UNIVERSITÄT
LUZERN

PROREKTORAT LEHRE UND INTERNATIONALE BEZIEHUNGEN

ZENTRUM LEHRE

WEBINAR
ETHIK, DATENSCHUTZ & VERANTWORTUNG

BIAS, DEEPFAKES & ERINNERUNGSKULTUR

WEBINAR ETHIK: BIAS, DEEPFAKES & ERINNERUNGSKULTUR

Das Webinar beginnt um **12:15 Uhr**

Die Trainingsdaten beinhalten all unsere Vorurteile und prägen zunehmend unsere Weltsicht, unsere eingegeben Daten werden missbraucht und ausgegebene Daten können zu Deepfakes werden mit weitreichenden Folgen, von Diffamierung bis zu einer Veränderung der Erinnerungskultur.

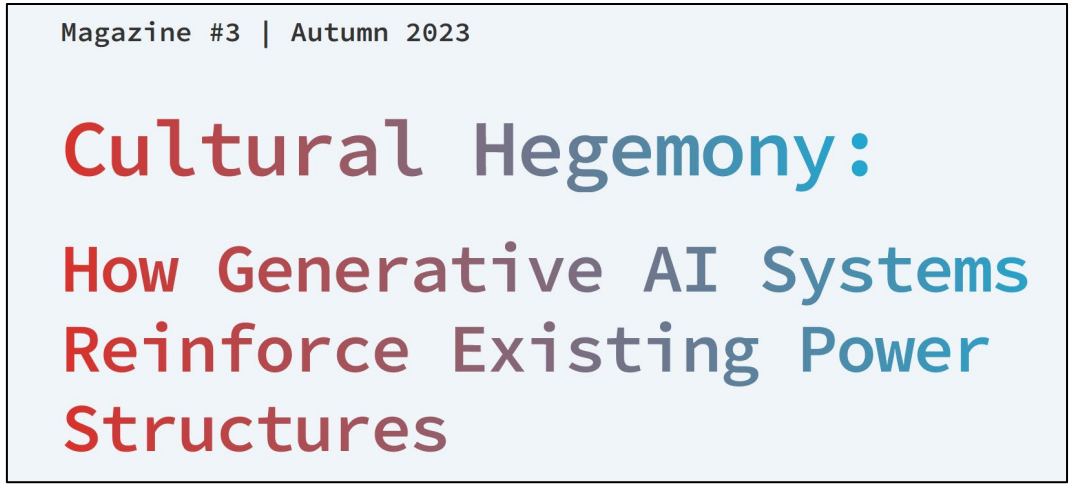
27. November 2025

UNIVERSITÄT
LUZERN

BIAS IN DEN TRAININGSDATEN

UNIVERSITÄT
LUZERN

BILDUNGSHEGEMONIE & DIGITALER KOLONIALISMUS?


Magazine #3 | Autumn 2023

Cultural Hegemony:

How Generative AI Systems Reinforce Existing Power Structures


Study shows pro-western cultural bias in the way AI decisions are explained

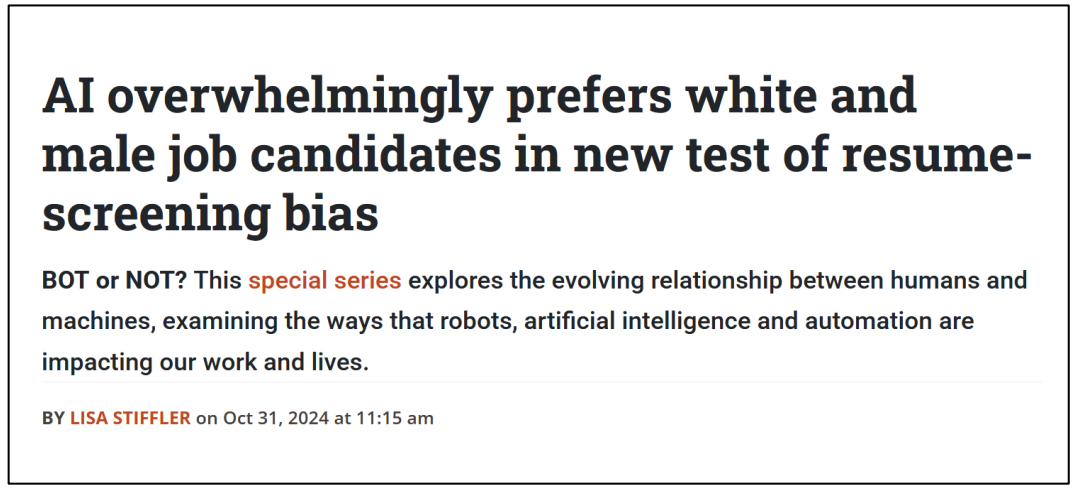
19 April 2024 - Mary Carman, Uwe Peters

Many existing explainable artificial intelligence systems produce explanations that are tailored to individualist, typically western, populations.


Addressing Western Bias in AI: A Call for Culturally Responsive Learning Materials

Christin Light ✨ ✨
Making learning sticky | Learning Experience Strategist | Behavioral science...

4. April 2024


AI overwhelmingly prefers white and male job candidates in new test of resume-screening bias

BOT or NOT? This special series explores the evolving relationship between humans and machines, examining the ways that robots, artificial intelligence and automation are impacting our work and lives.

BY LISA STIFFLER on Oct 31, 2024 at 11:15 am

Abstract

This study investigates the halo effect in AI-driven hiring evaluations using Large Language Models (LLMs) and Multimodal Large Language Models (MLLMs). Through experiments with hypothetical job applications, we examined how these models' evaluations are influenced by non-job-related information, including extracurricular activities and social media images. By analyzing models' responses to Likert-scale questions across different competency dimensions, we found that AI models exhibit significant halo effects, particularly in image-based evaluations, while text-based assessments showed more resistance to bias. The findings demonstrate that supplementary multimodal information can substantially influence AI hiring decisions, highlighting potential risks in AI-based recruitment systems.

The AI Influencer Illusion: How the Halo Effect Is Sabotaging Your AI Education

Aug 18, 2025 // Artificial Intelligence

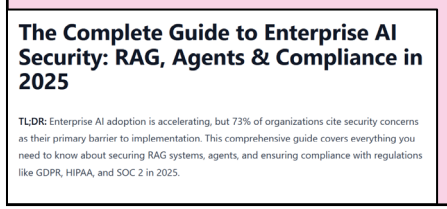
Most people teaching AI online learned everything they know from the same YouTube video you could watch in 15 minutes.

There, I said it. And before you dismiss this as cynicism, consider the psychology at play: you're probably trusting certain AI educators not because of their expertise, but because they're really good at looking like experts. This isn't necessarily their fault—it's how human psychology works in the age of social media.

The problem isn't that creators are intentionally misleading you. The problem is that a psychological phenomenon called the halo effect is quietly undermining your ability to distinguish between genuine AI expertise and polished content creation skills.

RAG

Retrieval Augmented Generation


The Complete Guide to Enterprise AI Security: RAG, Agents & Compliance in 2025

TL;DR: Enterprise AI adoption is accelerating, but 73% of organizations cite security concerns as their primary barrier to implementation. This comprehensive guide covers everything you need to know about securing RAG systems, agents, and ensuring compliance with regulations like GDPR, HIPAA, and SOC 2 in 2025.

GEGENMASSNAHMEN?

KI-spezifisch:

- RAG: Eigene Texte hochladen, auf die das LLM zugreifen soll
- Prompting mit Perspektiv-Vorgaben: „Nenne Quellen aus dem Globalen Süden / aus nicht-englischsprachigen Kontexten / von marginalisierten Autor:innen“

Immer gültig:

- Immer (auch) mit Originalquellen arbeiten
- Literaturlisten dekolonial reflektieren

HALO-EFFEKT

UNIVERSITÄT
LUZERN

KI-INDUZIERTE PSYCHOSEN: ONE-PERSON-ECHO-CHAMBER

Episode 253: “AI Psychosis”: Emerging Cases of Delusion Amplification Associated with ChatGPT and LLM Chatbot Use – A Psychiatric Review

They thought they were making technological breakthroughs. It was an AI-sparked delusion

SEP 5, 2025
By Hadas Gold

This man says ChatGPT sparked a ‘spiritual awakening.’ His wife says it threatens their marriage

By Pamela Brown, Clare Duffy and Shoshana Dubnow, CNN
7 min read · Updated 12:20 PM EDT, Wed July 2, 2025

UNIVERSITÄT
LUZERN

VERSTÄRKUNG VON ECHO CHAMBERS UND SOZIALEM GRABEN

Athens Journal of Politics & International Affairs 2025, 1: 1-16
https://doi.org/10.30958/ajpia.X-Y-Z

Role of Echo Chambers in the Polarization of Society

*By Desislava Angelova**

The process of digitization gives easy access to online communication. This evolution should, in theory, make political processes more democratic and increase political accountability since social networks provide instant contact with politicians and provide for accountability. Yet digitalization has also introduced new threats for democracy, with the advent of fake news, online bot networks, artificial intelligence (AI), and social media algorithms. These developments are linked to a rise of far-right parties in Europe and around the world in recent years. Populism and extremism can thrive in the [redacted] ecosystem because the relative costs of reaching a mass audience on social networks to spread their messages are low, and it is an efficient way of reaching out to potential supporters. This article examines how echo chambers form and function on social media platforms and how they are connected to driving societal polarization. The analysis is based on desk research and critical inquiry through the lenses of echo chambers, polarization, and the logic of virtual communication. The first aim of the article is to study how social media allows echo chambers to form and how online spaces are exploited by political actors. The second goal is to observe patterns of social polarization and explore how echo chambers are connected to this process.

Keywords: *Digitalization, virtual communication, social media algorithms, echo chambers, polarization, confirmation bias, selective exposure, homophily*

PERFEKTE FORM

- Format ist perfekt
- Rhetorisch stark
- Tritt mit endgültigen Aussagen auf
- Übertrifft an Inhalt meistens unsere eigene Expertise
- Erscheint freundlich und unterstützend
- KI ist zunehmend «unsichtbar» integriert in unsere alltäglichen Programme

- Skalierung: nahezu beliebig gestaltbar und übertragbar auf alle Themen
- Geschwindigkeit: digitale Kreationen können sich blitzartig verbreiten

GEGENMASSNAHMEN

- Auswahl
- Quellenkritik
- Methodenkritik
- Perspektivenvielfalt
- Stil-Prompting: «Formuliere wissenschaftlich mit *erscheint als* anstatt absolut mit *es ist so*
- Reflexion der Inhalte

Gegenmassnahmen sind nicht neu, sondern galten schon immer bei kritischem und reflektiertem Arbeiten.

Aber: KI kann bestehende Probleme im Bildungssystem verstärken.

DEEPFAKE

HISTORISCHE DEEPFAKES

World / Asia

Finance worker pays out $25 million after video call with deepfake ‘chief financial officer’

By Heather Chen and Kathleen Magramo, CNN

2 min read · Published 2:31 AM EST, Sun February 4, 2024

Slovakia: Deepfake audio of Denník N journalist offers worrying example of AI abuse

October 31, 2023 | By IPI contributor Karin Kőváry Sólymos

Manipulated audio clip involving journalist Monika Tódová shared ahead of general election portends possible future use of AI tools to discredit media

Pornographic deepfakes of Taylor Swift emerged on social media. Politicians say the story is all too common

Internet Culture

Sat 27 Jan 2024

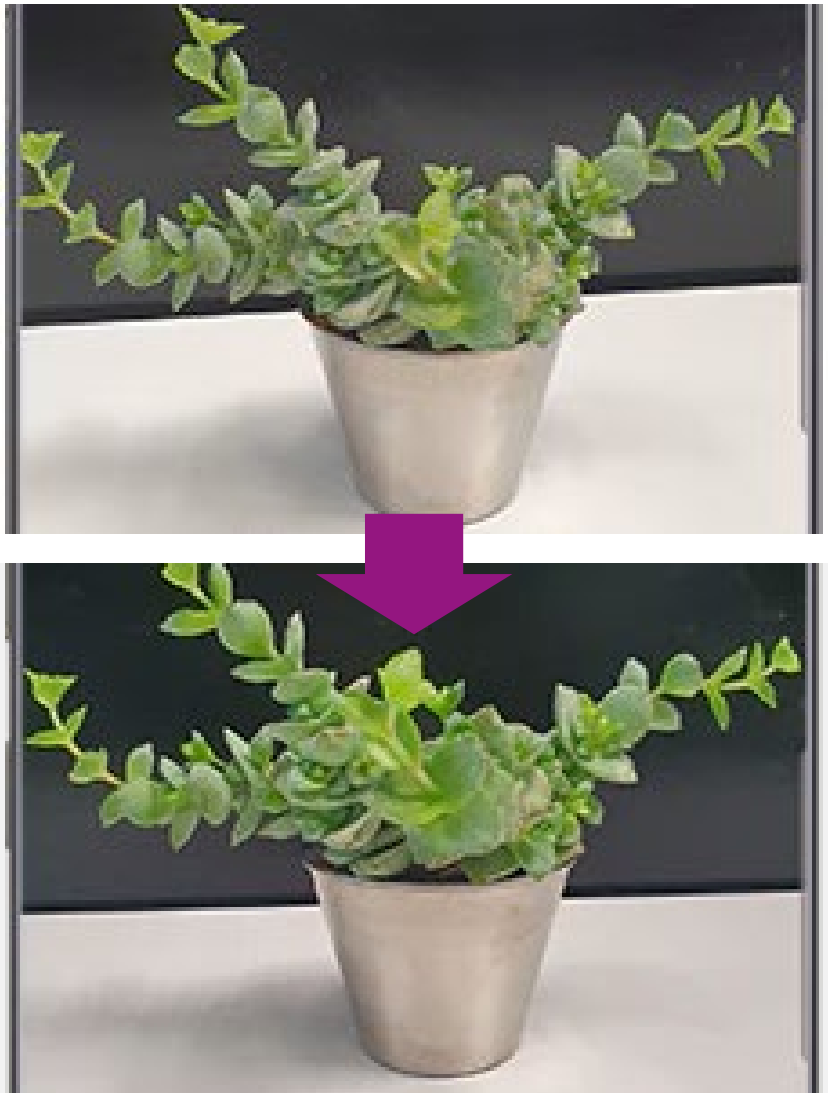
HOME > BLOG > DEEPFAKE STATISTICS AND TRENDS

Deepfake Statistics & Trends 2025: Growth, Risks, and Future Insights

Deepfakes are growing at an alarming rate—our 2025 analysis reveals key statistics, emerging trends, and the real risks businesses and individuals face.

2025-09-24

MEDIENBEARBEITUNG



AUTOMATISCHE OPTIMIERUNG VON AUFNAHMEN: FOTOS & VIDEOS

This feature uses AI to recognize subjects and scenes (like food, landscapes, or pets) and automatically adjust settings such as exposure, contrast, and color to improve the image quality. For example, it makes food look more appetizing and dark scenes brighter.

(Zusammenfassung durch Google)

What are the different scenes?

The rear camera is able to choose between twenty different modes and apply the one that suits every situation best.

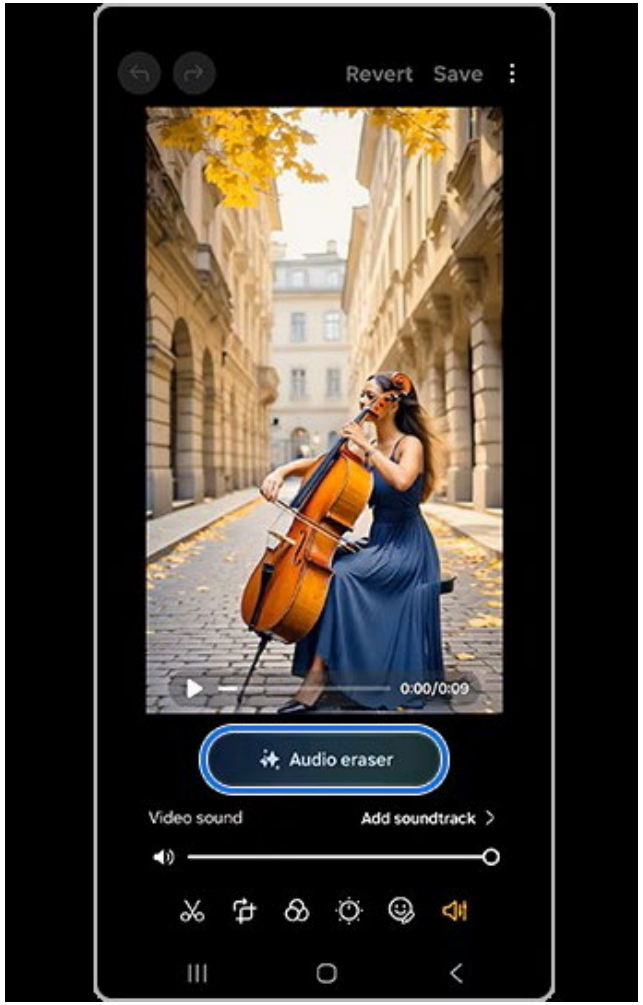
Food	Portraits	Flowers	Indoor scenes
Animals	Landscapes	Greenery	Trees
Sky	Mountains	Beaches	Sunrises and sunsets
Watersides	Street scenes	Night scenes	Waterfalls
Snow	Birds	Backlit	Text

On the S10 phones, ten more scenes are also available:

Clothing	Vehicle	Shoe	Dog
Face	Drink	Stage	Baby
People	Cat		

https://www.samsung.com/latin_en/support/mobile-devices/how-does-the-scene-optimizer-improve-my-photos/

AUTOMATISCHE OPTIMIERUNG VON AUFNAHMEN: SOUNDS



Voices – relatively clear voices of people around

Wind – outside wind or air gust noise

Music – background music

Nature – sound of nature like waves and birds

Crowd – sounds of groups of people, like cheering or applause

Noise – general unclassified noise

https://www.samsung.com/latin_en/support/mobile-devices/how-to-remove-unwanted-sound-from-videos-using-audio-eraser-on-your-galaxy/

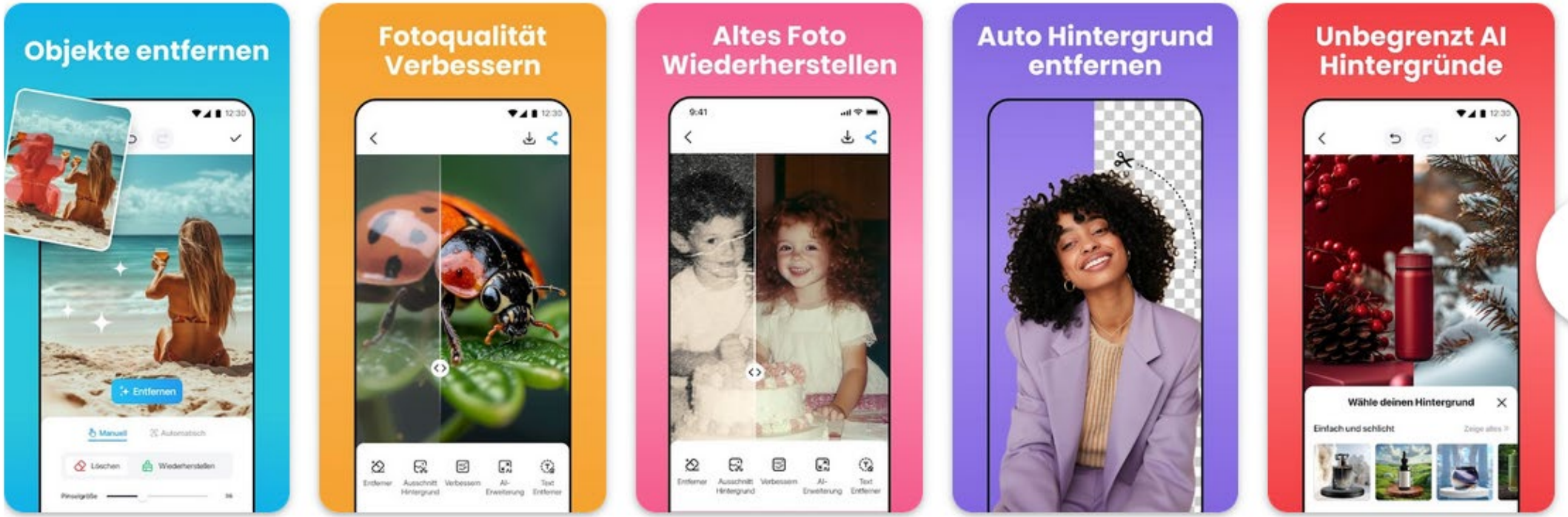
APPS ERWEITERN DIE MÖGLICHKEITEN

- Automatically fix lighting and contrast
- Remove unwanted objects or photobombers
- Sharpen blurry shots or enhance details
- Suggest aesthetic filters based on your photo content
- AI even refines skin tones, lighting, and facial details while keeping the look realistic
- Replace dull skies with dramatic sunsets
- Turn portraits into paintings or digital art
- Create motion effects

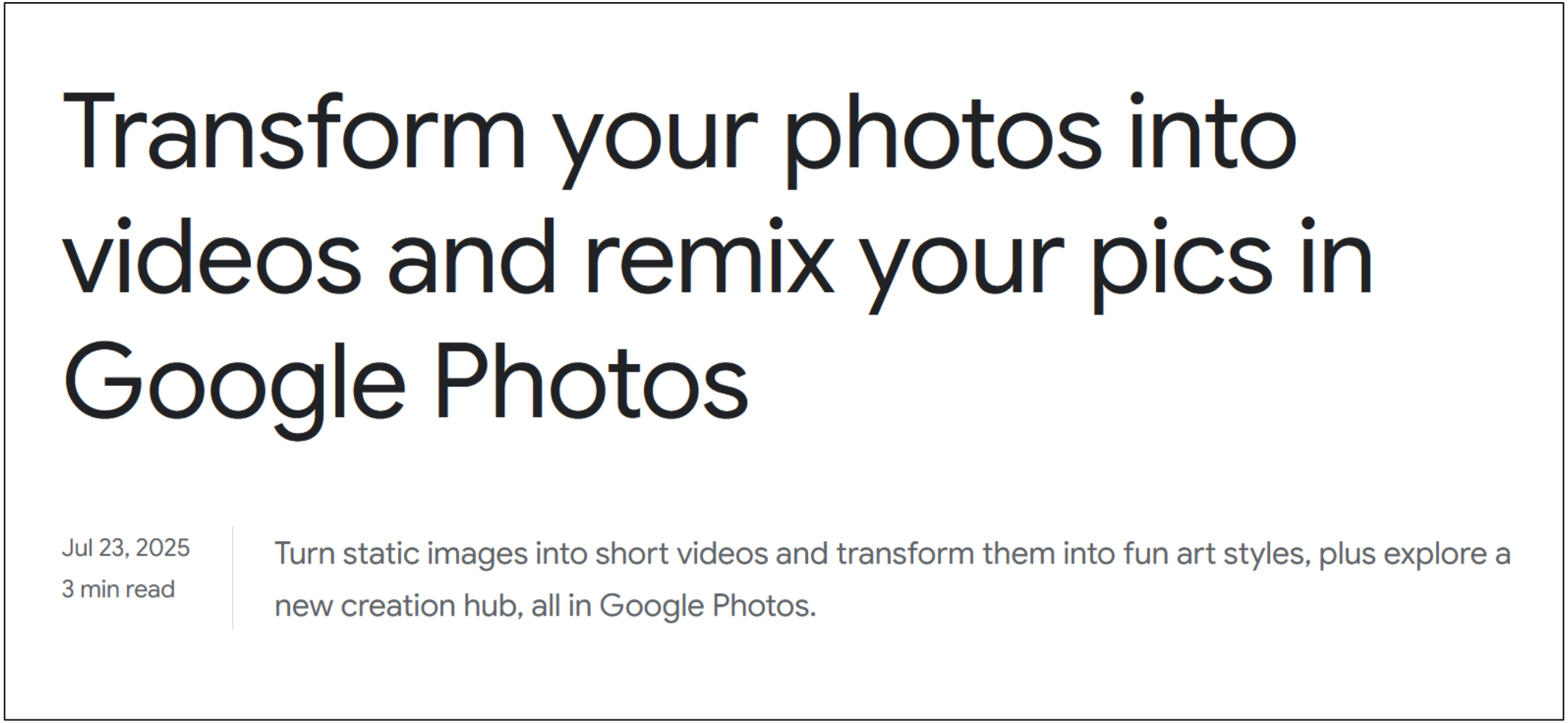
➢ Even better, these tools learn your editing preferences over time, making your future edits faster and more consistent.

https://ssa.academy/resources/the-ai-boost-how-artificial-intelligence-is-changing-smartphone-photography/

AI PHOTO EDITOR (GOOGLE PLAY)



MIT GOOGLE PHOTOS BILDER ZU VIDEOS UMWANDELN



MEDIENBEARBEITUNG KANN ERINNERUNGEN VERÄNDERN

MIT Libraries | DSpace@MIT

MIT Open Access Articles

Synthetic Human Memories: AI-Edited Images and Videos Can Implant False Memories and Distort Recollection

frontiers in Psychology

▸ Front Psychol. 2025 Jul 11;16:1645795. doi: 10.3389/fpsyg.2025.1645795

The algorithmic self: how AI is reshaping human identity, introspection, and agency

Jeena Joseph [1,*]

▸ Author information ▸ Article notes ▸ Copyright and License information

PMCID: PMC12289686 PMID: 40718563

AUSWIRKUNGEN AUF GERICHTSVERFAHREN UND SELBSTBILD

June 11, 2025

Safeguarding the Courtroom from AI-Generated Evidence: Federal Rule of Evidence 707 Approved by Judicial Conference

By Shane G. Ramsey

Advice > Body Dysmorphia Disorder

AI Filters: How Virtual Reality Shapes Mental Health Today

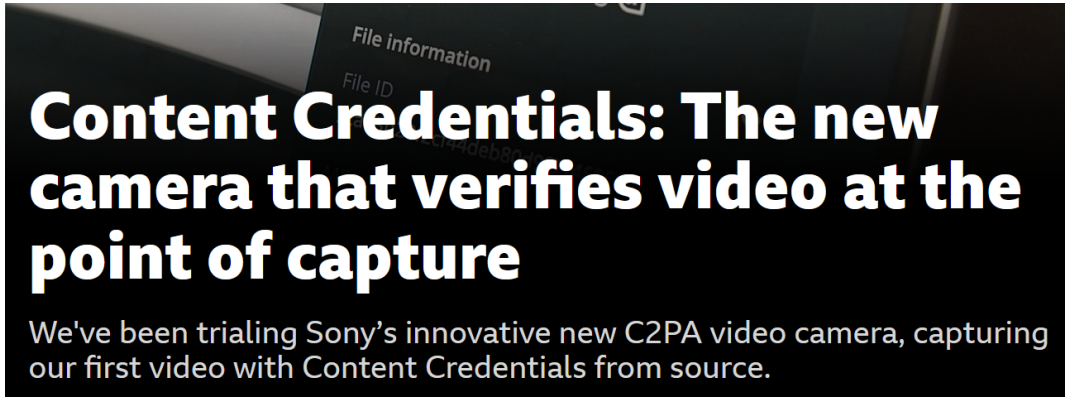
August 20, 2025

11 min read

AI filters and virtual reality technology significantly impact mental health by altering self-perception and potentially triggering body dysmorphic behaviors, with research showing direct links to decreased self-esteem and body image concerns that often require professional therapeutic intervention for healthy coping and digital wellness.

GEGENMASSNAHMEN?

- Medienbearbeitung: selbst nicht nutzen
- Deepfakes: spezifische Apps, aber unsicher
- Digital Watermarking
- Firmeninitativen
- Vertrauensvolle Quellen



GENERATIVE AI, OUR INSIGHTS

How to Detect and Prevent Deepfakes in 2025: A Complete Guide for the AI Era

MAY 14, 2025 BESTARION

The rise of generative AI has brought forth revolutionary innovations across industries, from automated content creation to personalized recommendations. However, alongside these benefits, a darker challenge has emerged—**deepfakes**. These synthetic videos, images, or audio recordings generated by AI can be nearly indistinguishable from real ones, making them powerful tools for misinformation, fraud, and identity theft.