

Какво са дийпфейковете и как да ги разпознаете?

Силвия Гаргова



Съдържание

- Произход и дефиниция
 - Кратка история
 - Видове deepfakes
 - Генериране
 - Разпознаване
 - Приложения
 - Примери
 - Cheapfakes
 - Игра
- 

Произход и дефиниция



От къде произлиза термина „deepfake“?

Произход:

DeepFake = Deep Learning + Fake

Неформална дефиниция:

Дийпфейк е снимка, видеоклип или аудиозапис, създаден с помощта на изкуствен интелект за реалистично симулиране или промяна на лицата, движенията и гласовете на хората.

Превод: Европейската комисия е вкарала на български език термините:

„дълбоко фалшифициране“

„дълбинен фалшификат“

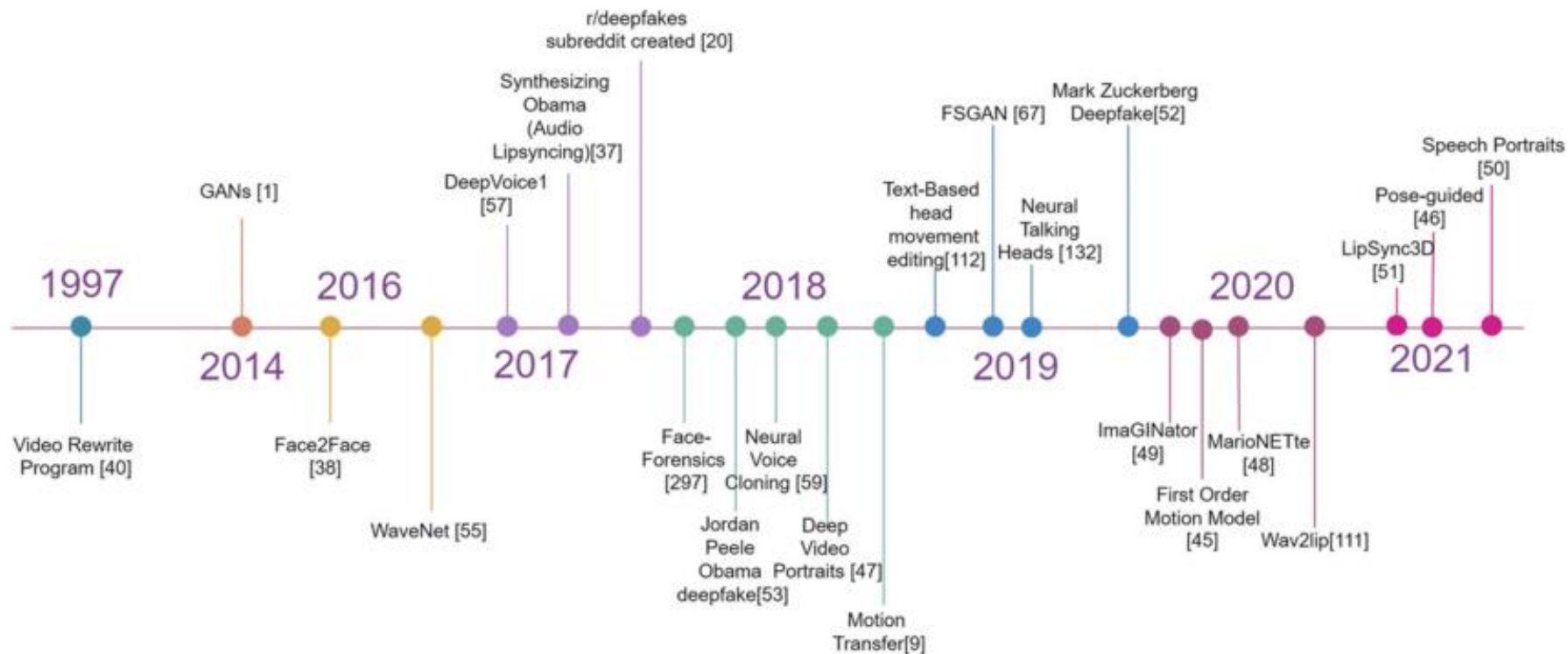
„дълбок фалшификат“



Кратка история



Кратка история на появата на дийпфейкове



Видове



Видове дийпфейкове

- изображения/видео
- аудио/глас
- текстови (използват се също термините readfakes, syntetic text)
- географски



Генериране



Как се създават дийпфейк изображения или видеа?

Техники за създаване на дийпфейк видеа:

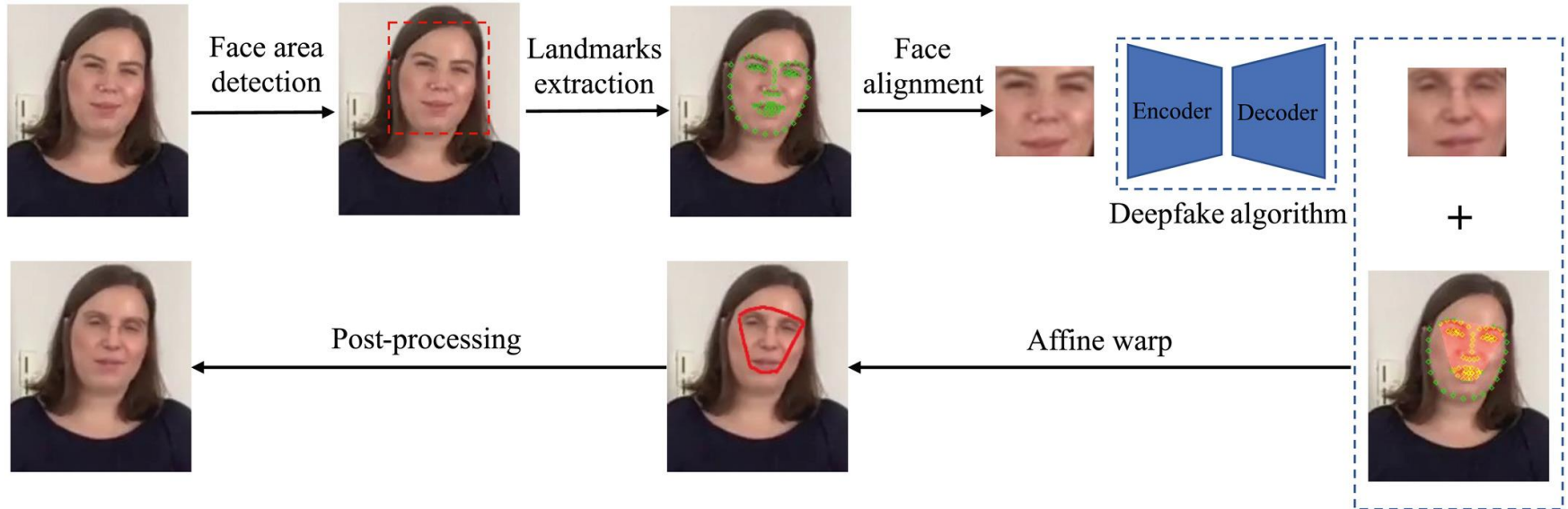
- размяна на лица (face swapping)
 - FakeApp
 - FaceShift
- реконструкция на лица (face reenactment/puppetry)
 - Face2face

Техники за създаване на дийпфейк изображения:

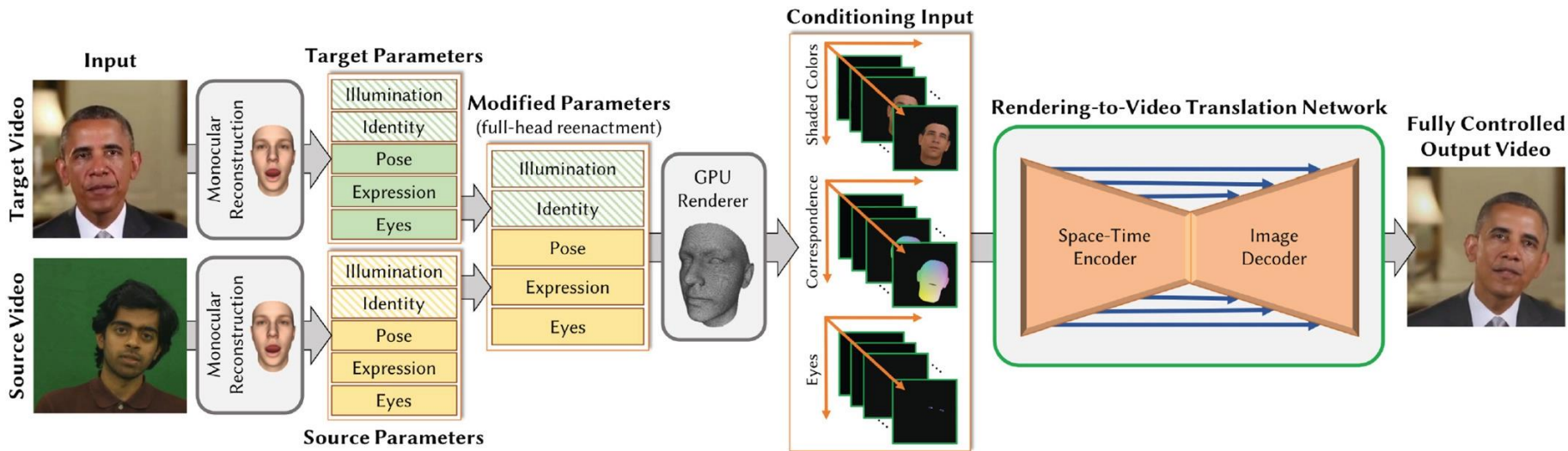
- Generative Adversarial Networks (GANs) – синтетични изображения



Размяна на лица (face swapping)

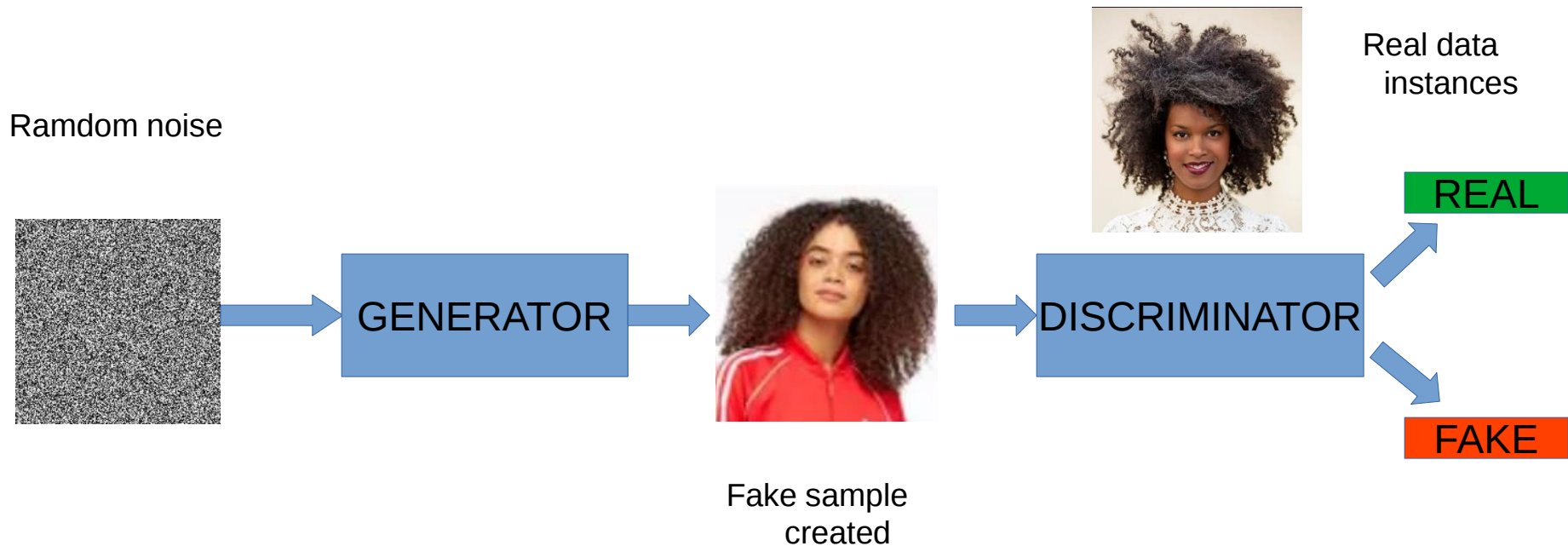


Реконструкция на лица (face reenactment/puppetry)



Kim, H., et al.: Deep video portraits. ACM Trans. Graph. 37(4), 1–14 (2018)

Generative Adversarial Networks (GANs)

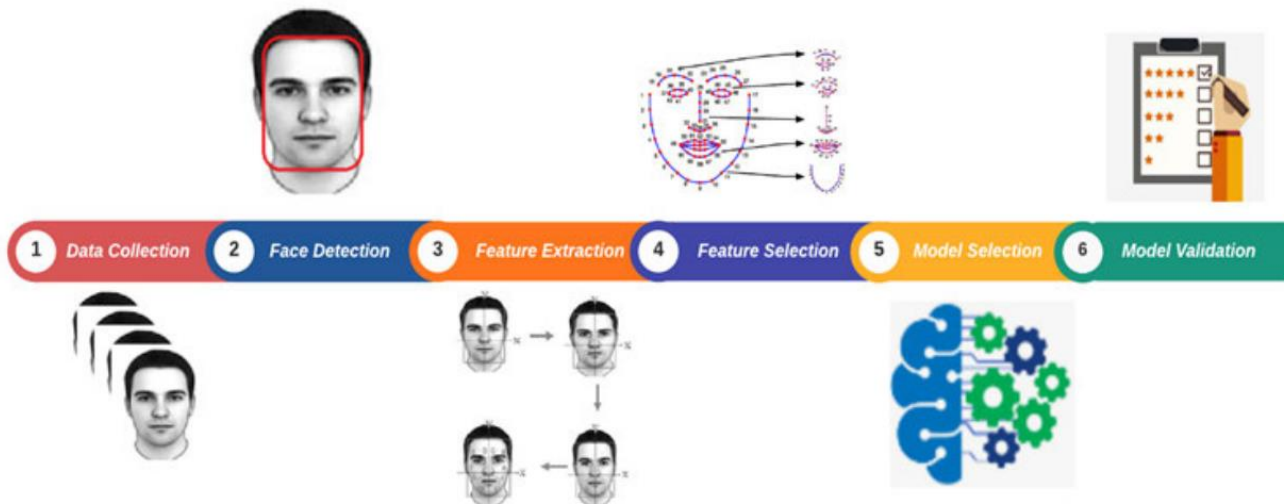


Разпознаване



Как се създават модели за разпознаване на deepfakes?

Процеса на разпознаване на deepfakes може да се обобщи в следните 6 стъпки:



Приложения



Положителни приложения

- кино индустрията
- образователни цели (създаване на многоезично съдържание)
- „възкресяване“ на исторически личности
- търговията (дава възможност да се видят продуктите без да се ходи във физически магазин)



Негативни аспекти

- създаване на дезинформация
- създаване на порнографско съдържание - през 2019 г. DeepTrace установява, че 96% от дипфейковете са видеа с порнографско съдържание.
- кражби на самоличност
- финансови измами

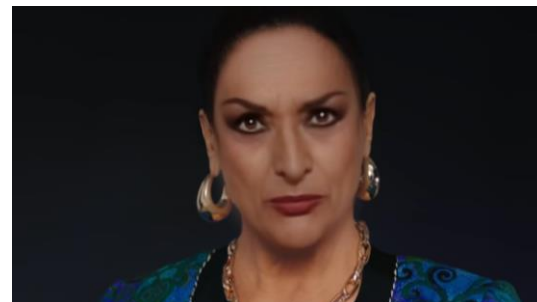
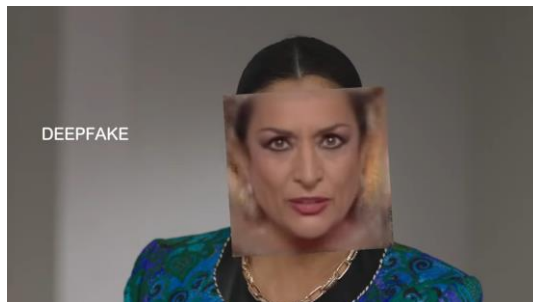


Примери



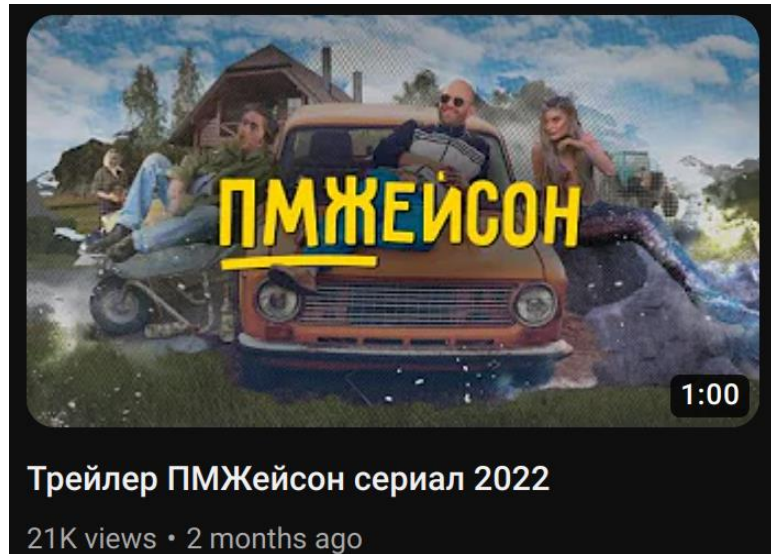
Дийпфейк технологии в рекламата

- Линк към рекламата: <https://www.youtube.com/watch?v=Yewm6TfLZ3Q>
- Как са създали дийпфейк образа: <https://www.youtube.com/watch?v=BQLTRMYHwvE&t=97s>



Първият дийпфейк сериал

https://www.youtube.com/@unreal_jason/videos



Географски дийпфейкс



Figure 4. Fake satellite images of a neighborhood in Tacoma with landscape features of other cities. (a) The original CartoDB basemap tile; (b) the corresponding satellite image tile. The fake satellite image in the visual patterns of (c) Seattle and (d) Beijing.

CheapFakes



TECHNOLOGIES

Recurrent Neural Network (RNN); Hidden Markov Models (HMM) and Long Short Term Memory Models (LSTM)

Generative Adversarial Networks (GANs)

Video Dialogue Replacement (VDR) model

FakeApp / After Effects

After Effects, Adobe Premiere Pro

Sony Vegas Pro

Free real-time filter applications

Free speed alteration applications

In-camera effects

Relabeling/ Reuse of extant video

DEEPPAKES

More expertise and technical resources required

Less expertise and fewer technical resources required

CHEAP FAKES

TECHNIQUES

Virtual performances (page 35)

Virtual performances

Voice synthesis (page 38)

Face swapping (page 35)

Lip-synching (page 38)

Face swapping: Rotoscope

Speeding and slowing (page 30)

Face altering/ swapping

Speeding and slowing

Lookalikes (page 27)

Recontextualizing (page 28)

EXAMPLES



Suwanjajorn et al. Face2Face: Synthesizing Obama



Posters and Howe's: Mark Zuckerberg

Deepfakes: Gal Gadot (not pictured because of image content)



Mario Klingemann: AI Art



Jordan Peele and BuzzFeed: Obama PSA



Paul Joseph Watson: Aosta Video

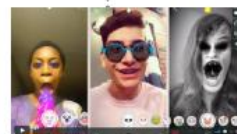


Bella Delphine: Hit or Miss Choreography

Rana Ayyub (not pictured because of image content)



Huw Parkinson: Uncivil War



SnapChat: Amsterdam Fashion Institute



Unknown: BBC NATO newscast

Игра



Кое изображение е истинско и кое дийпфейк?



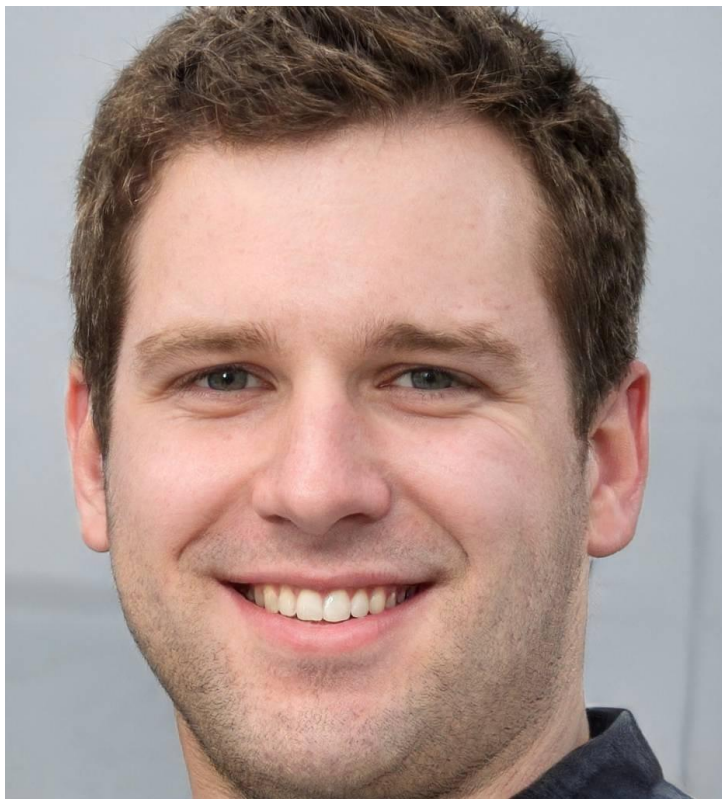
Дийпфейк



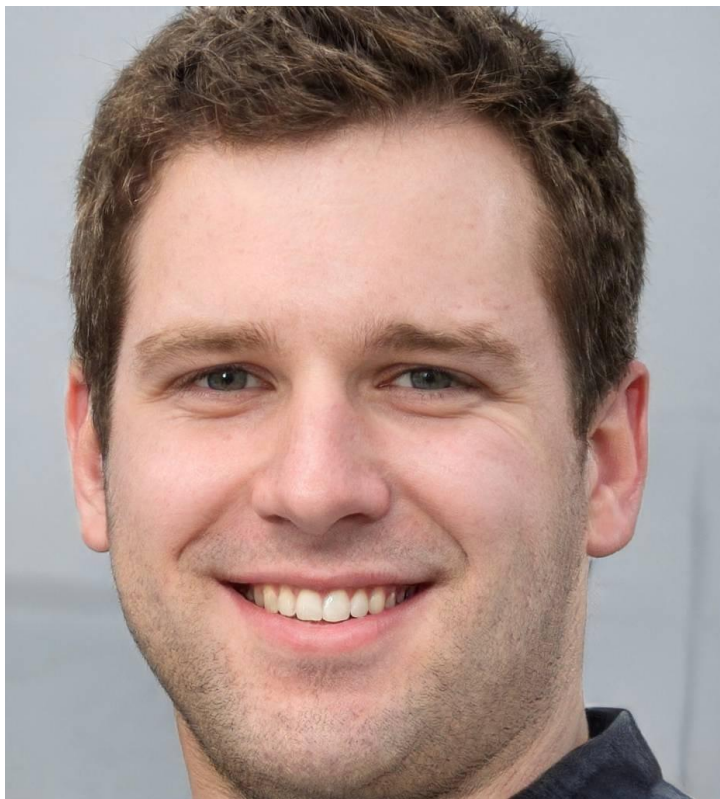
Истинско



Кое изображение е истинско и кое дийпфейк?



Дийпфейк



Истинско



Благодаря ви!

silvia.gargova@gate-ai.eu

Ако искате да получавате известия за предстоящи събития можете да се запишете в нашият мейлинг лист като сканиране QR кода:

