Nano Banana







ChatGPT 5.2



Google Nano Banana





Der muss üben!

https://deepfake-total.com/spot_the_deepfake/

Round 1

Is this audio file authentic or fake?

Fake! Authentic!

▶ 0:00 / 0:02

Rounds Played

Your Score (Human)

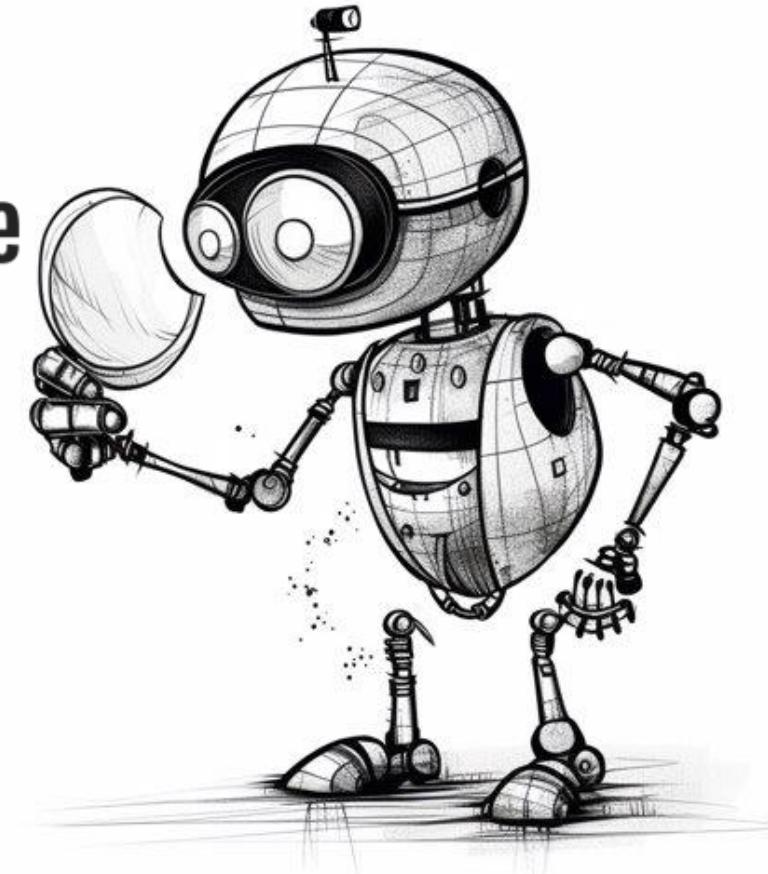
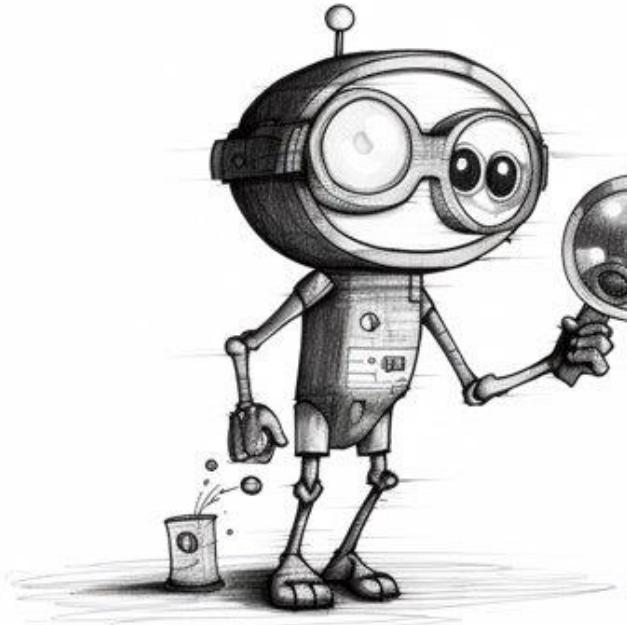
Deepfake Detection AI Score

1

1

0

KI-basierte Deepfake Erkennung



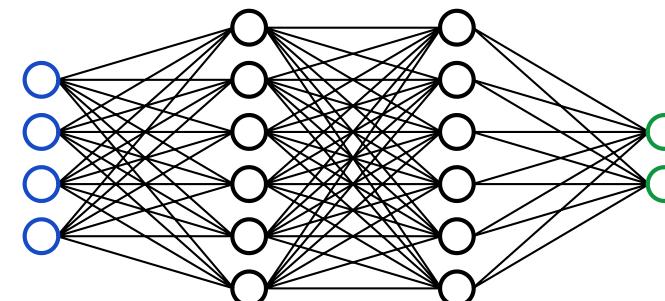
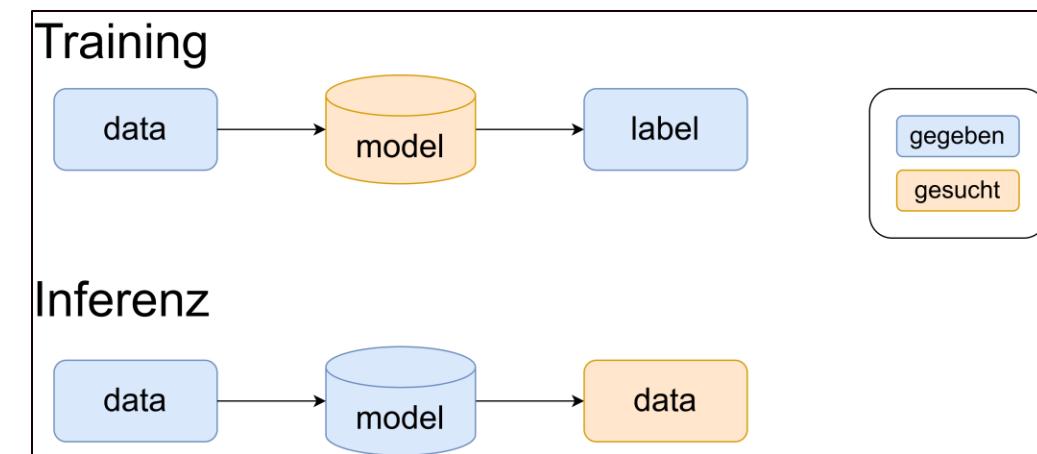
KI-driven detection of audio deepfakes

- Supervised Learning
 - Classification

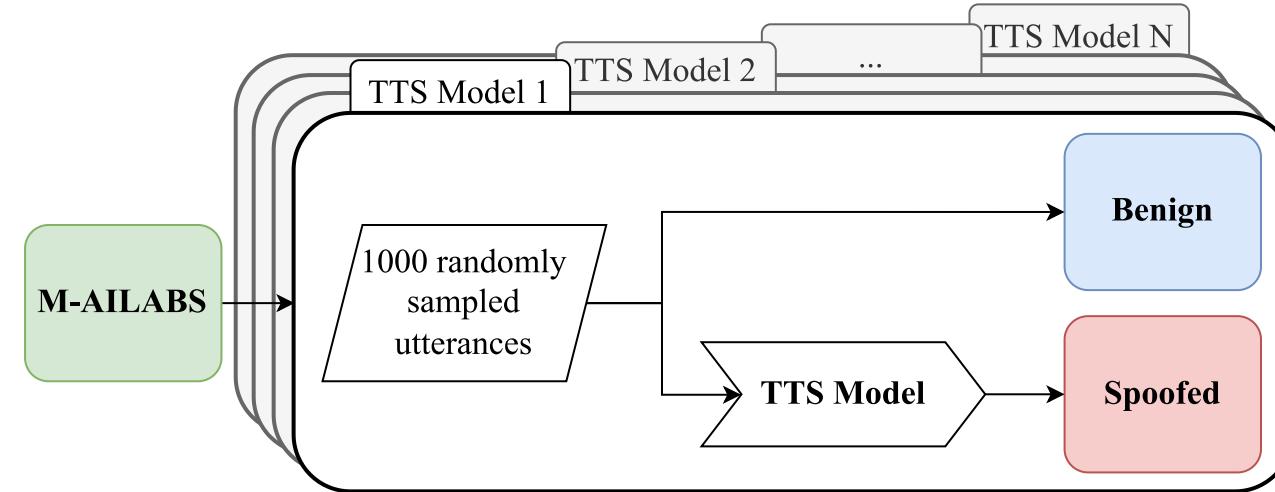
- Datensatz:
 - Real and Fake audio
 - Balanced
 - Large (10s of thousands minimum)

- Modelle:
 - Support Vector Maschine
 - KNN
 - ...
 - Neuronales Netzwerk

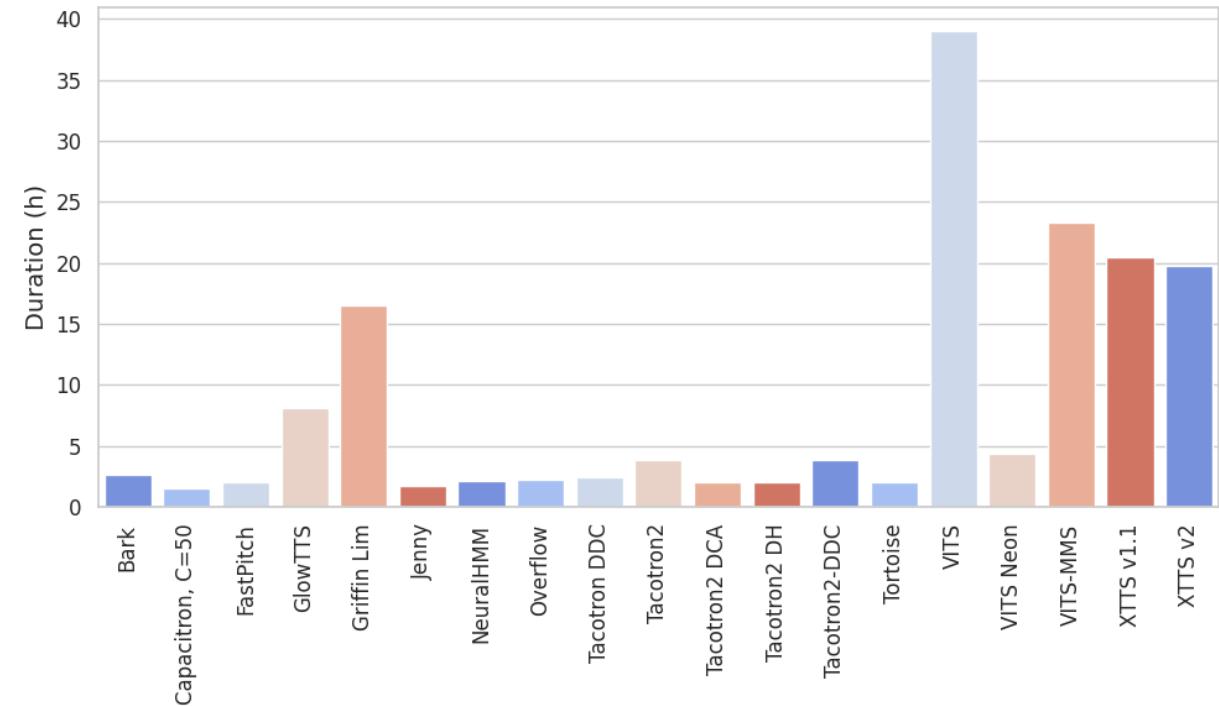
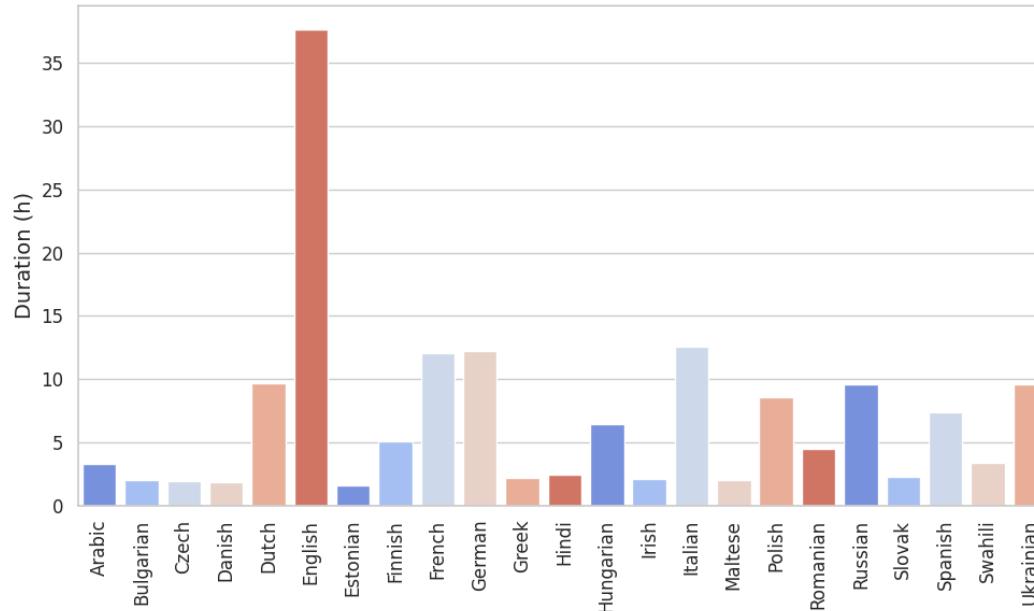
Data	Label
audio1.wav	real
audio2.wav	fake
audio3.wav	real
...	...



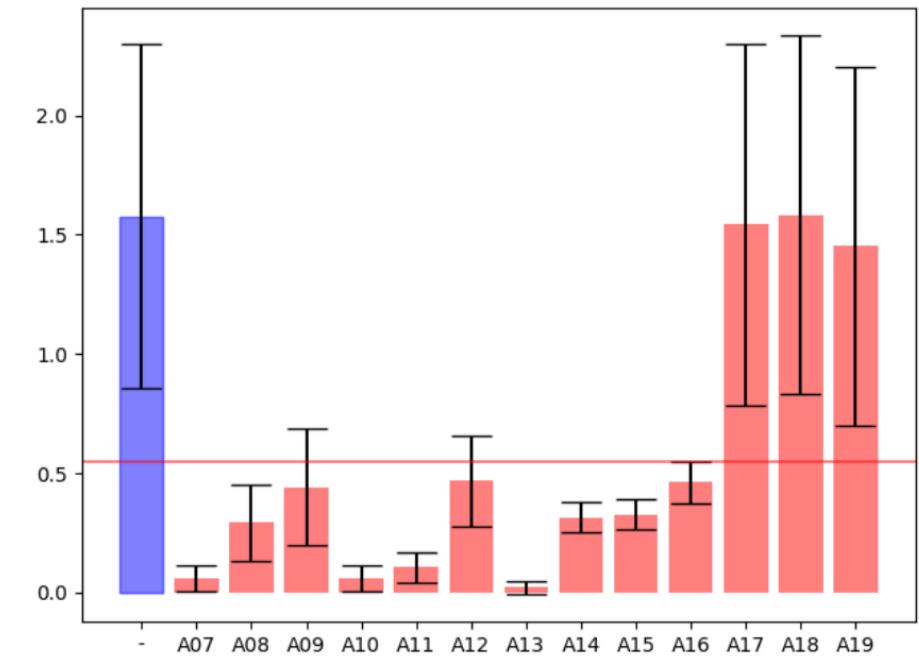
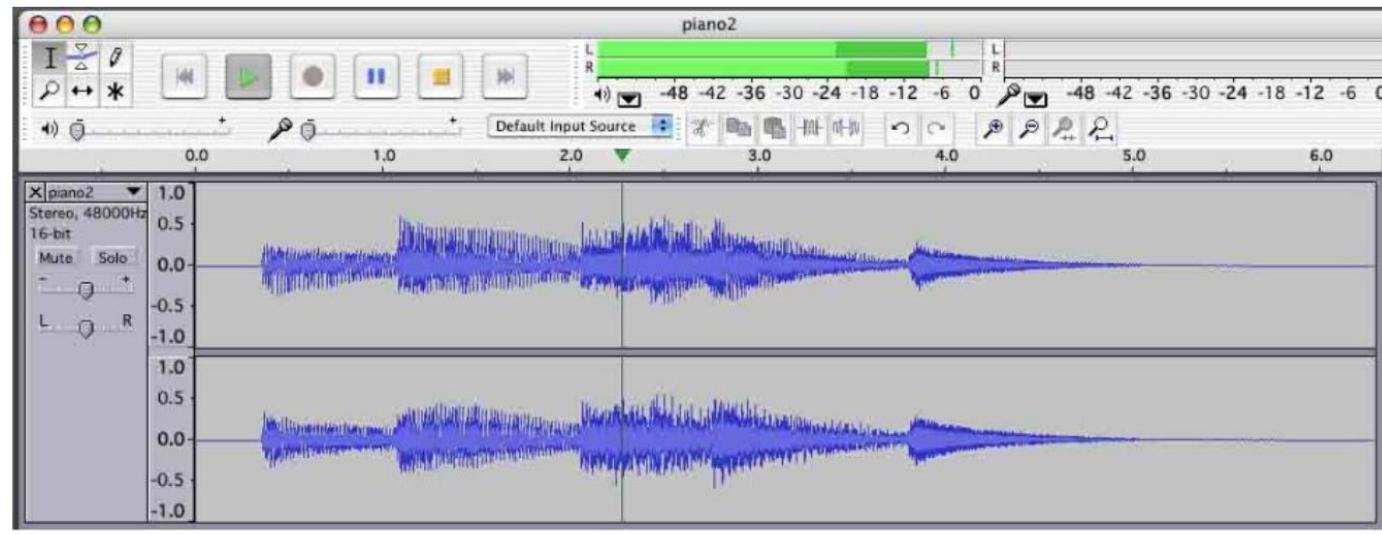
MLAAD: Multi Language Audio Antispoof Dataset



MLAAD: Multi Language Audio Antispoof Dataset



What do models learn?



Shortcut learning

Table 1: The results of training a dense neural network on only the *duration* of leading silence, compared against a random baseline. The network is able to significantly outperform the random baseline (see test EER, highlighted in blue). Results for the 2021 data (Deepfake and Logical Access) are copied verbatim from the submission platform (*progress* phase).

Model	Dev EER	Eval EER	LA21	DF21
FCNN	31.09 ± 3.8	15.12 ± 0.1	19.93	17.42
Random	50.0 ± 0.0	50.0 ± 0.0	-	-

<https://arxiv.org/pdf/2106.12914.pdf>

Model Name	Feature Type	Input Length	ASVspoof19 eval		In-the-Wild Data
			EER%	tDCF	EER%
LCNN	cqtspec	Full	6.354 ± 0.39	0.174 ± 0.03	65.559 ± 11.14
LCNN	cqtspec	4s	25.534 ± 0.10	0.512 ± 0.00	70.015 ± 4.74
LCNN	logspec	Full	7.537 ± 0.42	0.141 ± 0.02	72.515 ± 2.15
LCNN	logspec	4s	22.271 ± 2.36	0.377 ± 0.01	91.110 ± 2.17
LCNN	melspec	Full	15.093 ± 2.73	0.428 ± 0.05	70.311 ± 2.15
LCNN	melspec	4s	30.258 ± 3.38	0.503 ± 0.04	81.942 ± 3.50

<https://arxiv.org/pdf/2203.16263.pdf>

Erkennung von Audio Deepfakes



<http://deepfake-total.com/>



Analyze suspicious audio files to detect deepfakes, and automatically share them with the security community.

 [Youtube](#)

 [Twitter / X](#)

 [File Upload](#)



Enter a Youtube URL



Analyze



Home

Explore

Notifications

Messages

Lists

Bookmarks

Communities

Premium

Profile

More

Post



Nicolas Müller
@Nicolas15039314

← Post



ste is petitioning
@chai_stе

...

🌟 BREAKING: A behind the scenes corridor recording of Starmer about the Rochdale Azhar Ali crisis has been leaked.



Readers added context they thought people might want to know

Search

Relevant people



ste is petitioning
@chai_stе

Follow

🍉 marxist spiritual medium · daft
leftist content for daft people

Germany trends

1 · UEFA Champions League · Trending

#FCBLAZ

16.6K posts

2 · Automotive · Trending

#Tesla

23.8K posts

3 · Trending

#XRATIOAI

1,662 posts

4 · Technology · Trending

#instagramdown

346K posts

5 · Trending

\$zkhive

18.5K posts

6 · Trending

\$BOODEN



DEEPCODE TOTAL

Analyze suspicious audio files to detect deepfakes, and automatically share them with the security community.

 [Youtube](#)

 [Twitter / X](#)

 [File Upload](#)



https://twitter.com/chai_sté/status/1757717290865283282



Analyze

TWITTER_chai_ste_1757717290865283282.mp3

SSL-W2V2 Analysing seconds 0 to 30  Fake-O-Meter: 93.3%

Help us improve
Tell us if you think this audio is fake or authentic:

Authentic Don't know Fake



Home

Explore

Notifications

Messages

Lists

Bookmarks

Communities

Premium

Profile

More

Post



Nicolas Müller
@Nicolas15039314

← Post

Olaf Scholz reposted



Bundeskanzler Olaf Scholz ✅
@Bundeskanzler

...

Ohne Sicherheit ist alles andere nichts.

[Translate post](#)



Search

Fraunhofer
AISEC

Relevant people



Bundeskanzler Olaf Scholz ✅
@Bundeskanzler

Follow

Bundeskanzler der Bundesrepublik Deutschland. Europäer. Demokrat. MdB-Account: [@olafscholz](#) Federal Chancellor of Germany

Germany trends

1 · Trending

\$QORPO

5,673 posts

...

2 · Trending

\$PRINT

4,146 posts

...

3 · Trending

#Nockherberg

3,474 posts

...

4 · Trending

#NEDGER

3,426 posts

...

5 · Trending

\$linq

1,210 posts

...

DEEPCODE TOTAL

Analyze suspicious audio files to detect deepfakes, and automatically share them with the security community.

 Youtube

 Twitter / X

 File Upload



<https://twitter.com/Bundeskanzler/status/1761315033802224008>



Analyze

TWITTER_Bundeskanzler_1761315033802224008.mp3

SSL-W2V2 Analysing seconds 0 to 30  Fake-O-Meter: 22.4%

Help us improve
Tell us if you think this audio is fake or authentic:

Authentic Don't know Fake

[README](#)

Apache-2.0 license



LLaMA-Omni: Seamless Speech Interaction with Large Language Models

Authors: [Qingkai Fang](#), [Shoutao Guo](#), [Yan Zhou](#), [Zhengrui Ma](#), [Shaolei Zhang](#), [Yang Feng*](#)

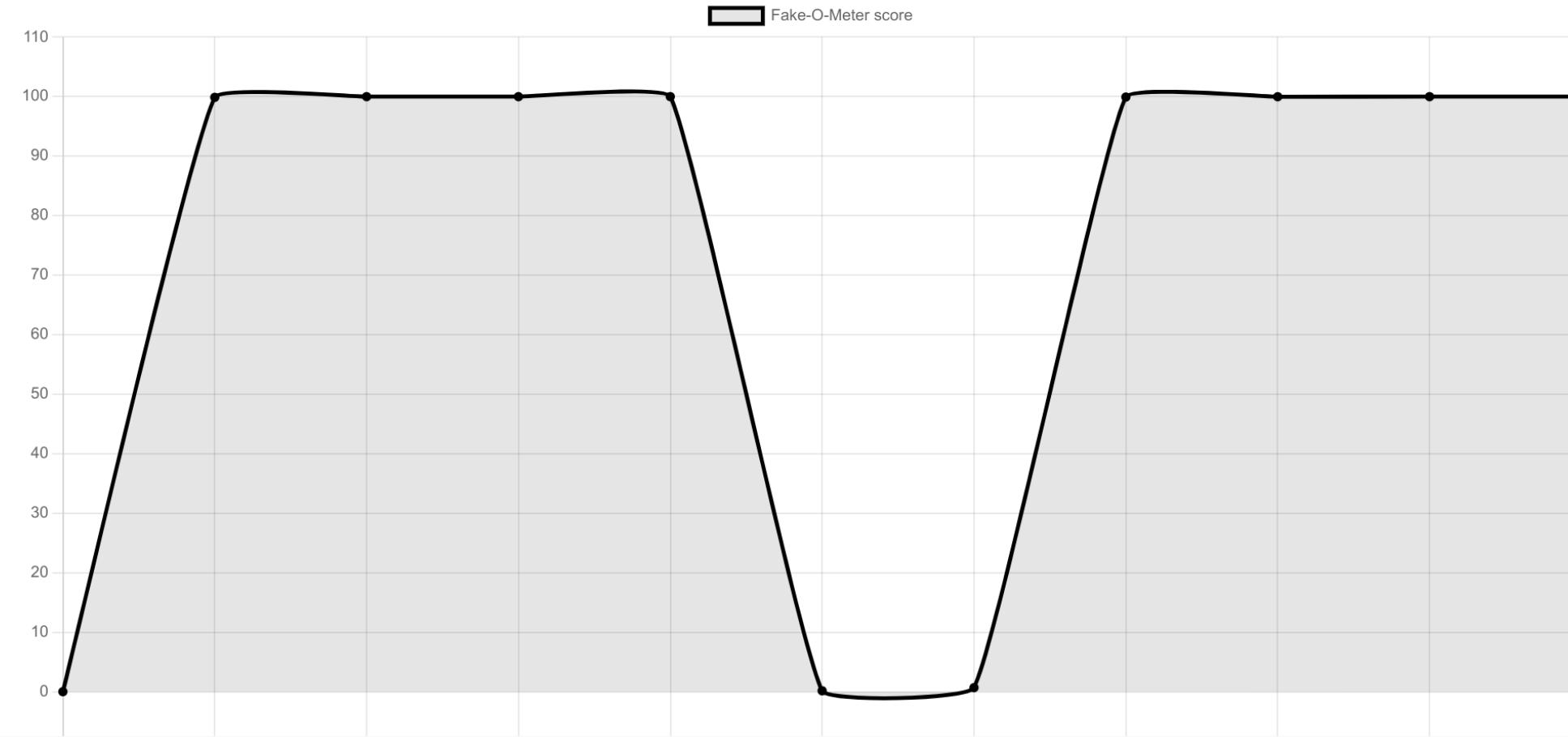
[arXiv 2409.06666](#)[Hugging Face Model](#)[Github Code](#)

[DeepfakeTotal](#)[Analysing seconds 0 to 90](#)

Fake-O-Meter: 72.8%

[More Info](#)

The following chart illustrates the Fake-O-Meter score across the audio over different time slices, each a length of 3.23 seconds. This allows for a more fine-grained analysis of the audio file. The X-axis (left to right) represents the time, and the Y-axis (bottom to top) represents the Fake-O-Meter score. The higher the value, the more likely it is that the audio at this specific time is AI-generated.

Analysis of the audio file over time

Examples

The following tables present a few examples of both synthetic audio (table 1) and authentic data (table 2). For each detection model (columns), the model score for the YouTube video (row) is shown via a color bar. Low scores (close to 0, green) indicate that the model considers the audio file authentic, while high scores (close to 1, yellow and red) indicate that the model considers the audio file a fake.

A perfect audio spoof detection model has high scores for the videos in the first table, visually represented by red bars. However, its scores would be low in the second table, i.e. predominantly green bars.

Table 1: Synthetic Audio

The following are videos where the audio track is created by an AI. These videos can be considered 'audio deepfakes'. Ideally, these scores are high (i.e. red color bar).

Youtube Video (click for details)	DeepfakeTotal
Jordan Peterson about the german government (DeepFake AI)	
Google Text to Speech With WaveNet AI	
TTS Forward-Tacotron + WaveRNN	
WORST Things Doctors Overlooked!	
Demonstration der Deepfake-Technologie im Kontext von Fake News	
What "unwritten rule" should everyone know and follow?	
Donald Trump Reads The Tragedy of Darth Plagueis The Wise #shorts	
Donald Trump Says Jeffrey Epstein Didn't Kill Himself Deepfake	
Two AIs talking to each other [Original]	

Erkennungsraten

- **TestA:** 97,9%
- **TestB:** 91,9%
- **TestC:** 76,1%
- **TestD:** 82,9%
- **TestE:** 96,8%
- **TestF:** 98,0%
- **TestG:** 89,9%



Google Synth ID

Füge Wasserzeichen zu KI-erstellten Inhalten hinzu:

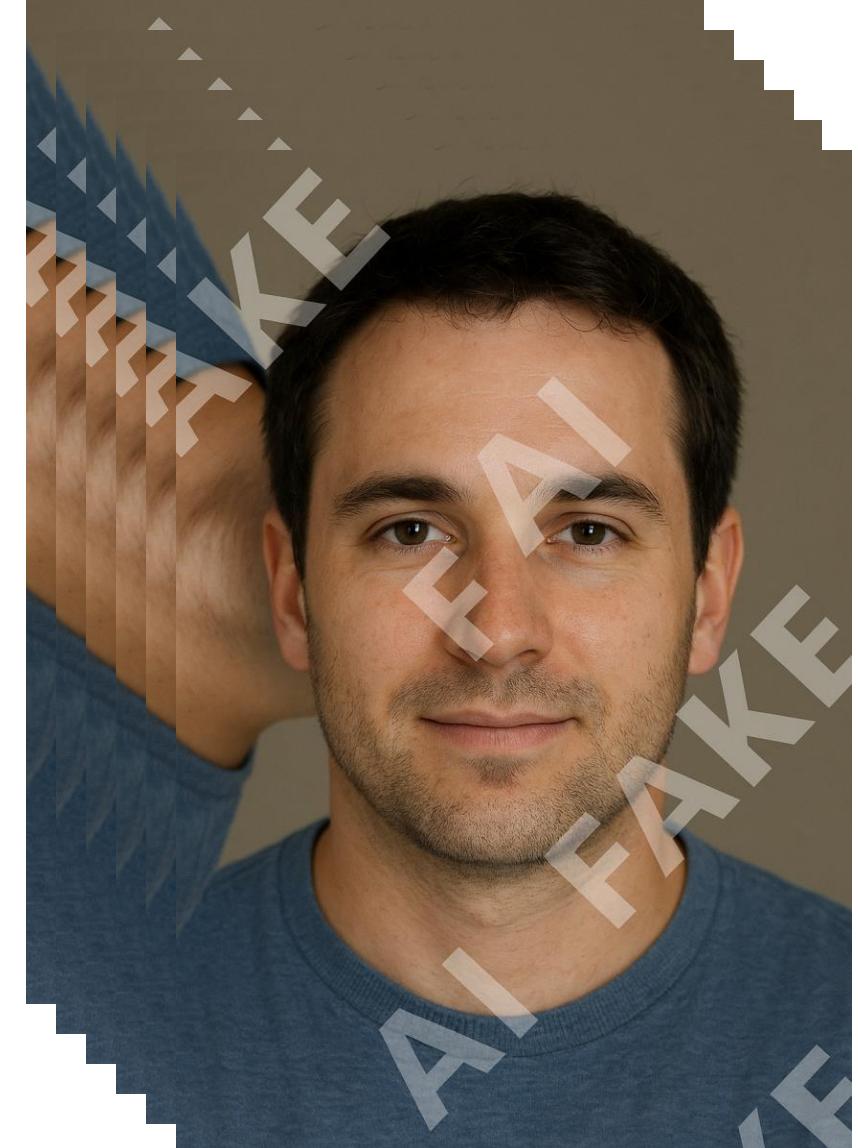
- **Bild:** Wasserzeichen wird ins Bild eingebettet
- **Video:** Wasserzeichen wird in jedes einzelne Frame eingebettet
- **Audio:** In Spektrogramm umwandeln, Wasserzeichen einbetten, zurückkonvertieren
- **Text:** Wahrscheinlichkeit für Wortwahlen verändertn. Wenn ein Text viele „bevorzugte“ Wörter enthält, gilt er als watermarked.

Wasserzeichen nur erkennbar wenn man weiß, wonach man sucht



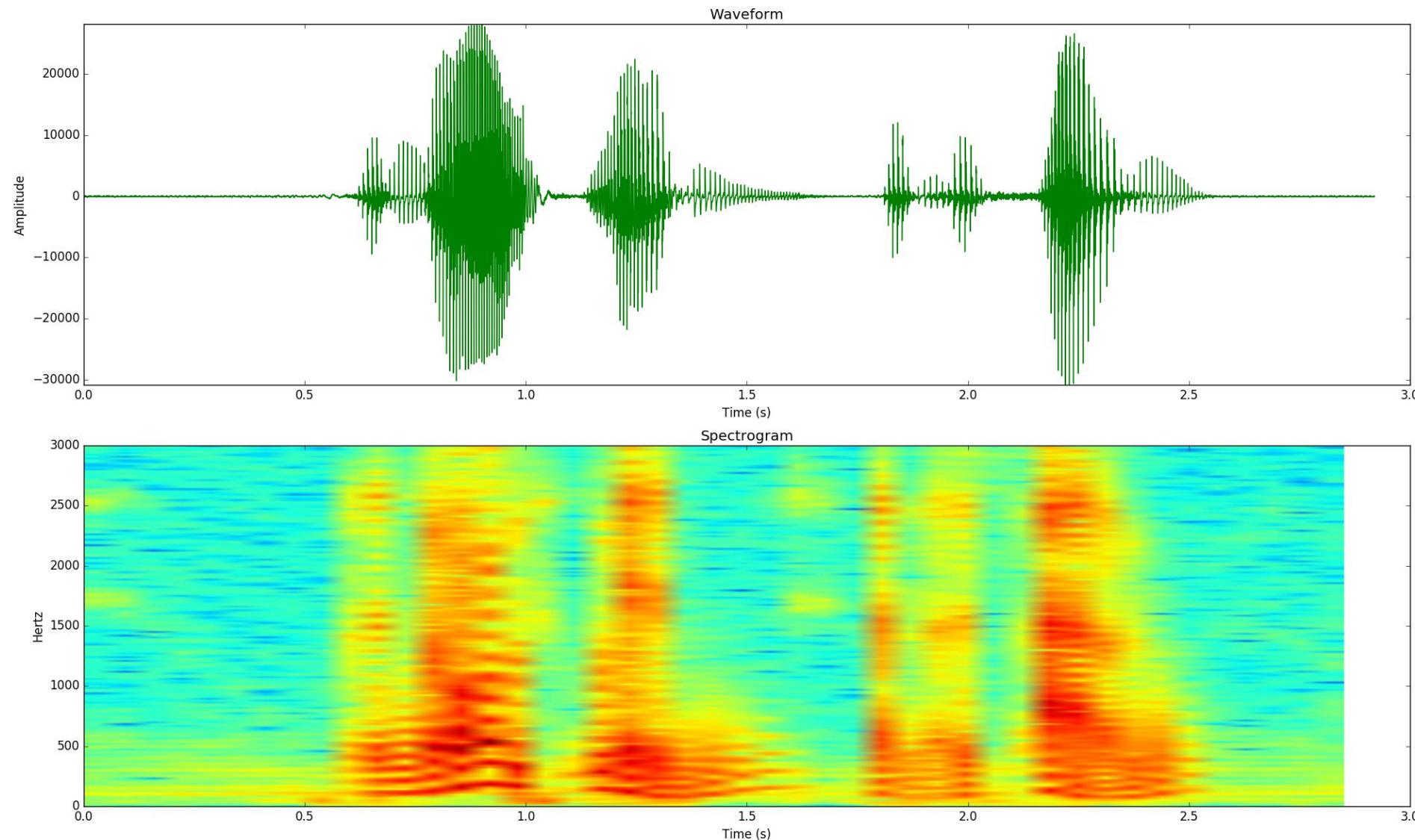
<https://deepmind.google/science/synthid/>

Watermarking Images and Video



<https://deepmind.google/science/synthid/>

Watermarking Audio



<https://wwwhomes.uni-bielefeld.de/gibbon/ShanghaiSummerSchool2018/>

Watermarking Text

I hope this email finds you well. I'm excited to share with you updates on the upcoming event.

We've just secured several great speakers who will be sharing their expertise and experiences. These speakers are leaders in their fields and have a wealth of knowledge to offer. In addition to the speakers, we also have other engaging activities such as interactive

[MORE VIDEOS](#)

SynthID: A tool for watermarking and identifying AI-generated content

<https://deepmind.google/science/synthid/>

Watermarking Text

oe this email finds you well. I'm excited to share with you
ates on the upcoming event.

incredible

great

spectacular

fantastic

we just secured several speakers who will b
ertise and experiences ers are leaders in th
e a wealth of knowledg ddition to the speak
have other engaging activities such as interactive works

MORE VIDEOS

Watermarking Text

be this email finds you\ incredible
ates on the upcoming d to share with you
we just secured several great speakers who will b
ertise and experiences. These speakers are leaders in th
e a wealth of knowledge to offer. In addition to the speak
have other engaging activities such as interactive works
t sessions and networking opportunities. These op
MORE VIDEOS

Watermarking Text

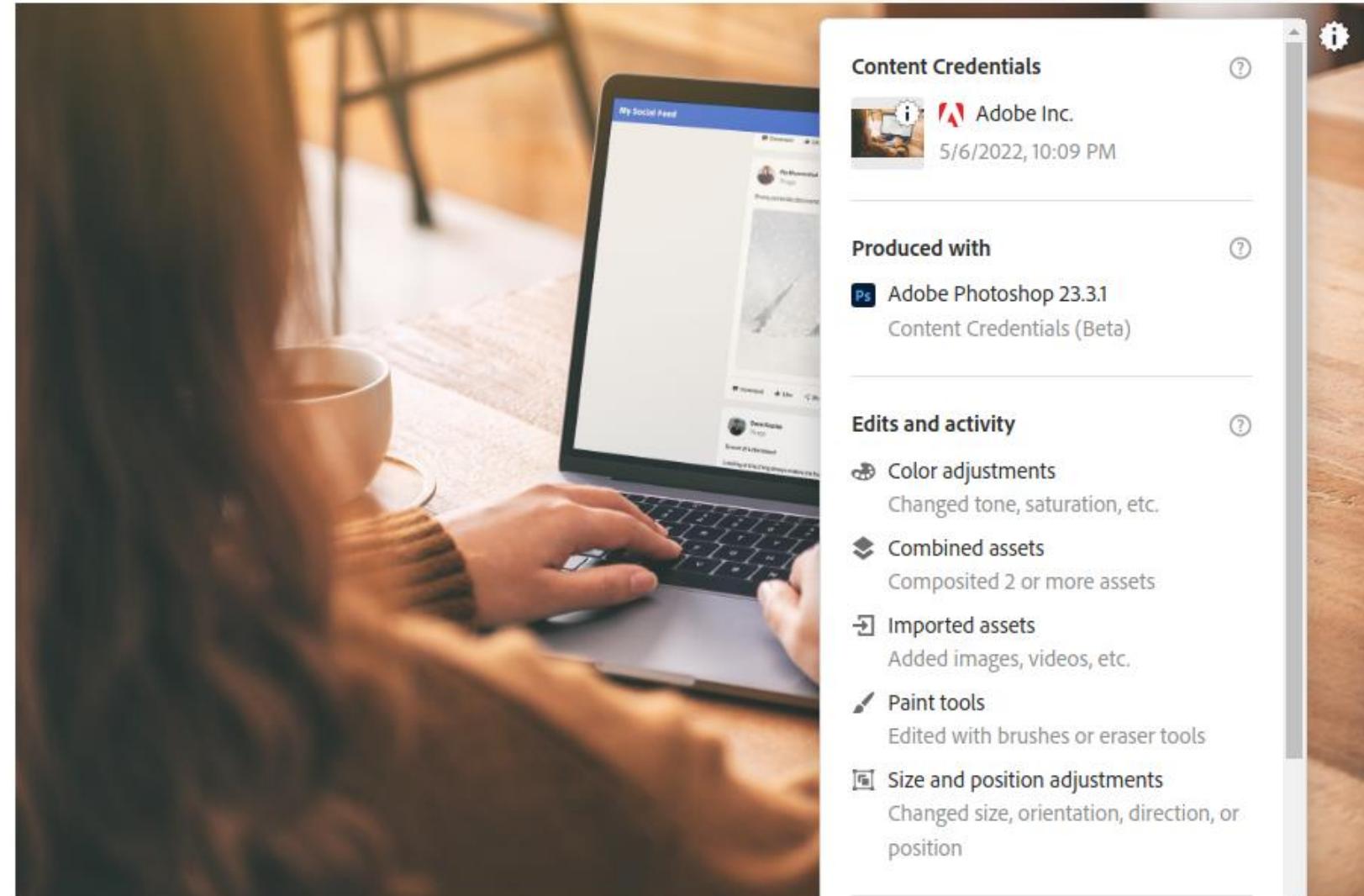
I hope this email finds you well. I'm excited to share with you some updates on the upcoming event.

We've just secured several incredible speakers who will be sharing their expertise and experiences. These speakers are leaders in their field and have a wealth of knowledge to offer. In addition to the speakers, we will also have other engaging activities such as interactive workshops, break-out sessions and networking opportunities. These activities will provide attendees with the opportunity to dive deeper into the topics, connect with peers, and build valuable relationships.

I'm confident this event will be a great success, and I'd love to have you as a speaker or workshop leader. I think your knowledge and experience would be a valuable addition to the event. If you're interested, please let me know your availability and we

Content Authenticity Initiative

- "Instead of guessing what's fake, let's prove what's true"
- Kryptographische Signatur
- "Herkunft", nicht "Wahrheit"



<https://contentauthenticity.org/>

Andreas Hoinisch · 2.

Präzision verkaufen, Innovation vorantreiben – Ihr Partner für sma...
14 Std. ·

+ Folgen

Egal wo ich hier lese:

Alle haben eine Rolex, einen Porsche und arbeiten wann immer sie wollen von wo sie wollen. ... mehr



Andreas Hoinisch ✅ · 2.

Präzision verkaufen, Innovation vorantreiben – Ihr Partner für sma...
14 Std. ·

+ Folgen

Egal wo ich hier lese:

Alle haben eine Rolex, einen Porsche und arbeiten wann immer sie wollen von wo sie wollen. ... mehr



Inhaltsnachweise

X

Für diese Medieninhalte sind Informationen zur Quelle oder zum Verlauf verfügbar. [Mehr erfahren](#)

- ◆ KI wurde verwendet, um Bild zu erstellen.
- 💻 Verwendete App oder verwendetes Gerät: ChatGPT
- ⌚ Anmeldedaten für Inhalte bereitgestellt von: OpenAI





Promoting transparency in AI

Adobe is committed to promoting transparency around content generated with AI tools like Adobe Firefly.

When downloading content generated with Firefly:

Content Credentials will be included

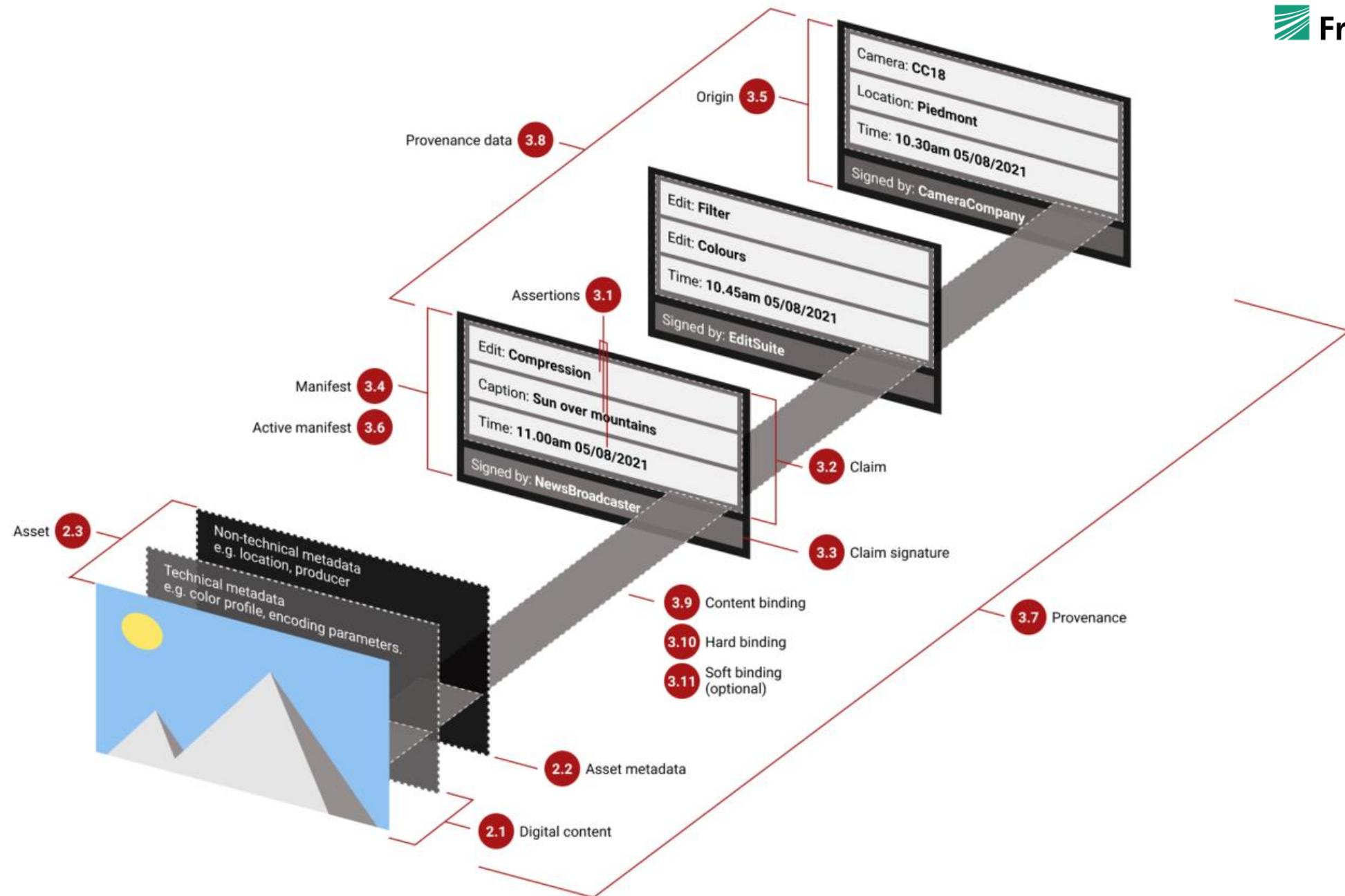


Adobe will include Content Credentials with all AI-generated content to let people know it was generated with AI. [Learn more about Content Credentials](#)

Don't show this again

Cancel

Continue

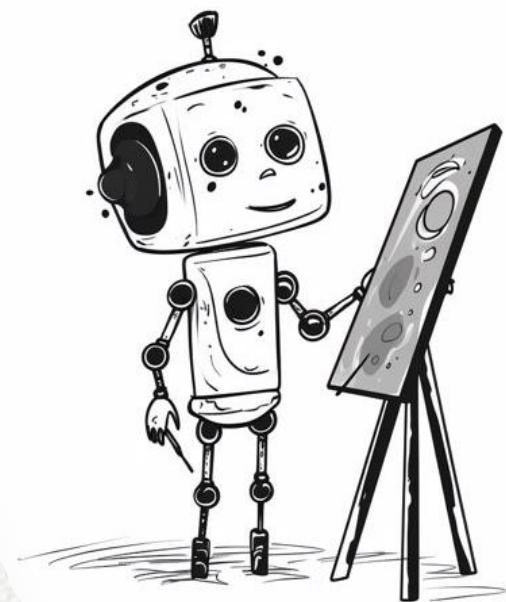




Ethischer Nutzen von KI

Be-me-voice (Google Parrotron)

Viele gutartige Use-Cases





Kontakt

Dr. Nicolas Müller

—

Cognitive Security Technologies

nicolas.mueller@aisec.fraunhofer.de

Ressourcen:

- deepfake-total.com/
- deepfake-demoaisec.fraunhofer.de/



Fraunhofer-Institut für Angewandte
und Integrierte Sicherheit AISEC