DOES THE EU AI ACT SUFFICIENTLY PROTECT USERS OF SOCIAL MEDIA PLATFORMS AGAINST DEEPFAKE MISINFORMATION?

FAKULTÄT FÜR SPRACH- UND LITERATURWISSENSCHAFTEN
UNIVERSITÄT BAYREUTH

FINJA OTREMBA
BT725424@MYUBT.DE

UNIVERSITÄT BAYREUTH

DANGERS OF DEEPFAKES

"THE THREAT OF DEEPFAKES [...] COMES NOT FROM THE TECHNOLOGY USED TO CREATE IT, BUT FROM PEOPLE'S NATURAL INCLINATION TO BELIEVE WHAT THEY SEE, AND AS A RESULT DEEPFAKES [...] DO NOT NEED TO BE PARTICULARLY ADVANCED OR BELIEVABLE IN ORDER TO BE EFFECTIVE IN SPREADING MIS/DIS INFORMATION."

HOMELAND SECURITY: INCREASING THREATS OF DEEPFAKE IDENTITIES

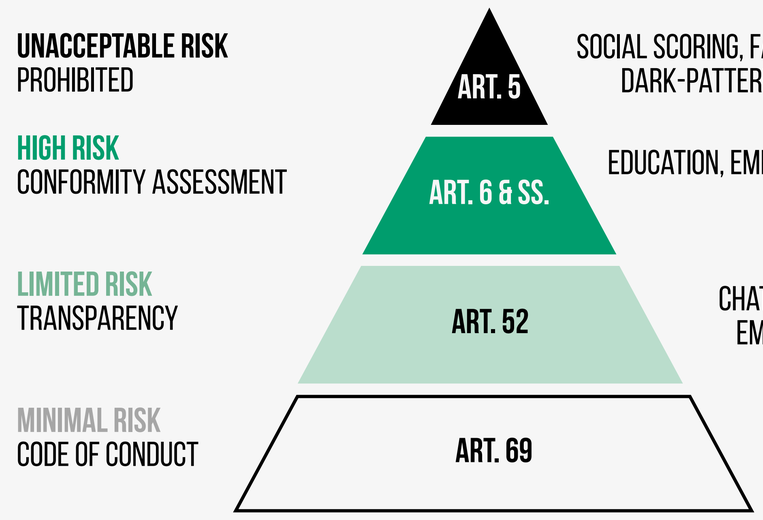
REAL-TIME DEEPFAKES

CAN BE USED TO IMITATE VOICES IN ORDER TO SCAM PEOPLE OVER THE PHONE, PRETENDING TO BE A FRIEND OR RELATIVE

MANIPULATED MEDIA

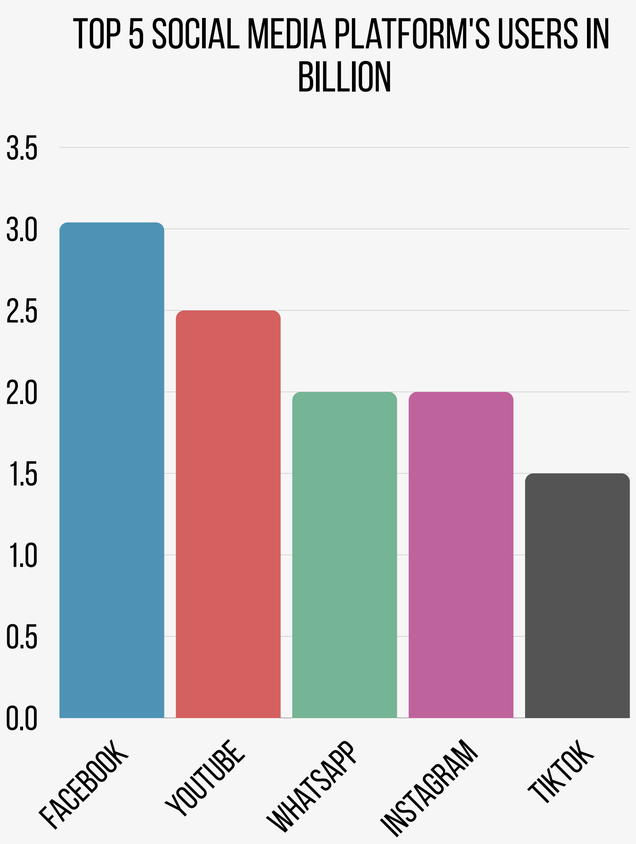
CAN BE USED TO LIP-SYNC NEW AUDIO TO PRE-EXISITING VIDEOS, PUTTING WORDS IN PEOPLES MOUTH THEY HAVE NEVER SAID, OR FACE-SWAPPING THEM WITH ANOTHER PERSON, PUTTING THEM AT A TIME AND PLACE, THEY´VE NEVER BEEN

DEEPFAKES CATEGORIZED IN THE EU AI ACT



SINCE DEEPFAKES ONLY FALL INTO THE "LIMITED RISK" CATEGORY, THEY NEED TO CONFORM TO THE TRANSPARENCY REGULATIONS OF THE EU AI ACT, INFORMING VIEWERS OF THE DEEPFAKE, THAT IT WAS AI GENERATED

5,24 BILLION USERS ON ALL SOCIAL MEDIA PLATFORMS



CAN YOU SPOT THE (DEEP) FAKE?

WHICH PERSON DO YOU THINK IS **NOT** REAL?

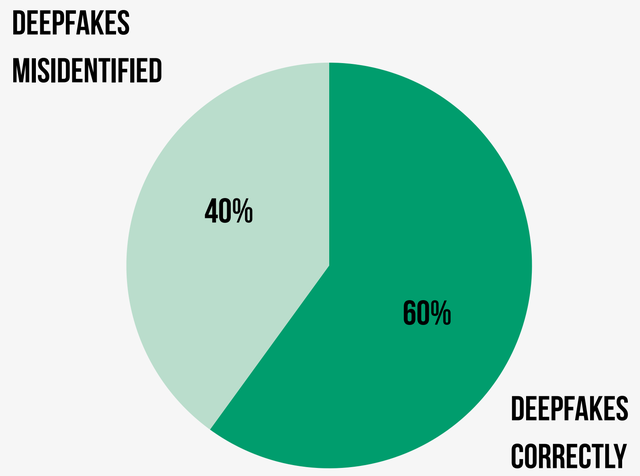
A

B

THE AVERAGE USER OF SOCIAL MEDIA ...

... VISITS AROUND 7 DIFFERENT SOCIAL PLATFORMS PER MONTH

... SPENDS ON AVERAGE 2 HOURS AND 23 MINUTES PER DAY ON SOCIAL MEDIA, WITH USERS BETWEEN THE AGES OF 16-24 SPENDING AN AVERAGE OF 3 HOURS AND 38 MINUTES



WHAT DO SOCIAL MEDIA PLATFORMS DO TO COMBAT AI-GENERATED CONTENT?

AI LABELS

FACEBOOK

ONLY HAS AN AI-LABELLING OPTION ON MOBILE DEVICES AND NOT WHEN USING A BROWSER

YOUTUBE

HAS A YES/NO OPTION FOR IF THE CONTENT WAS ALTERED TO A POINT, THAT THE SHOWN FOOTAGE NEVER ACTUALLY HAPPENED. ONLY IN THE FINE PRINT IS AI MENTIONED.

WHATSAPP

NO AI-LABEL WHEN SHARING MEDIA IN CHATS OR WHEN POSTING A STATUS.

INSTAGRAM

ONLY HAS AN AI-LABELLING OPTION ON MOBILE DEVICES AND NOT WHEN USING A BROWSER

TIKTOK

HAS AN AI-LABLE SETTING FOR POSTED CONTENT BOTH IN THE APP AND WEBSITE, HOWEVER YOU HAVE TO CLICK ONE "SHOW MORE" BUTTON IN THE POSTS SETTINGS FIRST.

THE EU AI ACTS REGULATIONS REGARDING DEEPFAKES DON´T FULLY COME INTO EFFECT UNTIL 2026. SOCIAL MEDIA PLATFORMS HAVE TIME UNTIL THEN TO INTRODUCE A LABELLING OPTION FOR AI-GENERATED OR AI-ALTERED CONTENT, IN ORDER TO INFORM THE USERS THAT THE POST THEY ARE SEEING IS NOT REAL.

CONCLUSION

WHILE SOCIAL MEDIA PLATFORMS HAVE ALREADY BEGUN INTRODUCING LABELS, THAT MARK AI-GENERATED CONTENT AS NOT REAL, THOSE LABELS ARE IN SOME CASES HARD TO FIND WHEN POSTING AND INCONSISTENTLY AVAILABLE ACROSS APPS AND WEBSITES. HOWEVER, SOCIAL MEDIA PLATFORMS AND THE EU AI ACT CURRENTLY CAN ONLY DO SO MUCH. THEREFOR IT IS IMPORTANT FOR USERS OF THESE PLATFORMS TO BE AWARE THAT THE POST THEY ARE SEEING MAY NOT BE REAL AND CAN ACTIVELY SPOT DEEPFAKES

IN ORDER TO SPOT A DEEPFAKE IN THE EXPANSE OF SOCIAL MEDIA, LOOK OUT FOR:

- UNNATURAL FACIAL EXPRESSIONS
- AWKWARD FACIAL-FEATURE POSITIONING
- A LACK OF EMOTION
- UNNATURAL BODY MOVEMENT
- INCONSISTENT HAIR OR TEETH
- INCONSISTENT OR UNNATURAL BACKGROUND
- BLURRING OR MISALIGNMENT OF FACIAL-FEATURES OR BODY PARTS
- INCONSISTENT AUDIO AND NOISE