TLP:CLEAR
Contains AI

DeepFakes Unmasked

Candid Wüest

Security Advocate @ xorlab

xorlab

2025

Which photo was generated by AI?

DEEP FAKE

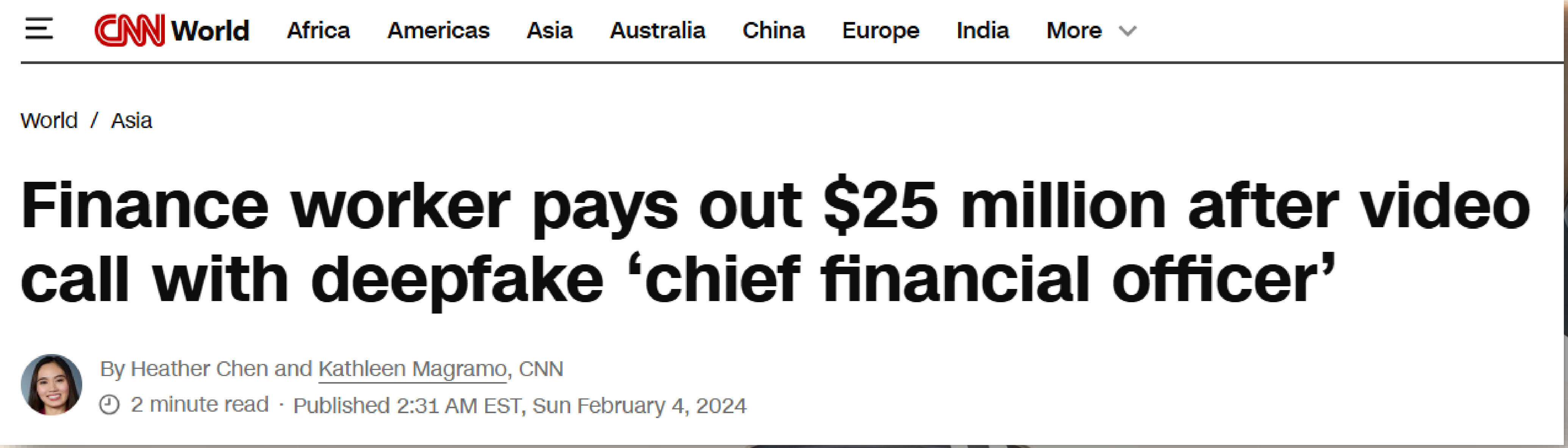
DEEP FAKE

DEEP FAKE

Source: https://thispersondoesnotexist.com

DeepFake

Business Email Compromise (BEC) Scams



World / Asia

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

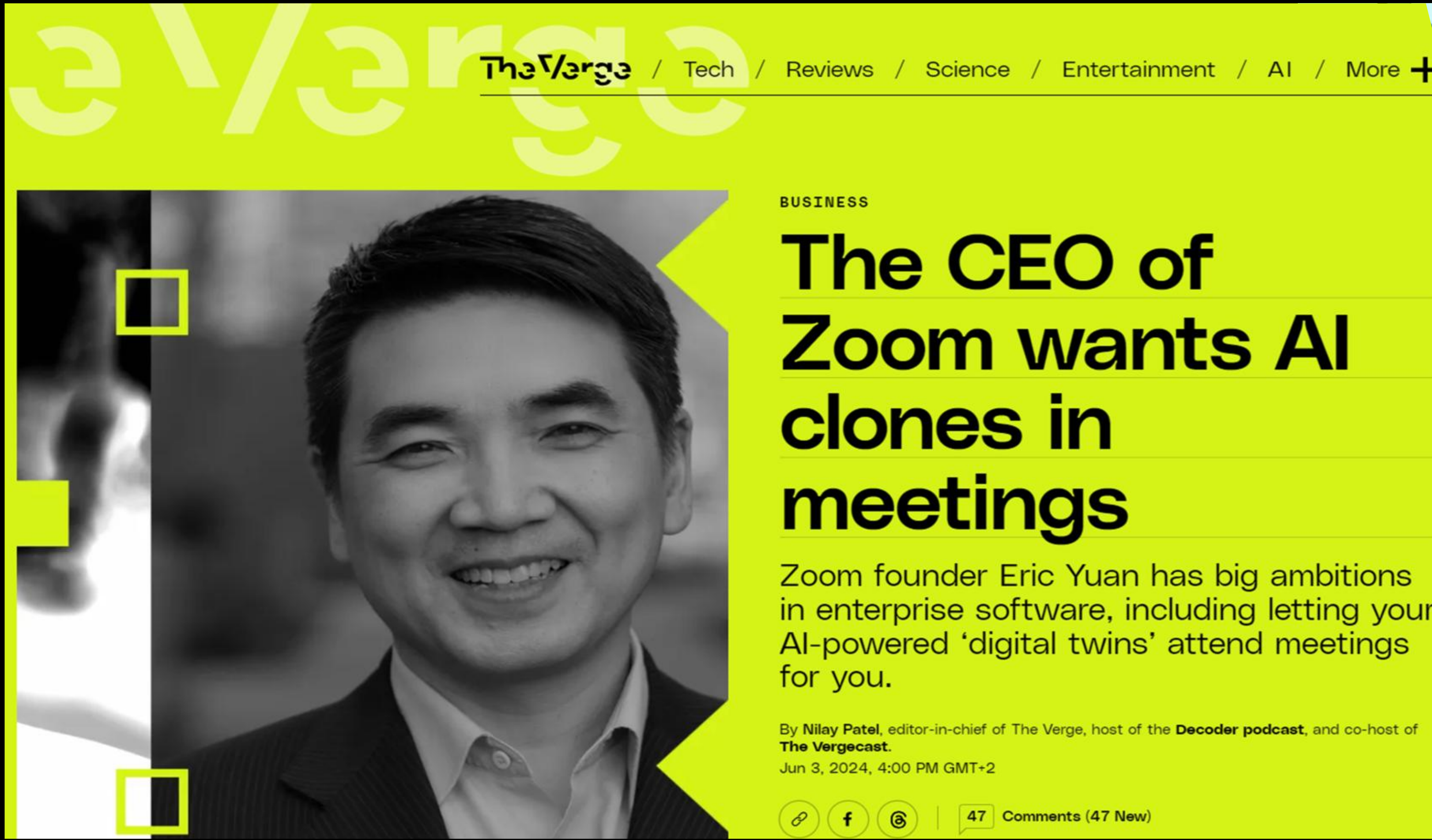
By Heather Chen and Kathleen Magramo, CNN

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

How to spot an imposter?

© Terminator2 - Tri-Star Pictures

Virtual - Virtual Meeting



The Verge / Tech / Reviews / Science / Entertainment / AI / More +

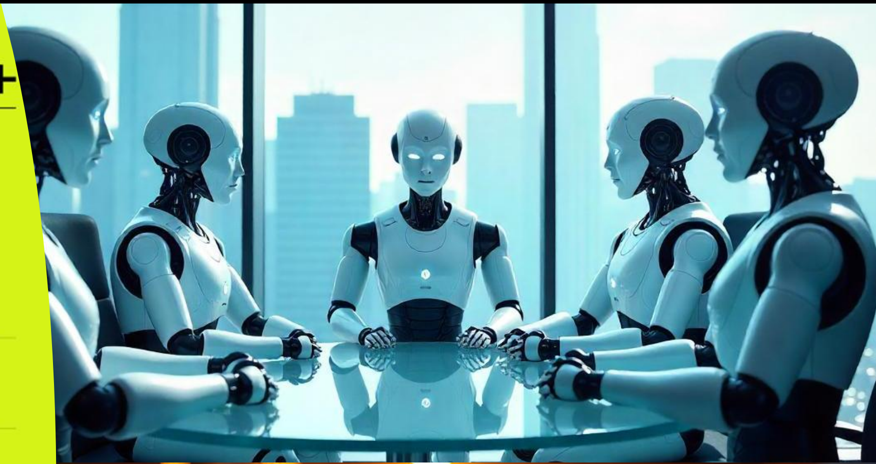
BUSINESS

The CEO of Zoom wants AI clones in meetings

Zoom founder Eric Yuan has big ambitions in enterprise software, including letting your AI-powered 'digital twins' attend meetings for you.

By **Nilay Patel**, editor-in-chief of The Verge, host of the **Decoder podcast**, and co-host of **The Vergecast**.

Jun 3, 2024, 4:00 PM GMT+2



Different Generation Methods

Partial

e.g. Face Swap

Full Fake

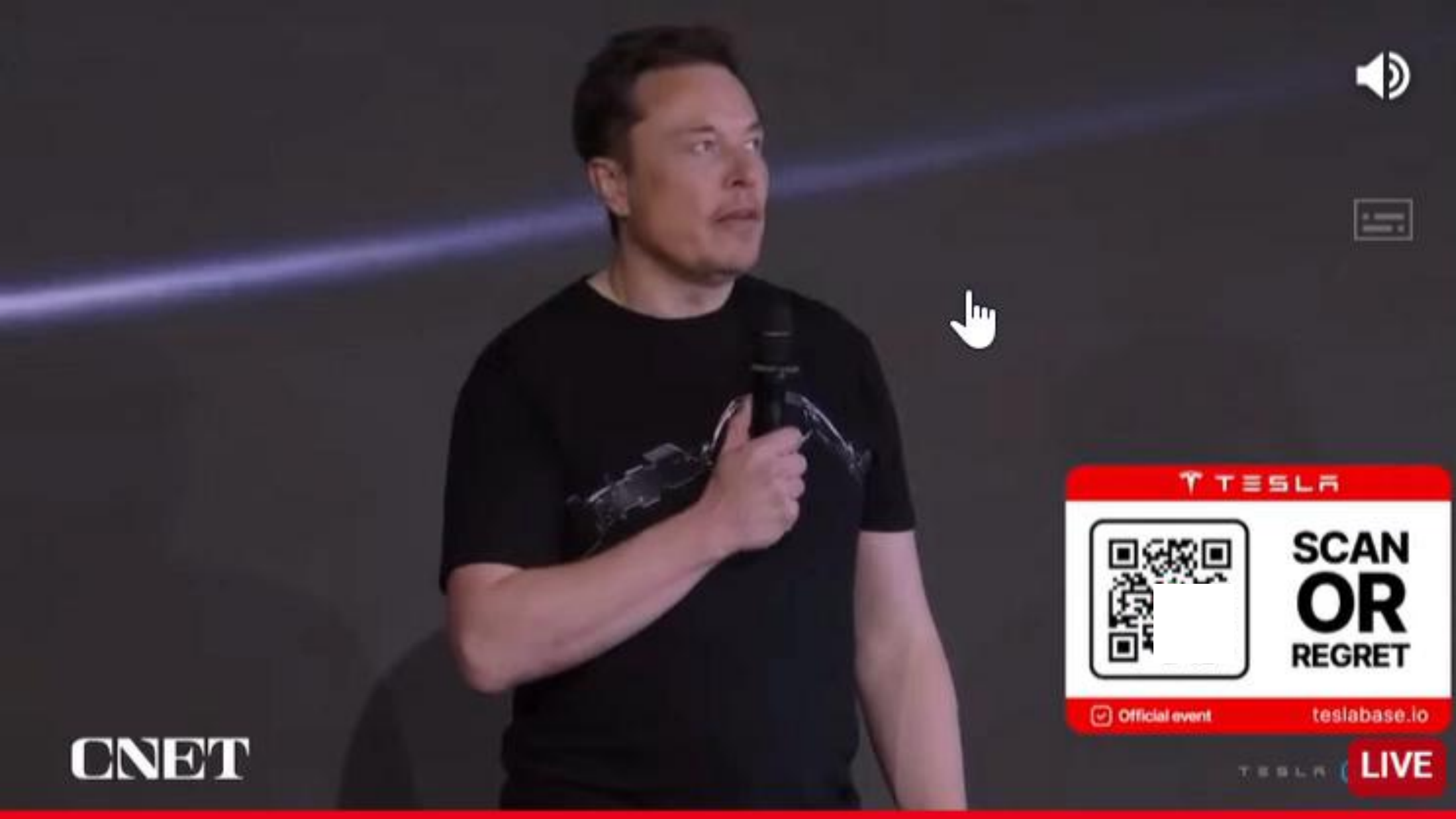
e.g. Twitter Ad

Pre-Generated

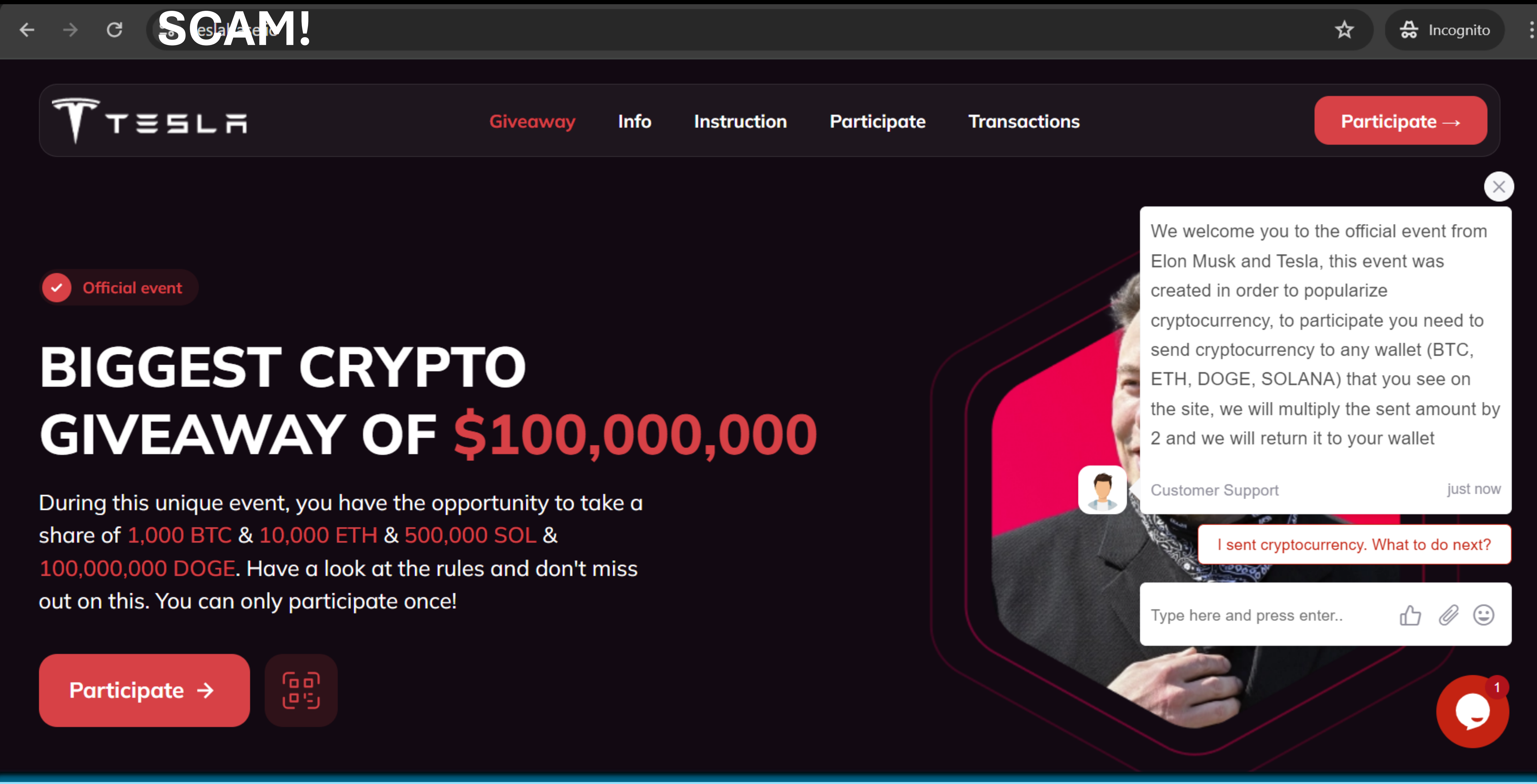
Real Time

e.g. to interact

☺ Parallel lines have so much in common. It’s a shame they’ll never meet.


TESLA
SCAN
OR
REGRET
☑ Official event
teslabase.io
CNET
TESLA
LIVE

SCAM!



BIGGEST CRYPTO GIVEAWAY OF $100,000,000

During this unique event, you have the opportunity to take a share of 1,000 BTC & 10,000 ETH & 500,000 SOL & 100,000,000 DOGE. Have a look at the rules and don't miss out on this. You can only participate once!

We welcome you to the official event from Elon Musk and Tesla, this event was created in order to popularize cryptocurrency, to participate you need to send cryptocurrency to any wallet (BTC, ETH, DOGE, SOLANA) that you see on the site, we will multiply the sent amount by 2 and we will return it to your wallet

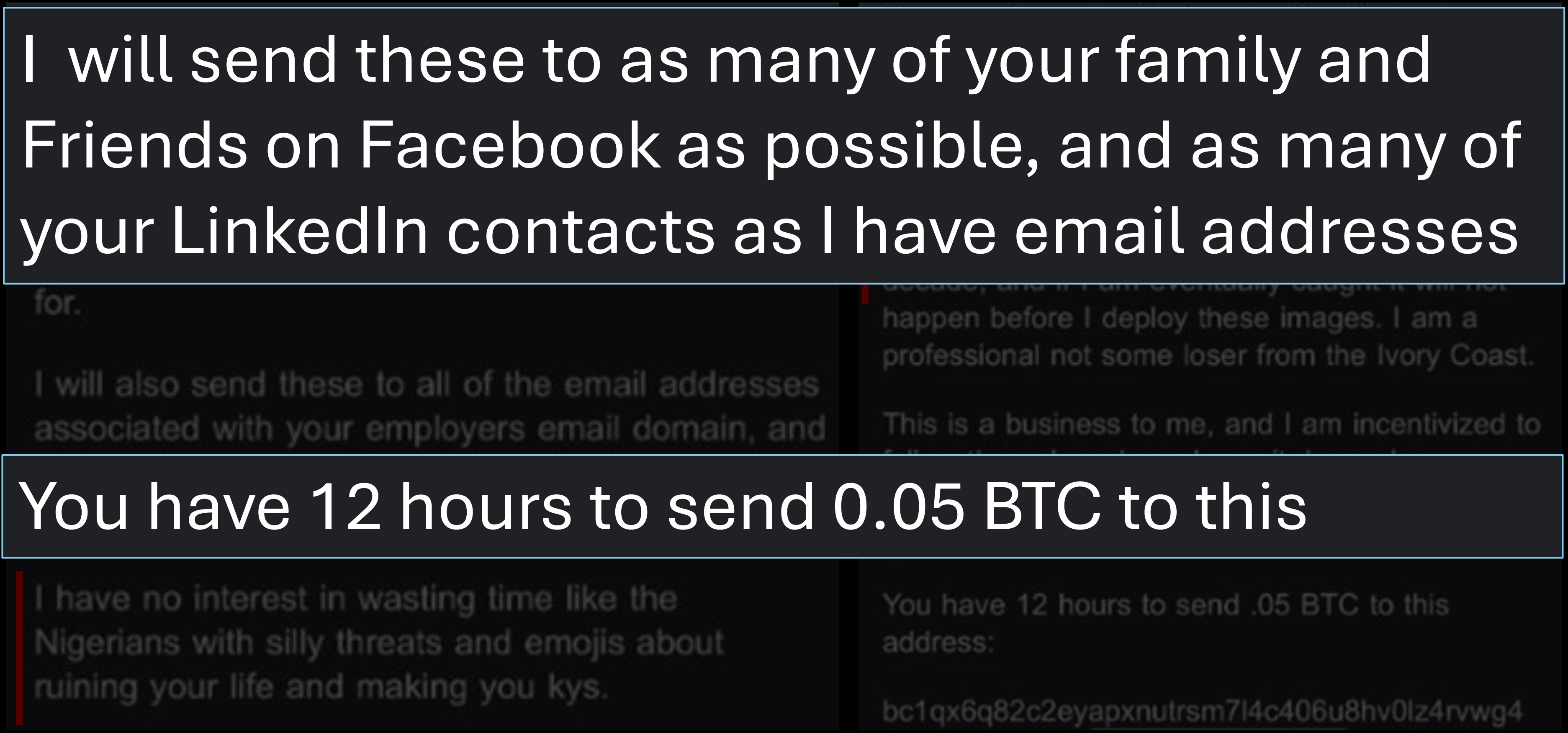
Customer Support just now

I sent cryptocurrency. What to do next?

Type here and press enter..

☺ Why does your serverless app return “Internal server error”?

DeepFake Sextortion Scams



I will send these to as many of your family and Friends on Facebook as possible, and as many of your LinkedIn contacts as I have email addresses

You have 12 hours to send 0.05 BTC to this

for.

I will also send these to all of the email addresses associated with your employers email domain, and

I have no interest in wasting time like the Nigerians with silly threats and emojis about ruining your life and making you kys.

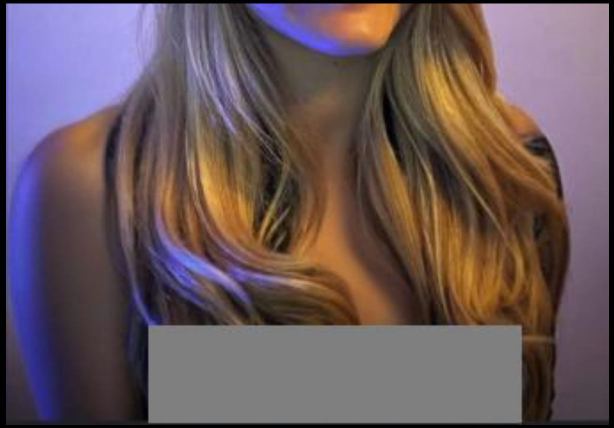
happen before I deploy these images. I am a professional not some loser from the Ivory Coast.

This is a business to me, and I am incentivized to

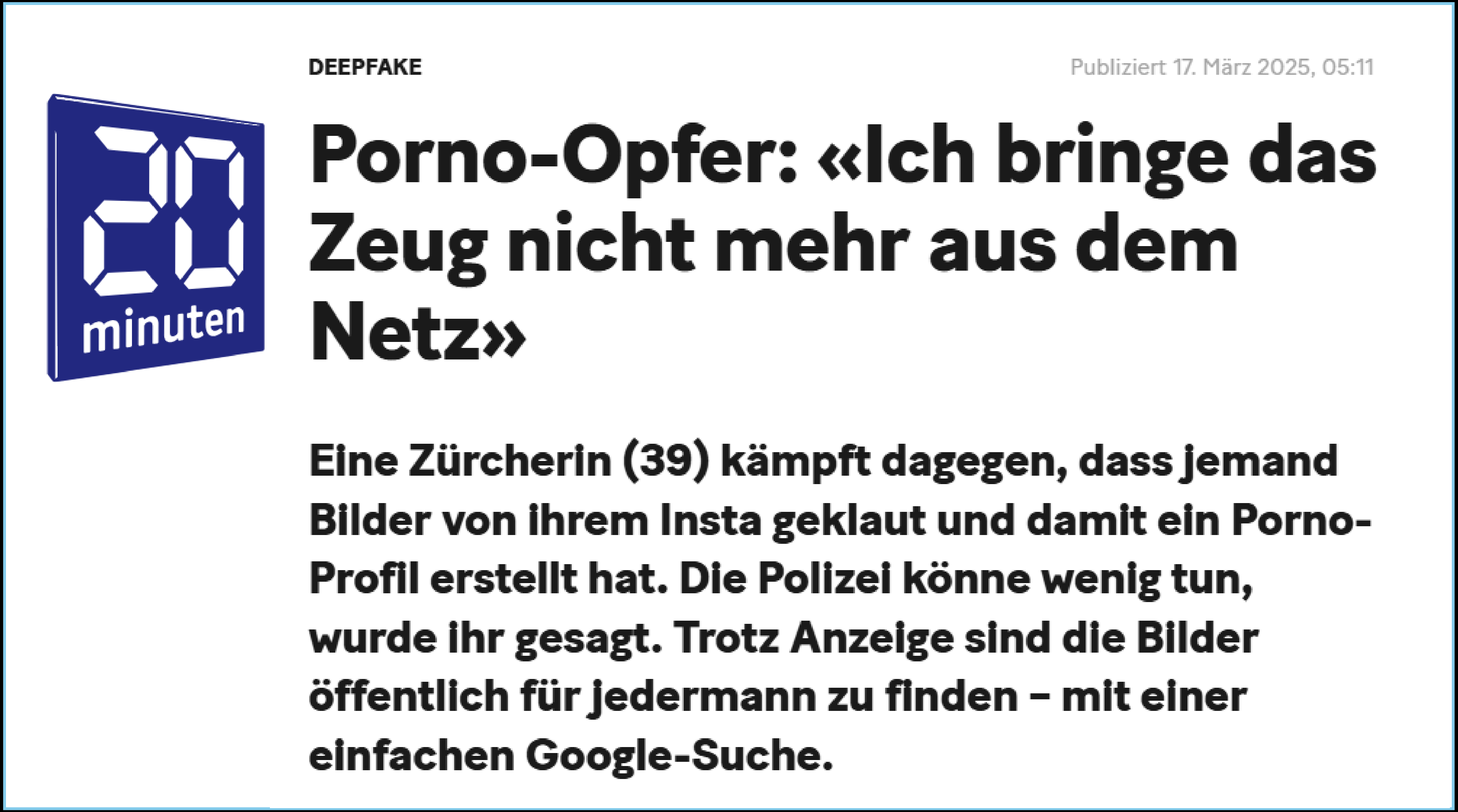
You have 12 hours to send .05 BTC to this address:

bc1qx6q82c2eyapxnutrsm7l4c406u8hv0lz4rvwg4

Becca Caddy
@beccacaddy



More and more “nudify” apps with GenAI



DEEPFAKE

Publiziert 17. März 2025, 05:11

Porno-Opfer: «Ich bringe das Zeug nicht mehr aus dem Netz»

Eine Zürcherin (39) kämpft dagegen, dass jemand Bilder von ihrem Insta geklaut und damit ein Porno-Profil erstellt hat. Die Polizei könne wenig tun, wurde ihr gesagt. Trotz Anzeige sind die Bilder öffentlich für jedermann zu finden – mit einer einfachen Google-Suche.

☺ I trained an AI to detect sarcasm. It said, “Oh wow, what a great idea.”

Romance with Brad Pitt costs €830’000

Source: X.com

☺ I told my wife she should embrace her mistakes. She gave me a hug

Don’t mock the victim!

“Four films to see with Brad Pitt (really) for free”

Netflix France

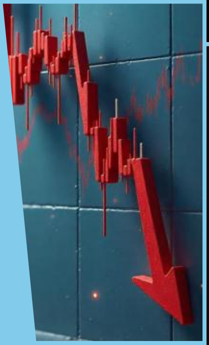
“Hi Anne, Brad told us he would be at the stadium on Wednesday... and you?”

Toulouse FC

Dis-information campaign examples

Corporate Stock drop

e.g. CEO dead



Insurance Fraud

e.g. car repair

Imposter

e.g. son in trouble scams

Deception

e.g. political elections / surrendering

☺ Team work is important; it helps to put the blame on someone else.

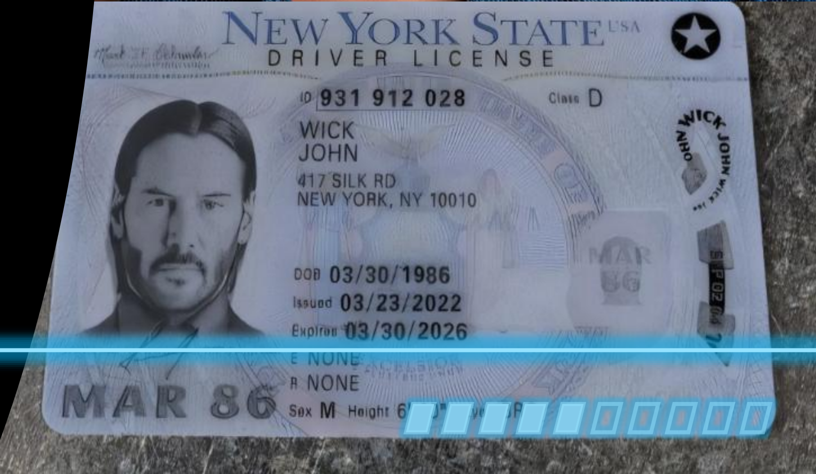
DeepFake as-a-service offers

OnlyFakes & Co. on TOR and Telegram

e.g. for Login or KYC bypass for crypto currency exchanges

US$5 per image to US$500 per minute of video

- Fully synthetic ID → fake account
- Photo → Fake ID card (sign-up KYC)
- Passport → selfie with ID (account verification)



NEW YORK STATE USA
DRIVER LICENSE
ID 931 912 028 Class D
WICK
JOHN
417 SILK RD
NEW YORK, NY 10010
DOB 03/30/1986
Issued 03/23/2022
Expires 03/30/2026
E NONE
R NONE
MAR 86
Sex M Height 6

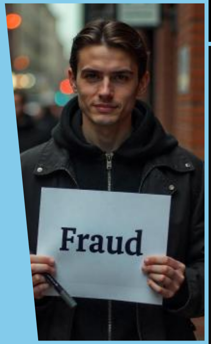
☺ My password has been hacked. Now I have to rename the cat.

Many Different Scams

BEC / CEO Fraud

Dis-information

Corporate Fraud



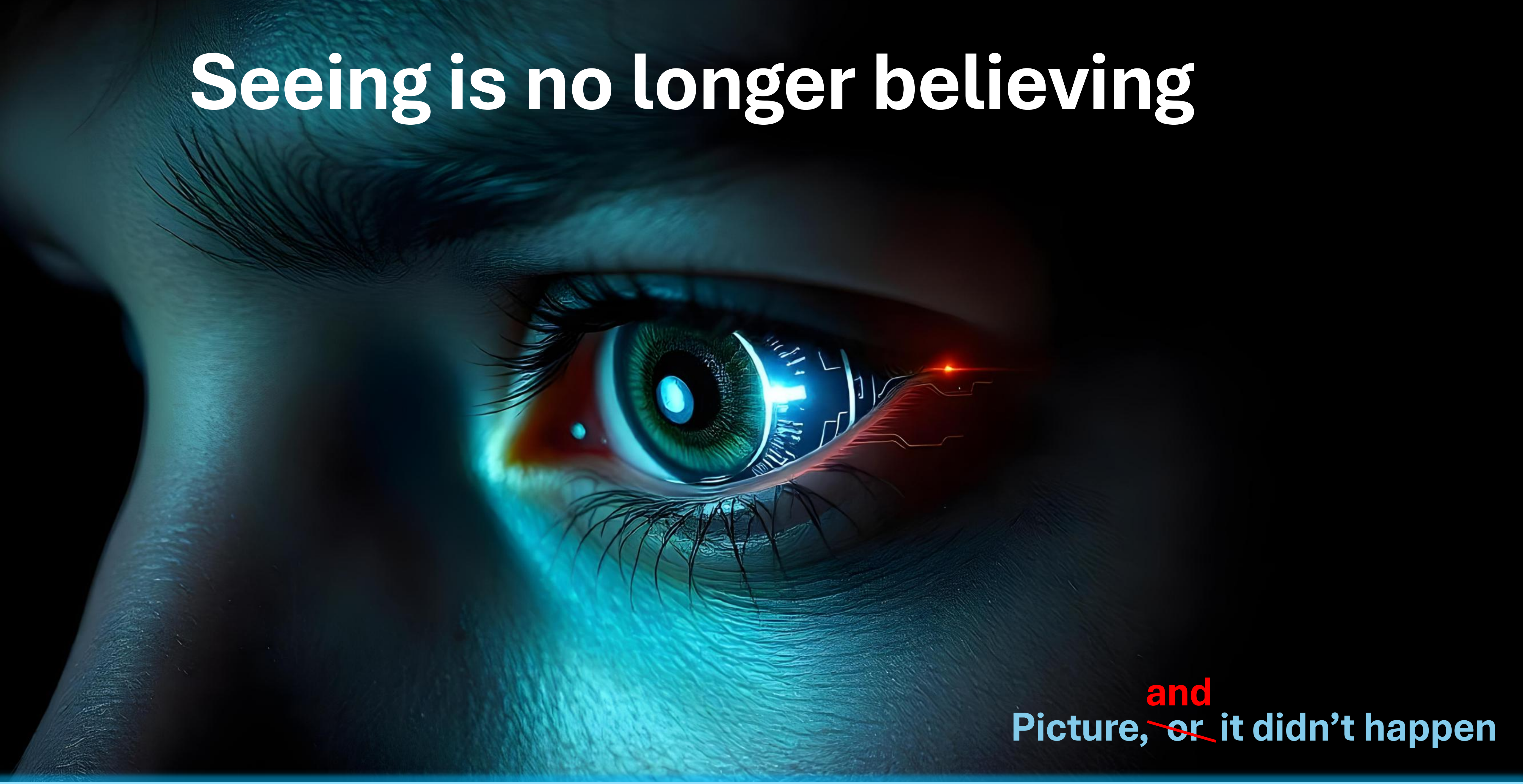
Deep Fake Porn

Imposter Scams

Auth Bypass

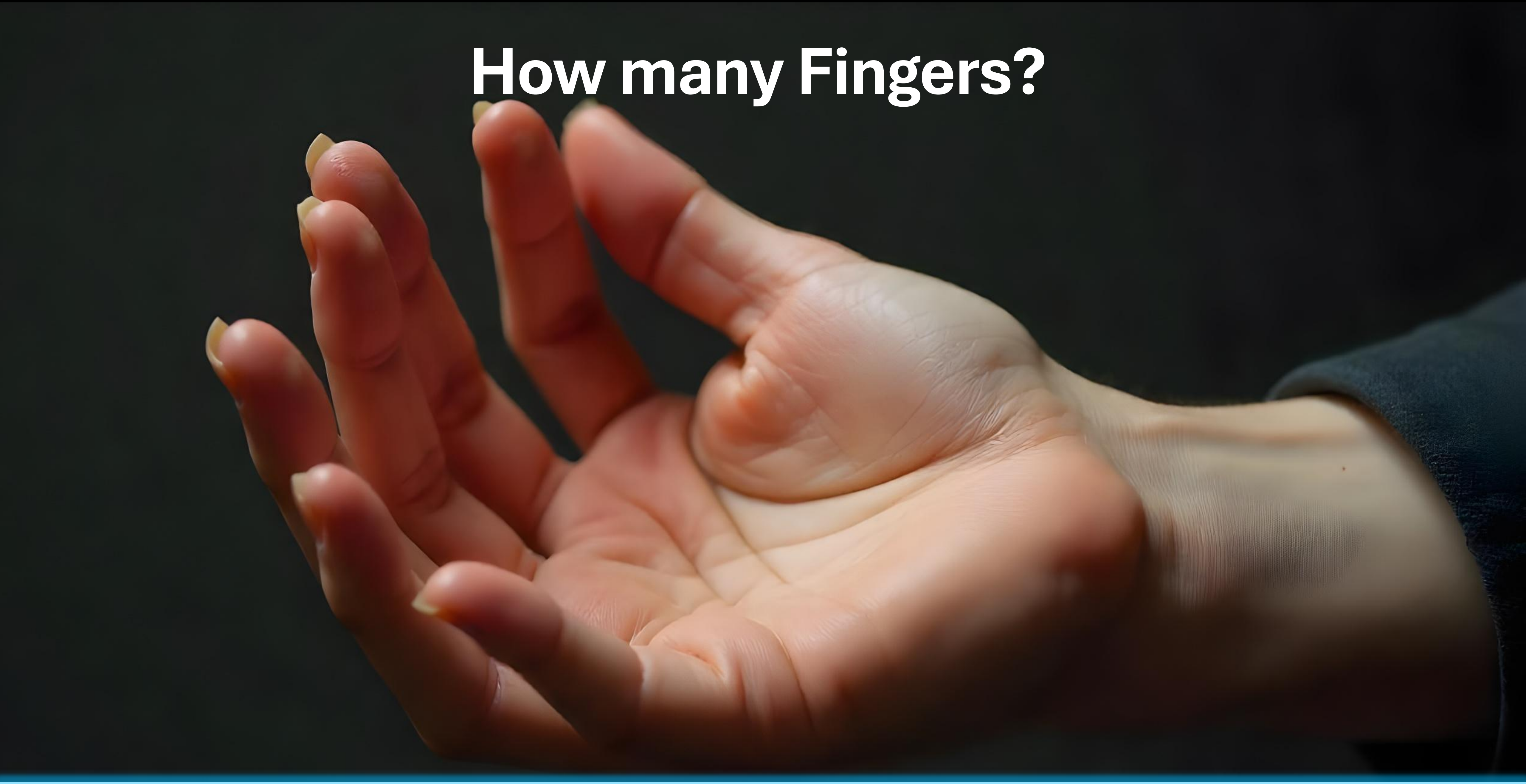
☺ I've had amnesia as long as I can remember.

Seeing is no longer believing

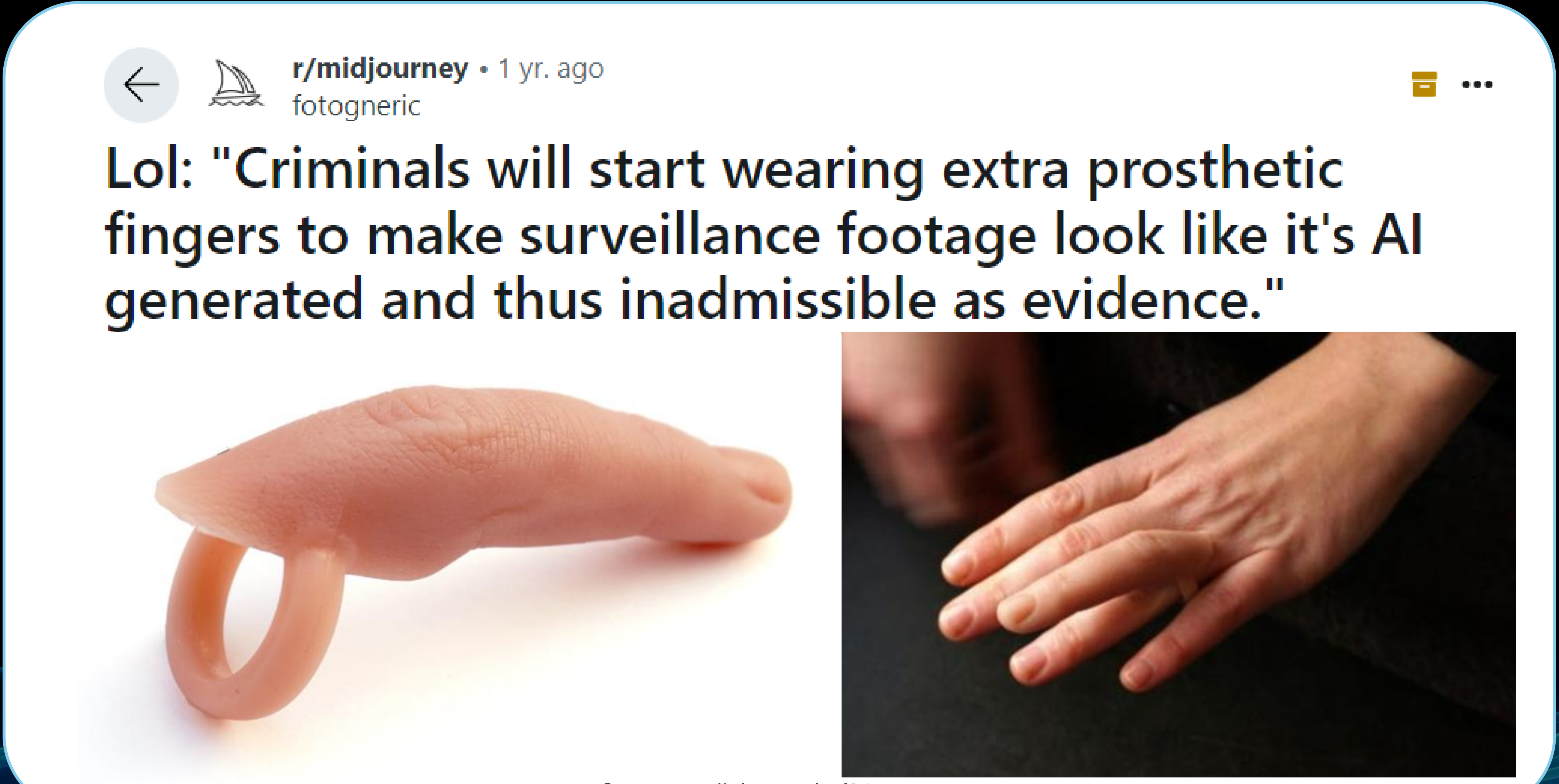


Picture, ~~or~~ and it didn't happen

How many Fingers?



Fake the DeepFake



r/midjourney • 1 yr. ago
fotogneric

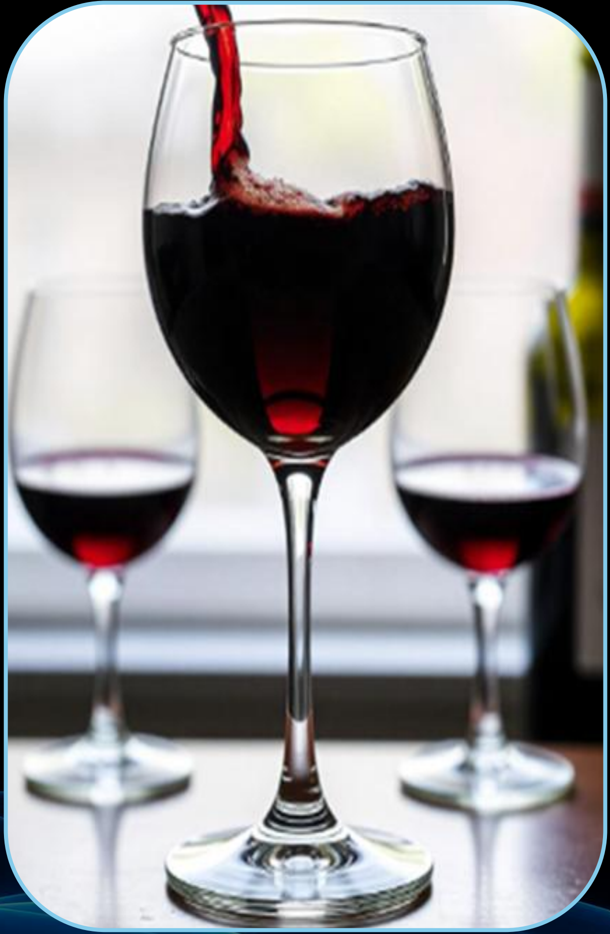
Lol: "Criminals will start wearing extra prosthetic fingers to make surveillance footage look like it's AI generated and thus inadmissible as evidence."

Source: nadjabuttendorf24.com

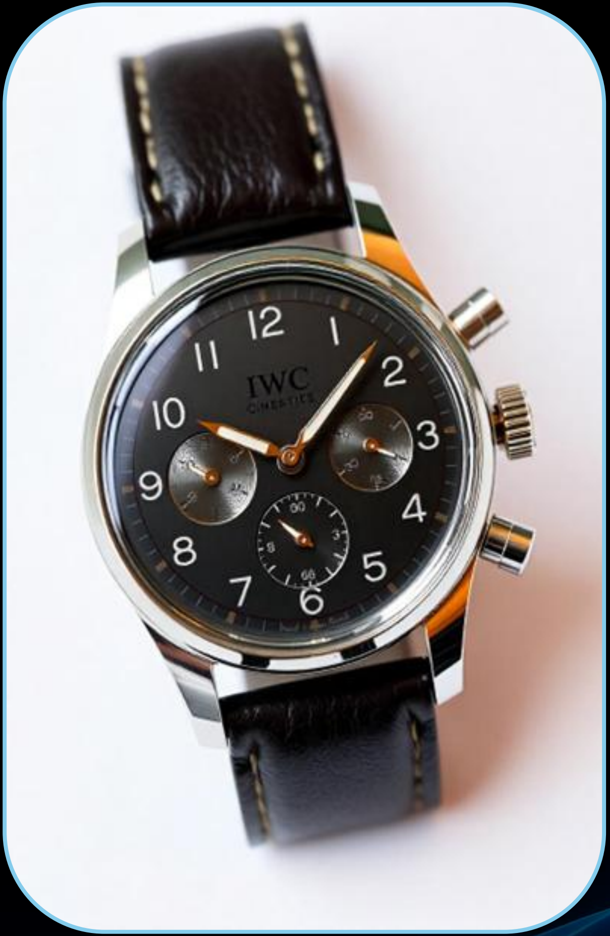
☺ The early bird might get the worm, but the second mouse gets the cheese.

The glass is half full (with AI)

A full wineglass?



Watch at 6 pm?



☺ I’m not messy. I’m organizationally challenged!

Cat and Mouse Game

☺ Sometimes you succeed and other times you learn.

DeepFake Mitigation Methods

DeepFake Detection

e.g. AI vs. AI

Source Signing

e.g. watermark

Fact Check

e.g. external validators

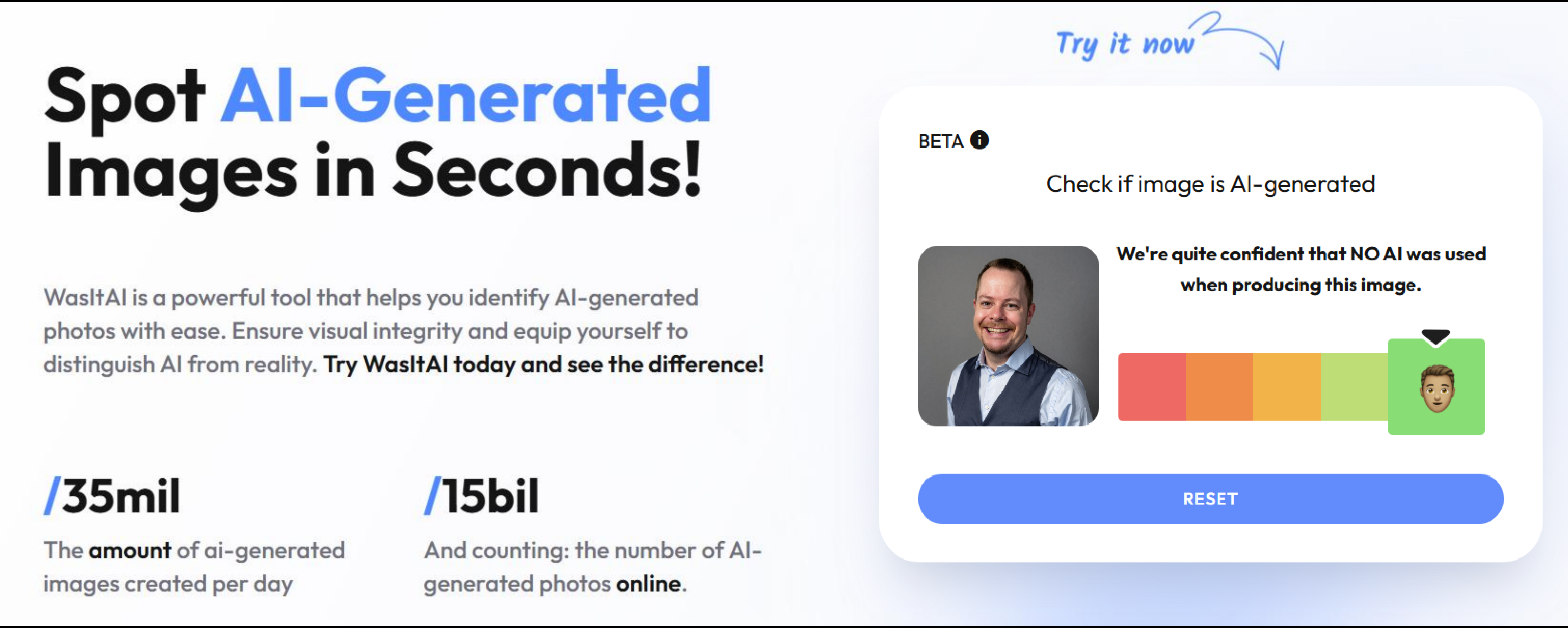
Process changes

e.g.limit actions

ID verification

☺ If time is money are ATM's time machines?

WasitAI.com PNG→JPG



Spot AI-Generated Images in Seconds!

WasItAI is a powerful tool that helps you identify AI-generated photos with ease. Ensure visual integrity and equip yourself to distinguish AI from reality. **Try WasItAI today and see the difference!**

/35mil

The **amount** of ai-generated images created per day

/15bil

And counting: the number of AI-generated photos **online**.

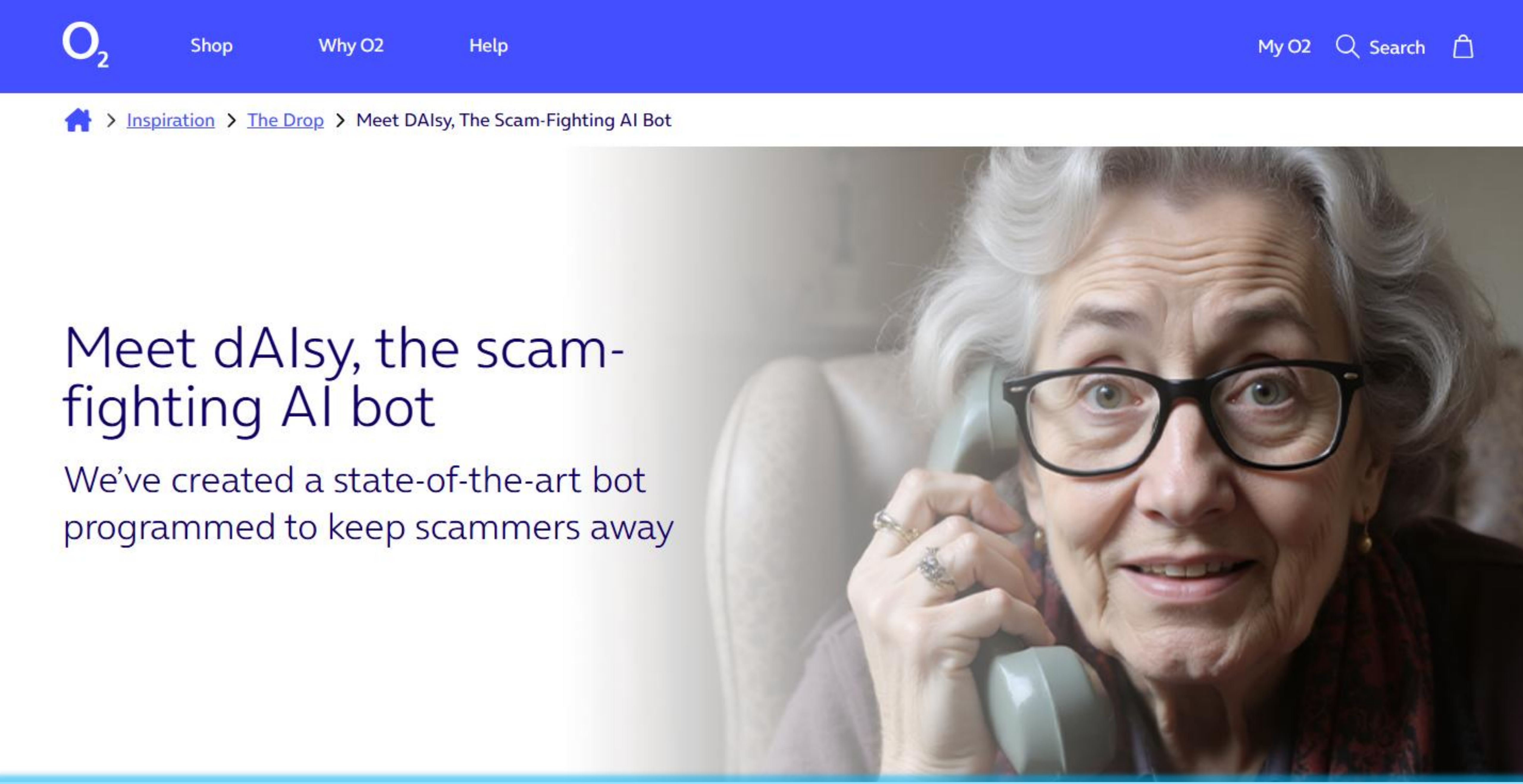
Try it now

BETA ⓘ

Check if image is AI-generated

We're quite confident that NO AI was used when producing this image.

RESET


Shop
Why O2
Help
My O2
Search
Inspiration > The Drop > Meet DAIsy, The Scam-Fighting AI Bot

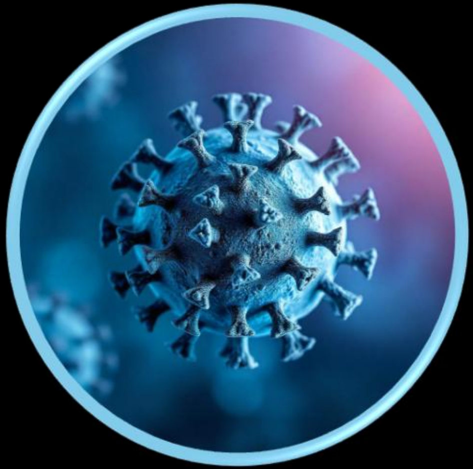
Meet dAIsy, the scam-fighting AI bot

We’ve created a state-of-the-art bot programmed to keep scammers away

Other Types of AI Attacks

AI Phishing

AI Pentest



AI Malware

Attacking AI

☺ I can totally keep secrets. It's the people I tell them to that can't

Additional AI Threats on the Horizon

Today

- Social media bots
- Personalized phishing
- Malware creation
- Auto pentesting
- Prompt injections

Soon

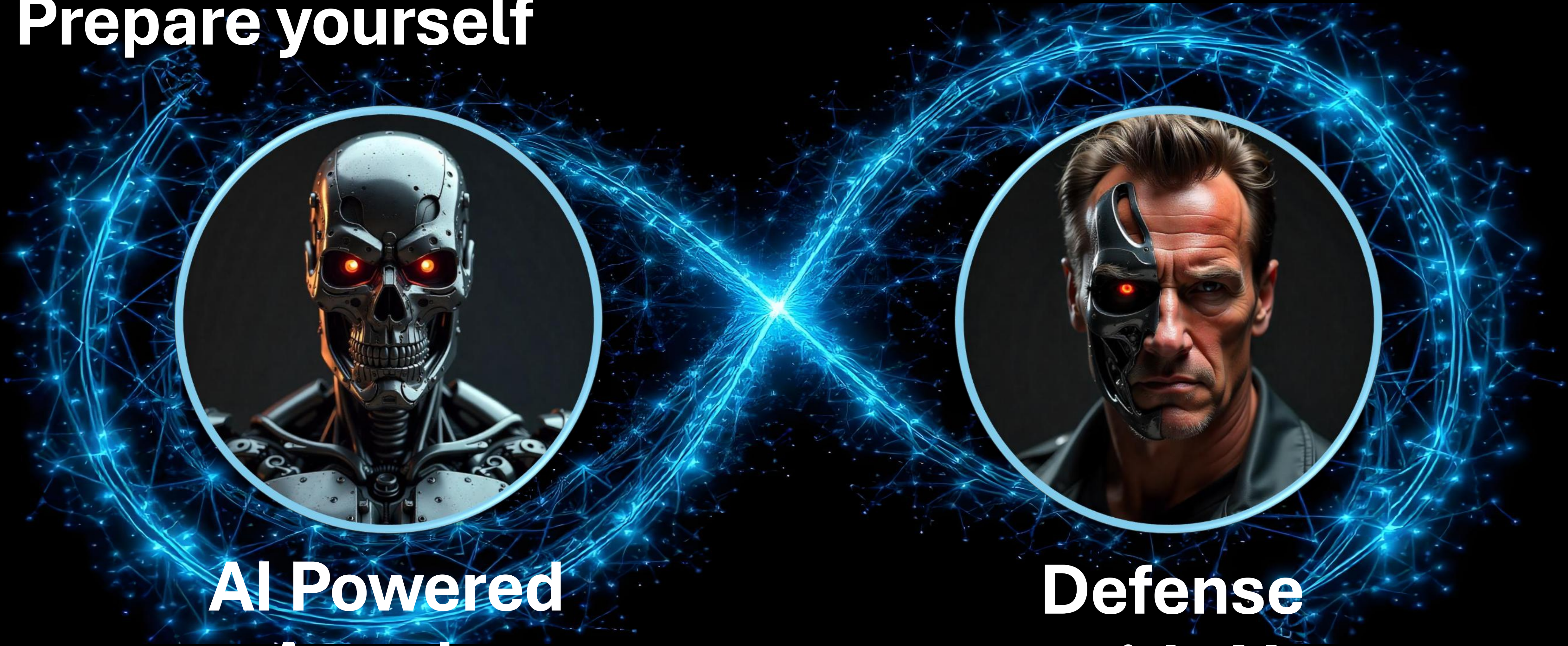
- Hijack the AI itself
- Auto AI-attack agents
- Extract AI models
- Large data poisoning
- AI-driven insiders

Future

- Mass real-time fakes
- Personalized malware
- Auto evasion bots
- Misinformation farms
- AI vs. AI fights

☺ What happens if you get scared half to death twice?

Prepare yourself



AI Powered Attacks

Defense with AI

Conclusion

1. DeepFakes boost Social Engineering
2. Detection is a cat & mouse game
3. People are not fully aware of the risks
4. Erosion of Trust
5. Use AI to protect efficiently

☺ There are two rules for success: 1) Don't tell all you know.

xorlab

Thank you for your attention!

Candid Wüest

Get the slides

My LinkedIn