AI-Generated Content: Opportunities and Challenges

Abdenour Hadid
Professor
Sorbonne Center for Artificial Intelligence (SCAI)
Sorbonne University Abu Dhabi, UAE

SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

3 May 2024

SORBONNE UNIVERSITÉ ABU DHABI

IMPROVE 2024

AI-Generated Content: Opportunities and Challenges

1. Background
2. Introduction to AIGC
3. On the Detection of AIGC
4. Our Proposed Methods
5. Discussion

SCAI SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ ABU DHABI

AI-Generated Content: Opportunities and Challenges

1. Background
2. Introduction to AIGC
3. Detection of AIGC
4. Our Proposed Methods
5. Discussion

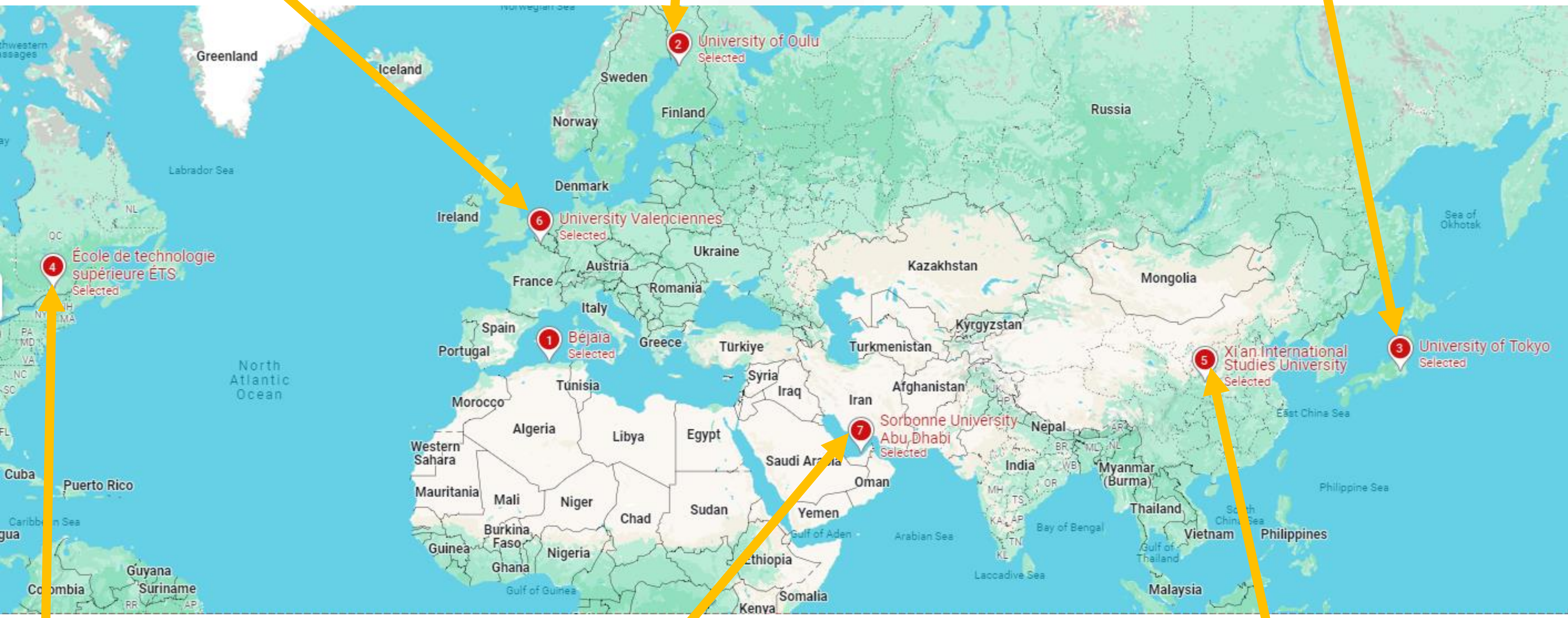
SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ ABU DHABI

Université Polytechnique
HAUTS-DE-FRANCE

UNIVERSITY OF OULU

東京工業大学
Tokyo Institute of Technology



ÉTS

SORBONNE UNIVERSITÉ ABU DHABI

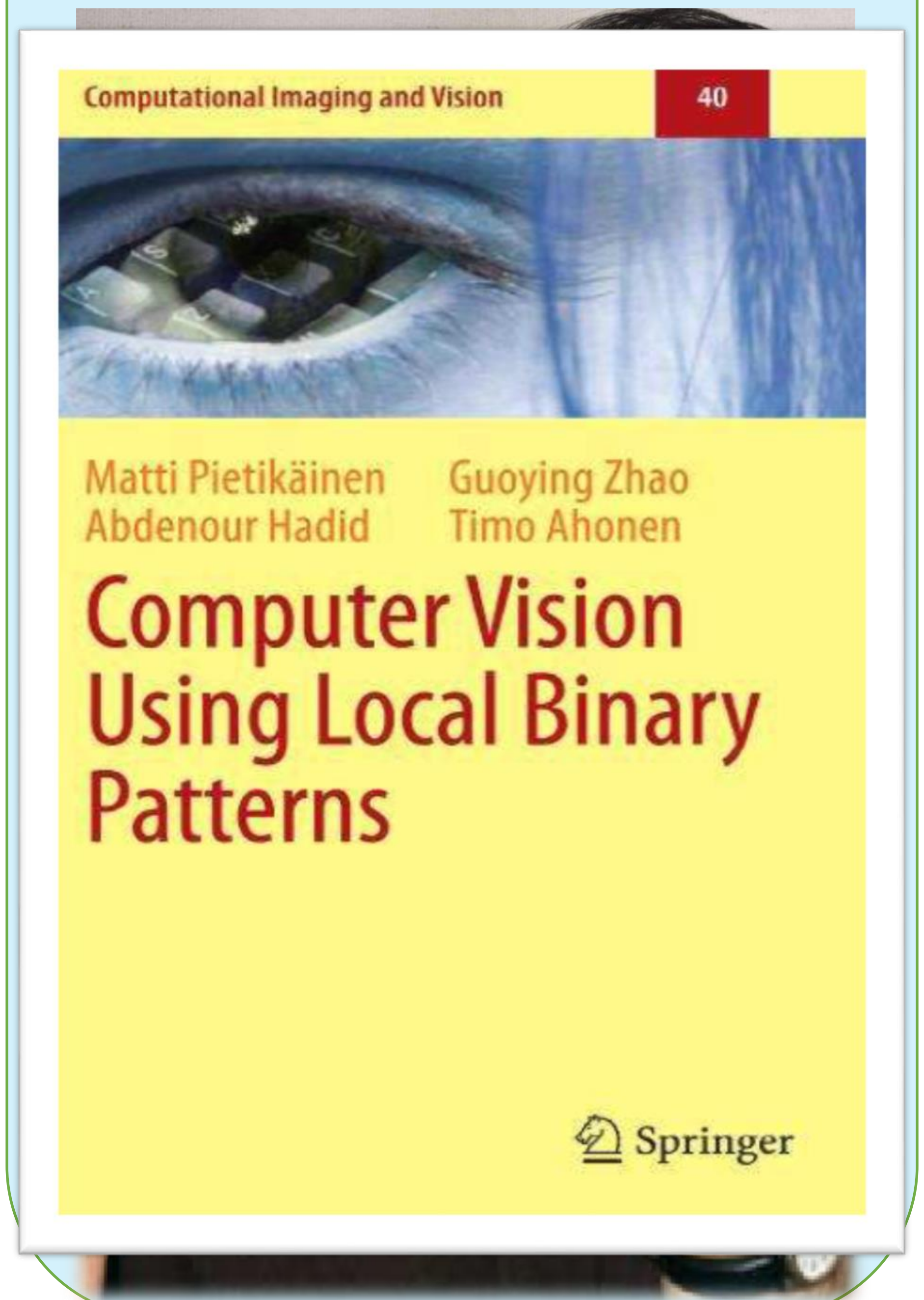
NORTHWESTERN POLYTECHNICAL UNIVERSITY
1938
西北工业大学

SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ ABU DHABI

AI in Security

Local Binary Patterns

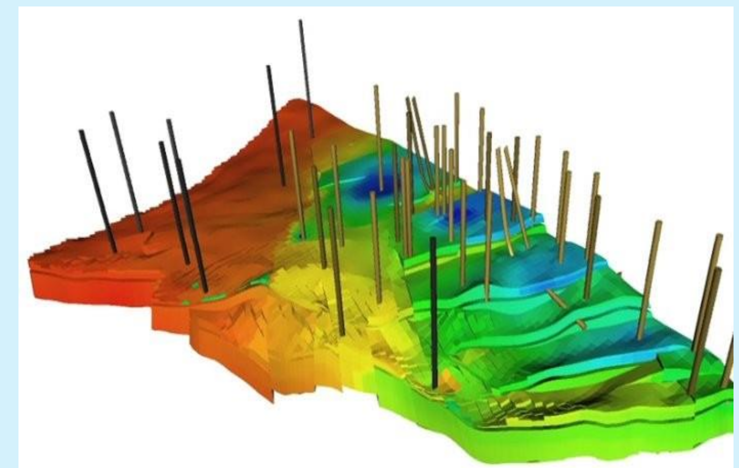


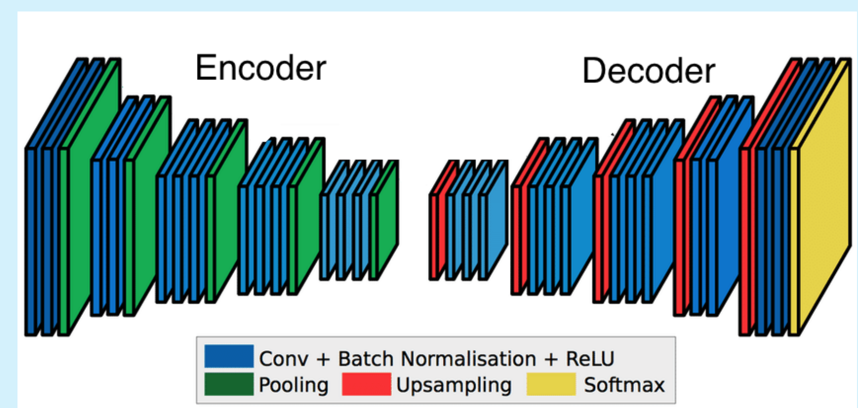
Key Publication:
A. Hadid, N. Evans, S. Marcel and J. Fierrez, "[Biometrics Systems Under Spoofing Attack: An evaluation methodology and lessons learned]()," in *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 20-30, Sept. 2015.

AI in Geoscience

Synthetic Data

TotalEnergies





Key Publication:
Abdenour Hadid, Tanujit Chakraborty, and Daniel Busby
"[When Geoscience Meets Generative AI and Large Language Models: Foundations, Trends, and Future Challenges]()"
ArXiv, 2024.

AI in Healthcare

Privacy?

Key Publication:
Abdenour Hadid,
[From Mind-Reading to Health-Reading Machines: Towards Contactless Health Diagnosis using Generative Artificial Intelligence,]
Nafath Journal, 2024.

AI in Mobiles

Lightweight?

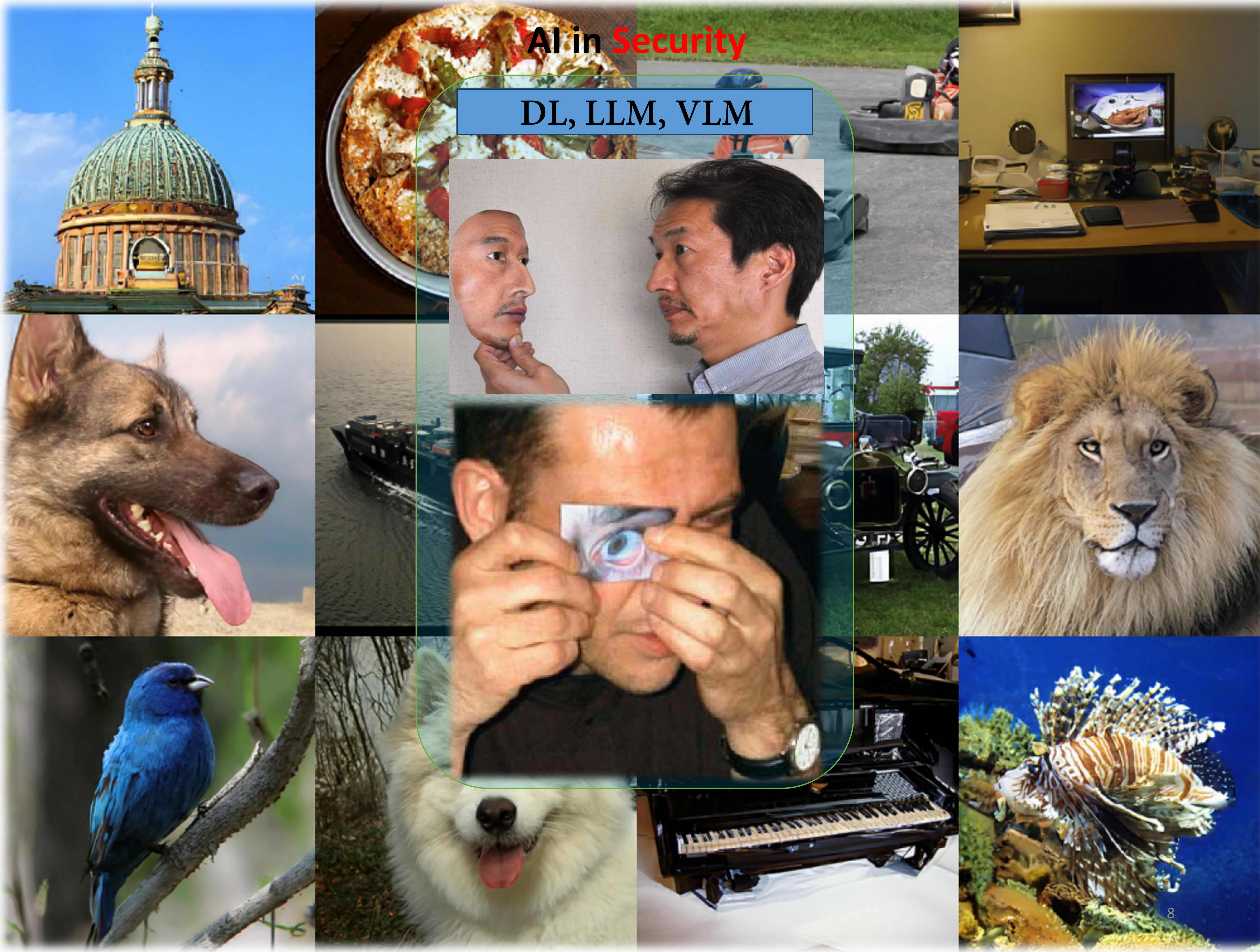
PATENTED TECHNOLOGY

AI in Autonomous Driving

Fusion?

Key Publication:
Abdenour Hadid *et al.*
[AI Powered Smart Roads for Future Smart Cities]
International Workshop on Information Technology & Communication for a Smart City - IWITIC'2021

AI in Security



Google Scholar

Abdenour Hadid

Full Professor, Sorbonne Center for Artificial Intelligence (SCAI)
Adresse e-mail validée de ieee.org

Artificial Intelligence   Computer Vision   Deep Learning

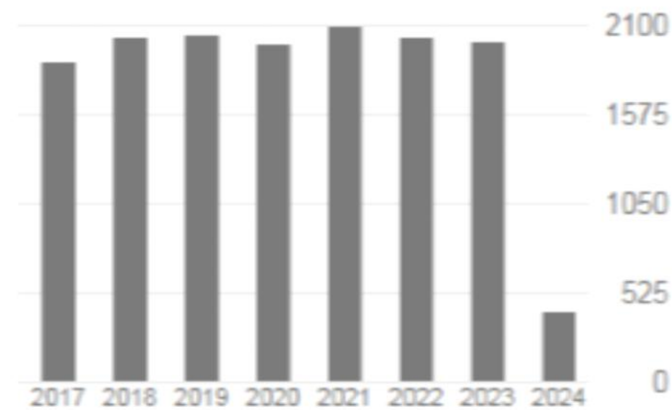
SUIVRE

OBTENIR MON PROPRE PROFIL

TITRE	CITÉE PAR	ANNÉE
Knowledge-Based Convolutional Neural Network for the Simulation and Prediction of Two-Phase Darcy Flows Z Elabid, D Busby, A Hadid ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and ...		2024
Harnessing the Power of Large Vision Language Models for Synthetic Image Detection M Keita, W Hamidouche, H Bougueffa, A Hadid, A Taleb-Ahmed arXiv preprint arXiv:2404.02726		2024
Bi-LORA: A Vision-Language Approach for Synthetic Image Detection M Keita, W Hamidouche, HB Eutamene, A Hadid, A Taleb-Ahmed arXiv preprint arXiv:2404.01959		2024
Generation and detection of manipulated multimodal audiovisual content: Advances, trends and open challenges H Liz-Lopez, M Keita, A Taleb-Ahmed, A Hadid, J Huertas-Tato, ... Information Fusion 103, 102103	2	2024
COVID-19 Infection Percentage Estimation from Computed Tomography Scans: Results and Insights from the International Per-COVID-19 Challenge F Bougourzi, C Distante, F Dornaika, A Taleb-Ahmed, A Hadid, ... Sensors 24 (5), 1557		2024

Citée par   TOUT AFFICHER

	Toutes	Depuis 2019
Citations	24218	10570
indice h	58	45
indice i10	143	112



Accès public   TOUT AFFICHER

12 articles   46 articles
non disponibles   disponibles

https://scholar.google.com/citations?user=Obhn_AkAAAAJ&hl=en

SCAI SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ ABU DHABI

On the Detection of AI-Generated Images and Videos

1. ~~Background~~
2. **Introduction to AIGC**
3. On the Detection of AIGC
4. Our Proposed Methods
5. Discussion

SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ ABU DHABI

Nowadays, it is extremely easy to generate text, images, videos, music - most recent models being based on Diffusion-like models.

Text-to-Text
- ChatGPT
- Bard
- LLaMa (Meta)
- PaLM 2
- Claude
- ...many more

Text-to-Image
- Midjourney
- DALL-E 3
- Stable Diffusion
- Muse
- Imagen
- Bard

Image-to-Text
- ChatGPT
- Flamingo
- Visualart

Image-to-3D
- Dream Fusion
- Magic3D

Image or Video-to-3D
- CSM AI

Text-to-Audio
- AutoLM
- Jukebox

Text-to-Code
- Codex
- Alphacode

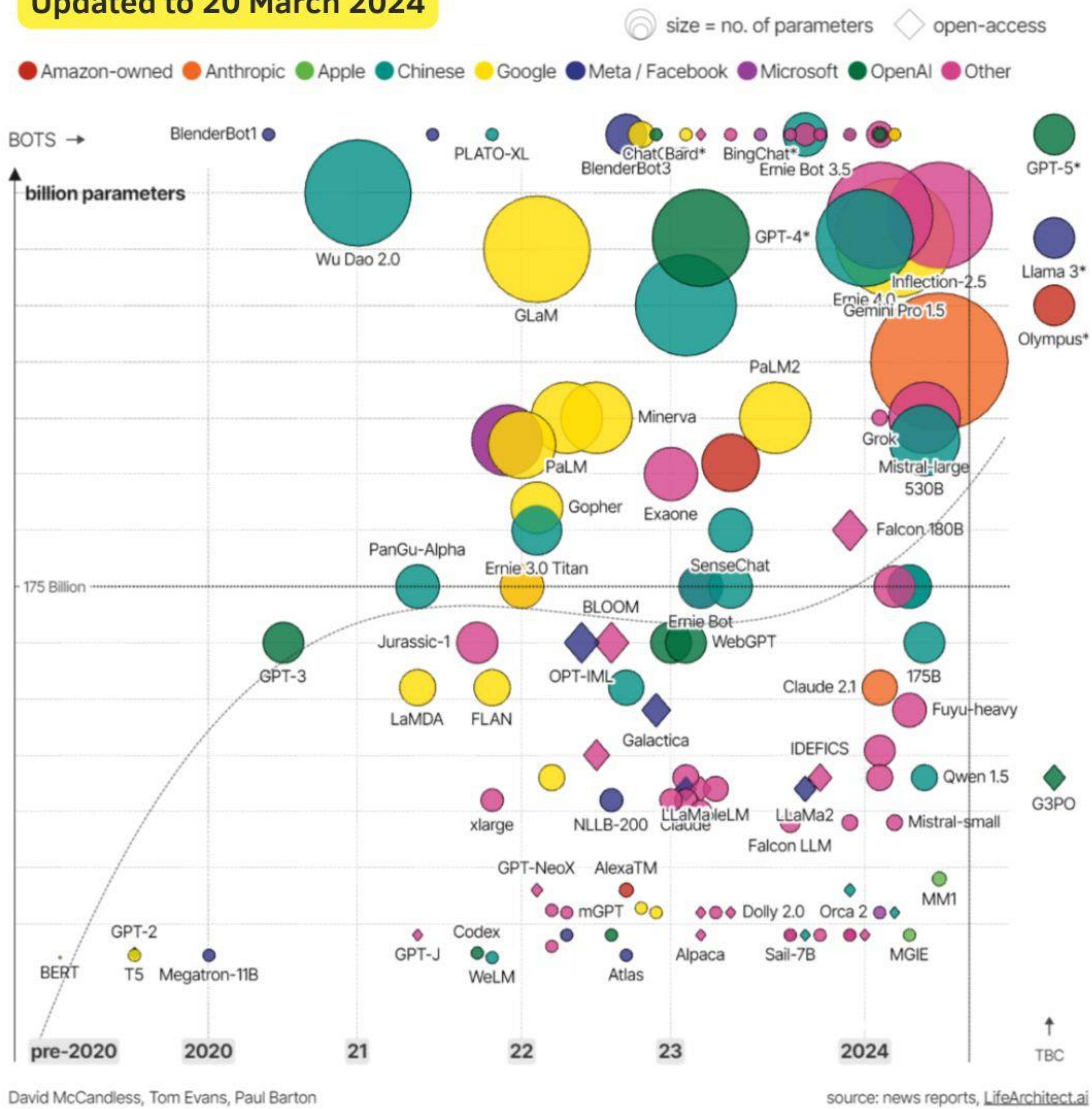
Image-to-Science
- Galatica
- Minerva

Text-to-video
- Runway
- Cuebric
- Phenaki

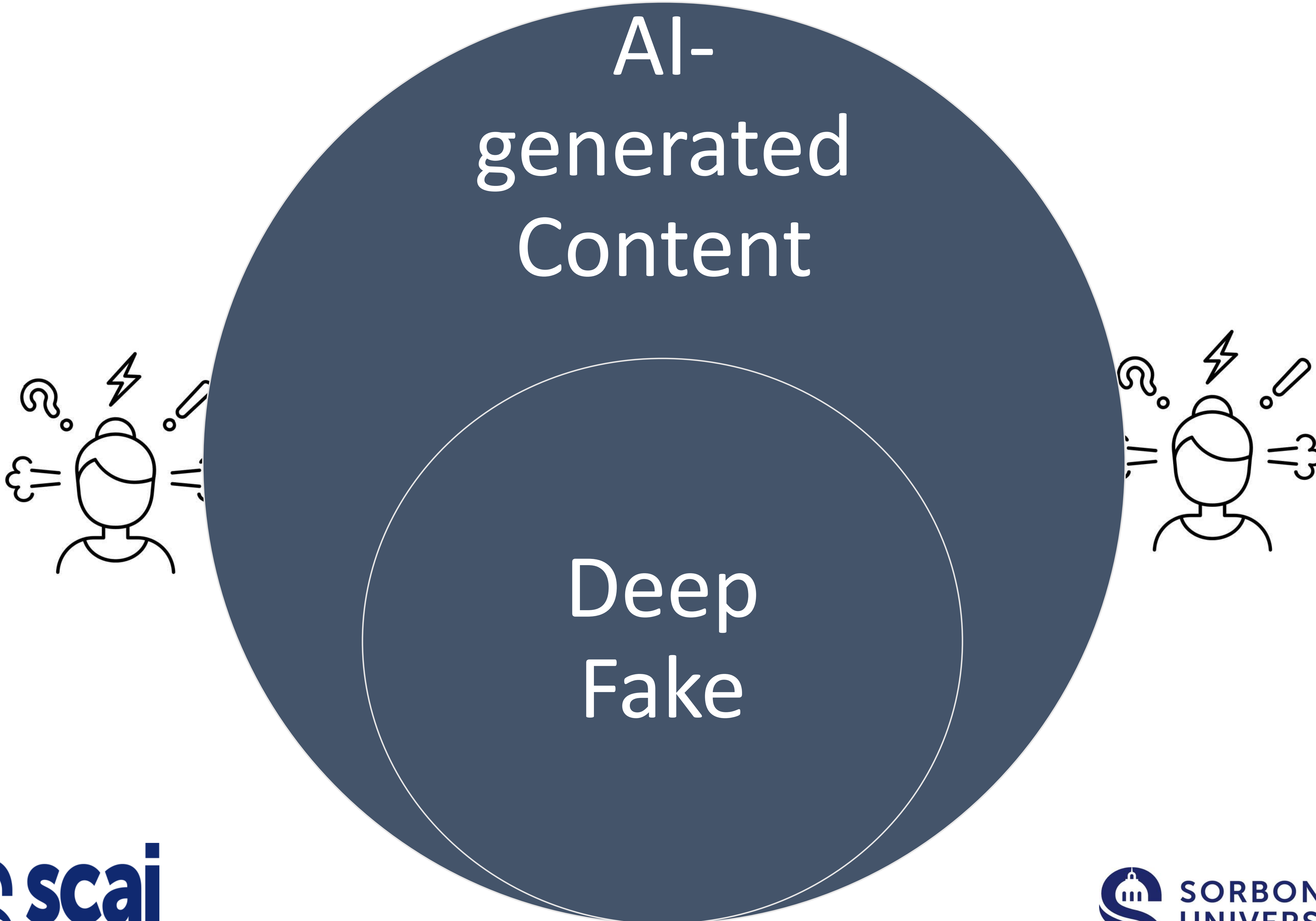
Audio-to-text
- Whisper

LLMs Landscape

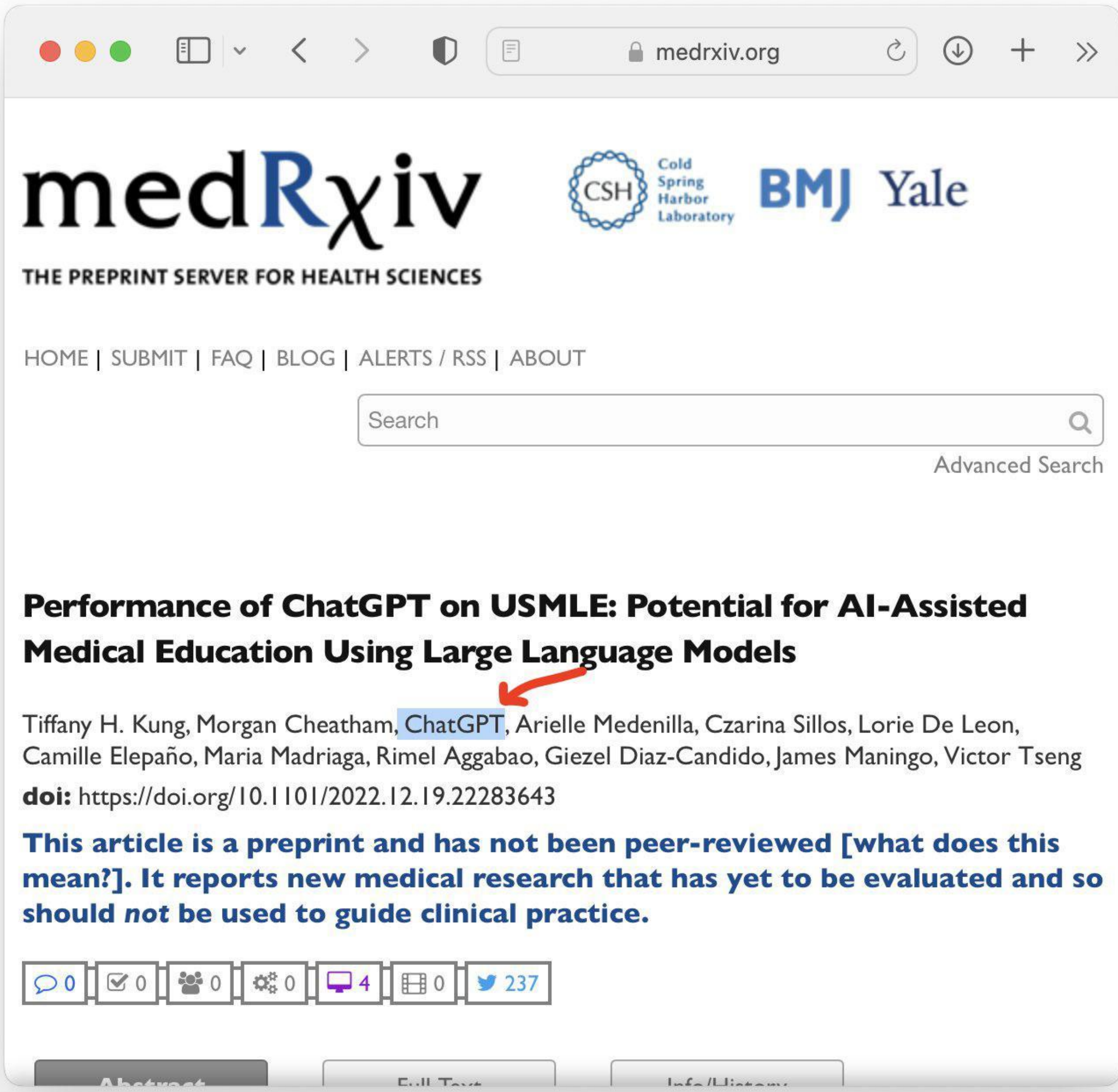
Updated to 20 March 2024



Creating highly good looking content (text, videos, images) with AI



Text Generation


medrxiv.org

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

HOME | SUBMIT | FAQ | BLOG | ALERTS / RSS | ABOUT

Search

Advanced Search

Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models

Tiffany H. Kung, Morgan Cheatham, ChatGPT, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, Victor Tseng

doi: https://doi.org/10.1101/2022.12.19.22283643

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

0 | 0 | 0 | 0 | 4 | 0 | 237


ELSEVIER

Nurse Education in Practice

Volume 66, January 2023, 103537

Editorial

Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?

Siobhan O'Connor [a] [1], ChatGPT [b]

Show more

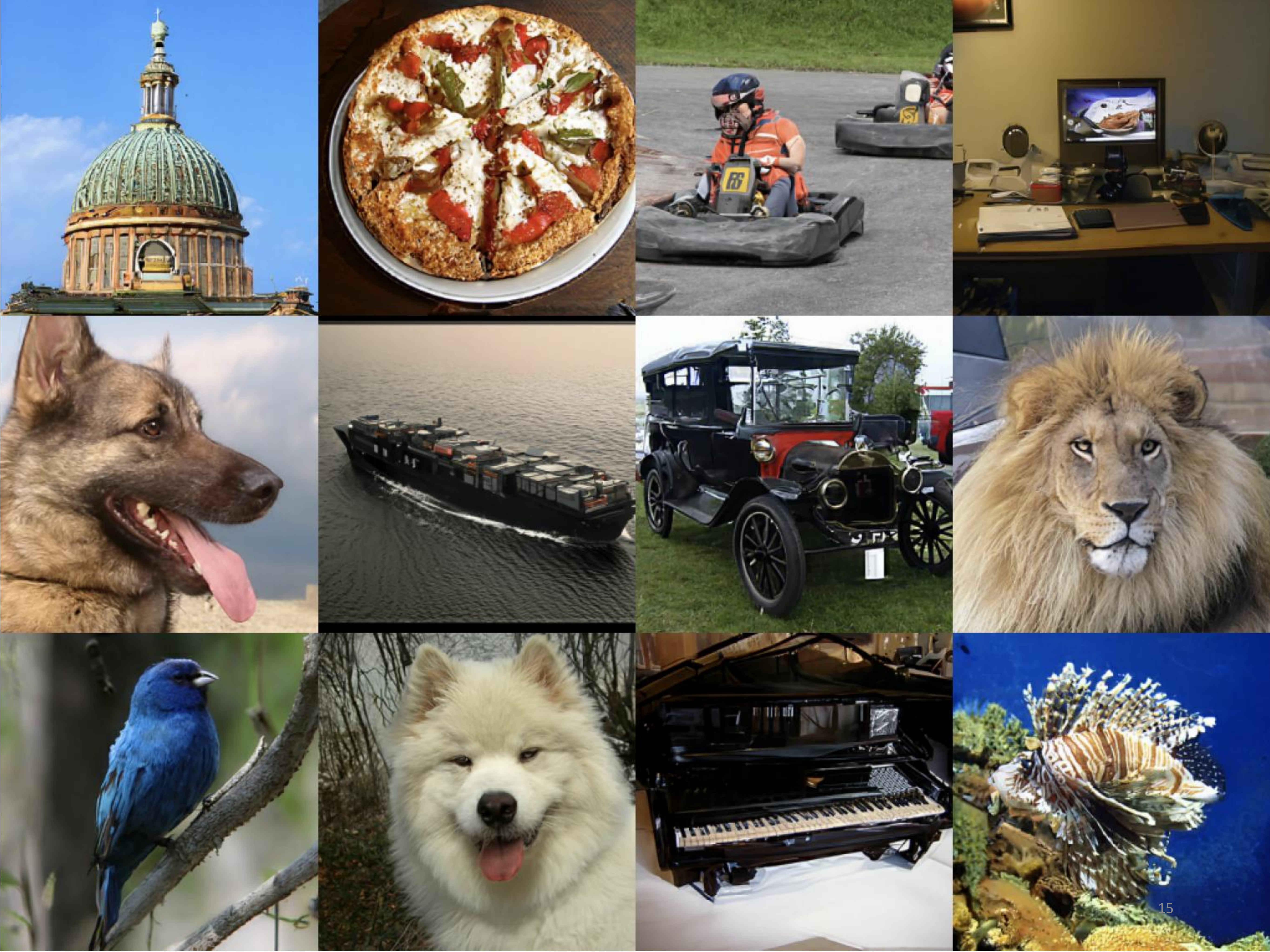
+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.nepr.2022.103537   Get rights and content

Section snippets

Funding

The author received no financial support for the research, authorship, and/or publication of this article....



EMO: Emote Portrait Alive - Generating Expressive Portrait Videos with Audio2Video Diffusion Model under Weak Conditions

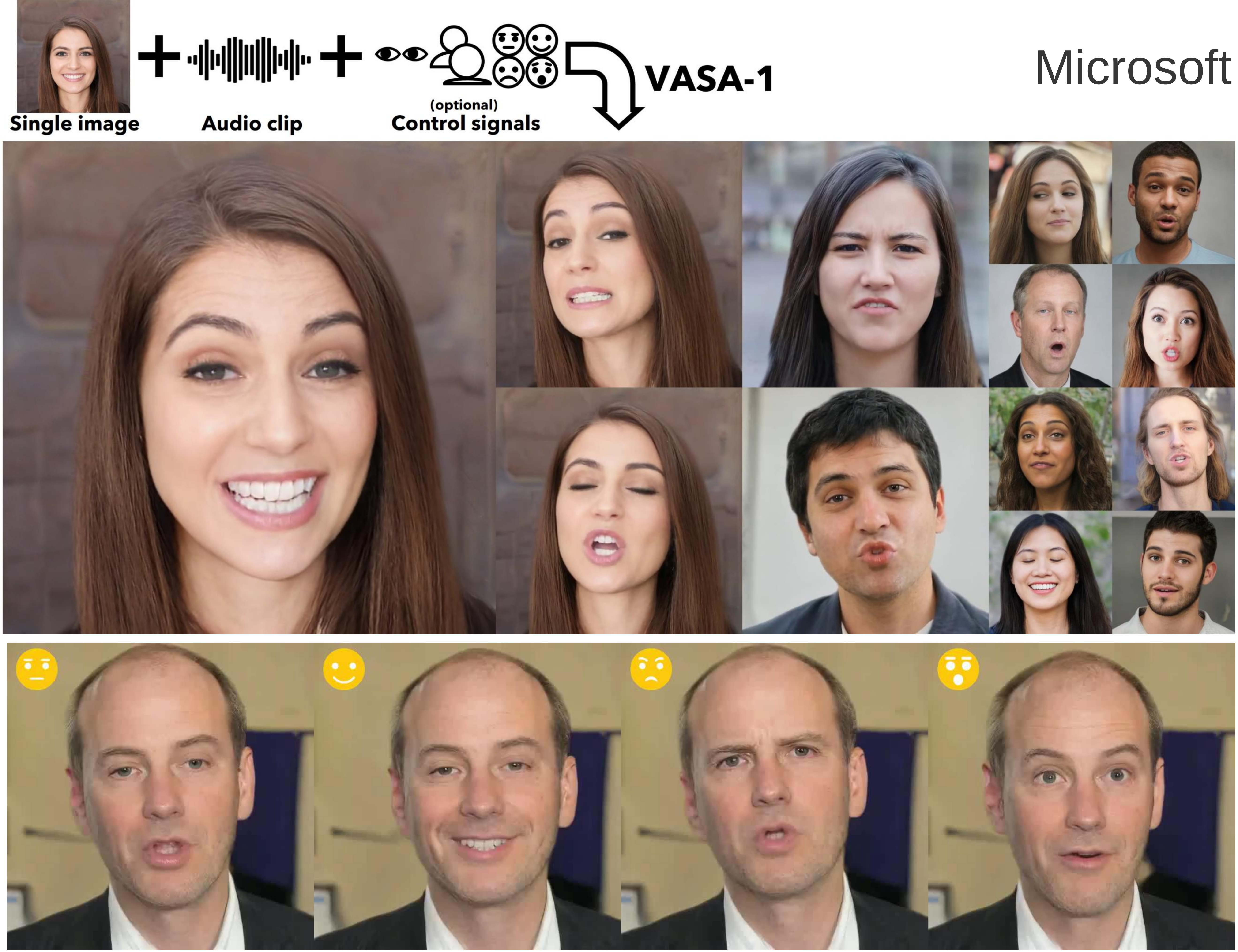
Linrui Tian, Qi Wang, Bang Zhang, Liefeng Bo
Institute for Intelligent Computing, Alibaba Group

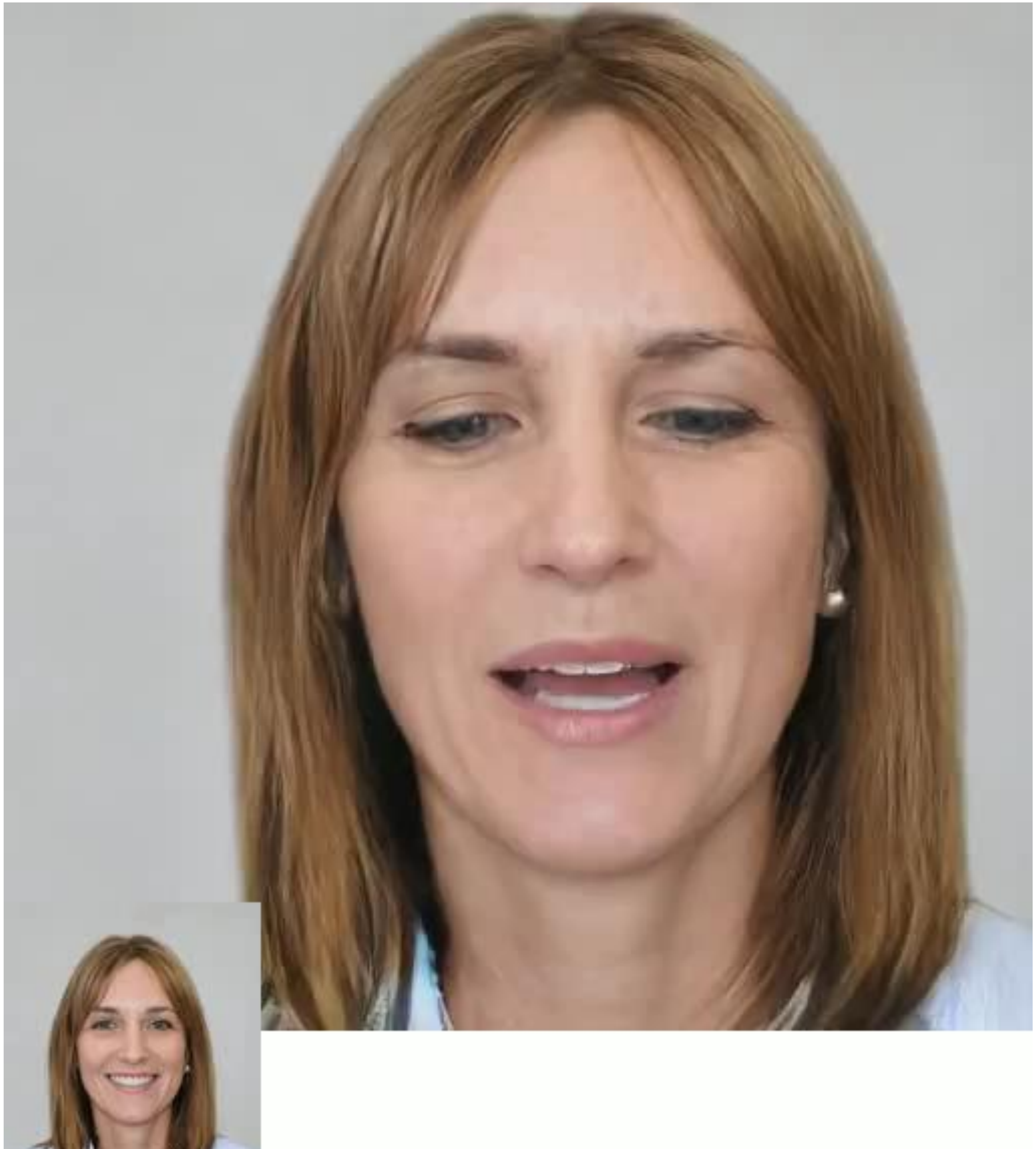
https://humanaigc.github.io/emote-portrait-alive/

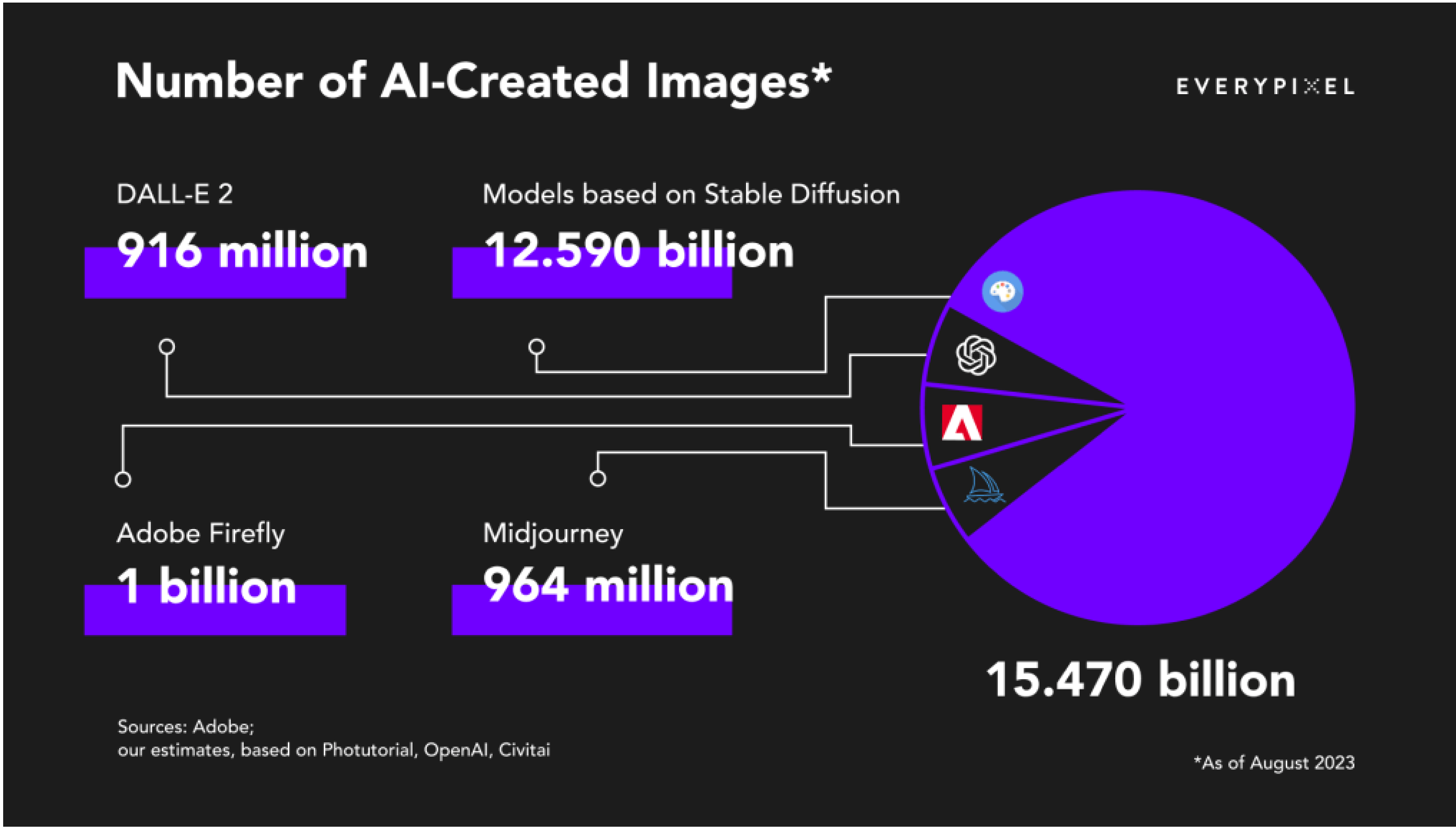
EMO: Emote Portrait Alive - Generating Expressive Portrait Videos with Audio2Video Diffusion Model under Weak Conditions

Linrui Tian, Qi Wang, Bang Zhang, Liefeng Bo
Institute for Intelligent Computing, Alibaba Group

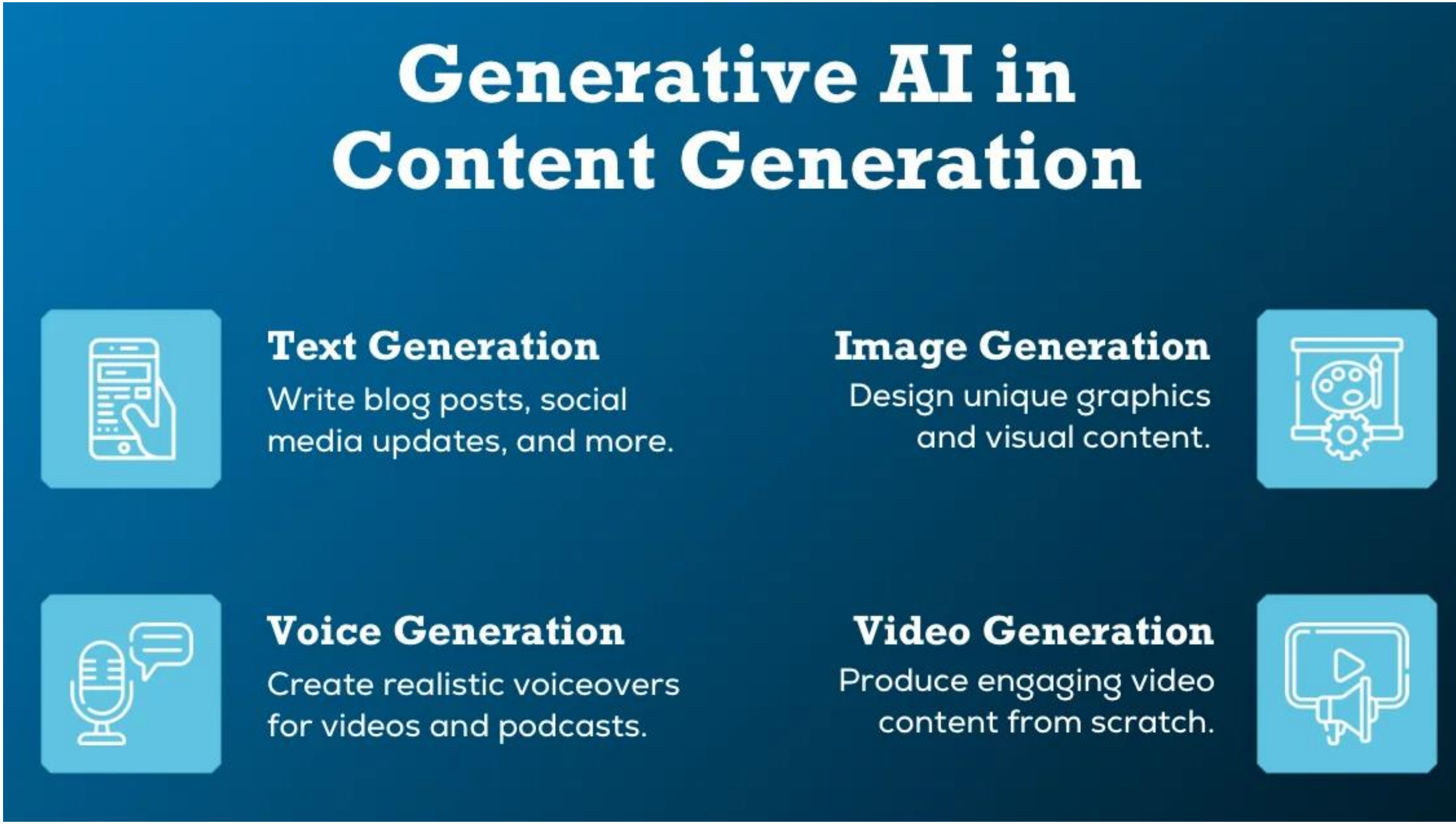
Microsoft






Number of AI-Created Images*
EVERYPIXEL
DALL-E 2
916 million
Models based on Stable Diffusion
12.590 billion
Adobe Firefly
1 billion
Midjourney
964 million
15.470 billion
Sources: Adobe;
our estimates, based on Photutorial, OpenAI, Civitai
*As of August 2023

Pros of AI-Generated Content



Biases and Hallucinations

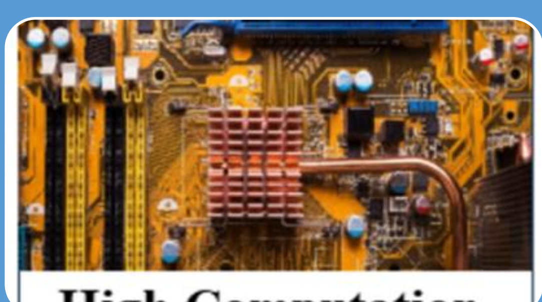
- Explainable models

Explainability &Trustworthiness

- Explainable Models

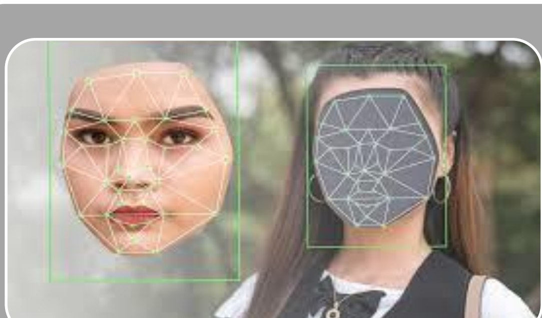
Generation Latency

- Data Efficient Modeling



High Computational Power

- AI and ML Accelerators



Impersonation, fraud and miss-use

- Robust detection models

AI-Generated Content: Opportunities and Challenges

1. ~~About my Background~~
2. ~~Introduction to AIGC~~
3. **On the Detection of AIGC**
4. Our Proposed Methods
5. Discussion

The Incident: A Deepfake Scam in Hong Kong


ONLINE
LIVE


FAKE


REAL


PESHAWAR
ZALMI
IMAGINED WITH AI


FAKE


FAKE


FAKE


FAKE


FAKE


FAKE


FAKE


REAL

Morgan Freeman?

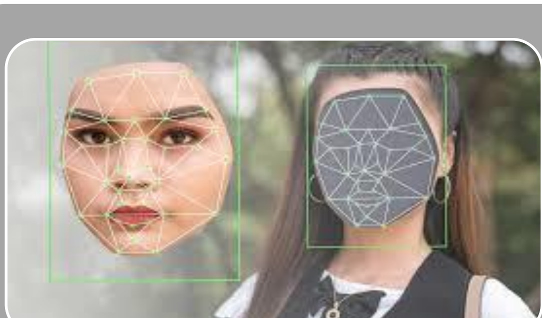
https://www.youtube.com/watch?v=oxXpB9pSETo

AI-Generated Content: Opportunities and Challenges

Abstract

It is becoming quite easy to generate realistic images and videos especially using diffusion-like models (e.g. DALL-E, GLIDE, Midjourney, Imagen, VideoPoet, Sora, Genie, etc.) due to their impressive generative capabilities. This creates a huge potential for a wide range of applications such as image editing, video production, content creation and digital marketing. Moreover, synthetically generated images and videos can be very useful for enhancing the training of AI models which usually require a large amount of data. However, these advances have also raised concerns about the potential misuse of these images and videos, including the creation of misleading content such as fake news and propaganda. So, one of the critical challenges associated with these advancements is the development of effective detection methods of synthetic images and videos. In the talk, we present the advances in automatically generating and editing images and videos, and discuss the limitations and challenges of such AI-generated content.

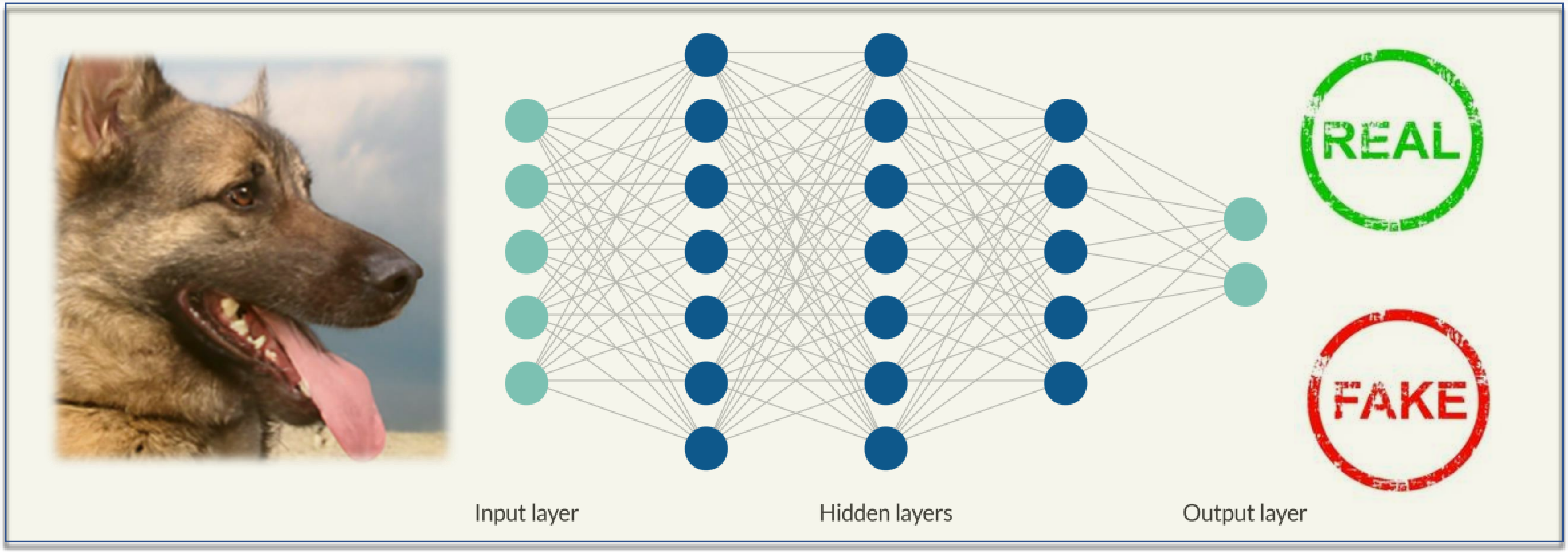
It is extremely easy to generate text, images, videos, music - most recent models being based on Diffusion-like models.



Impersonation, fraud and miss-use

- Robust detection models

The Detection of AI-Generated Content



Objective: develop a model $\mathcal{M}$ that learns

$$f : \mathbf{I} \rightarrow \mathbf{Y}$$

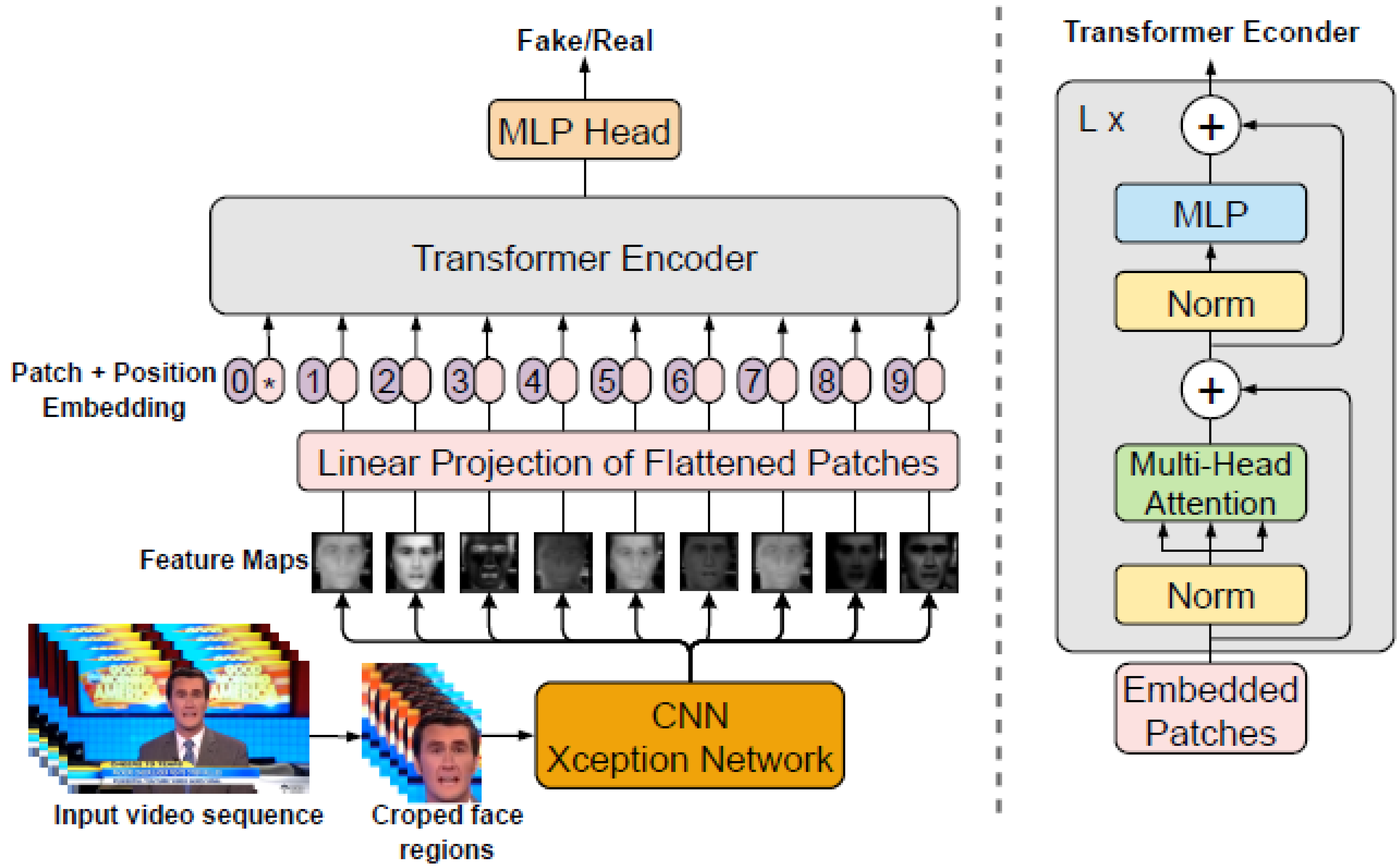
$$I_i \in \mathbf{I} = \mathbb{R}^{d \times d} \qquad y_i \in \mathbf{Y} = \{0, 1\}$$

$$D = \{(I_i, y_i) | 1 \leq i \leq n\}$$

$$\hat{y} = f_\theta(I)$$

Deepfake Detection Using Spatiotemporal Transformer

B Kaddar, SA Fezza, Z Akhtar, W Hamidouche, A Hadid, J Serra-Sagrist,
ACM Transactions on Multimedia Computing, Communications and Applications, 2023



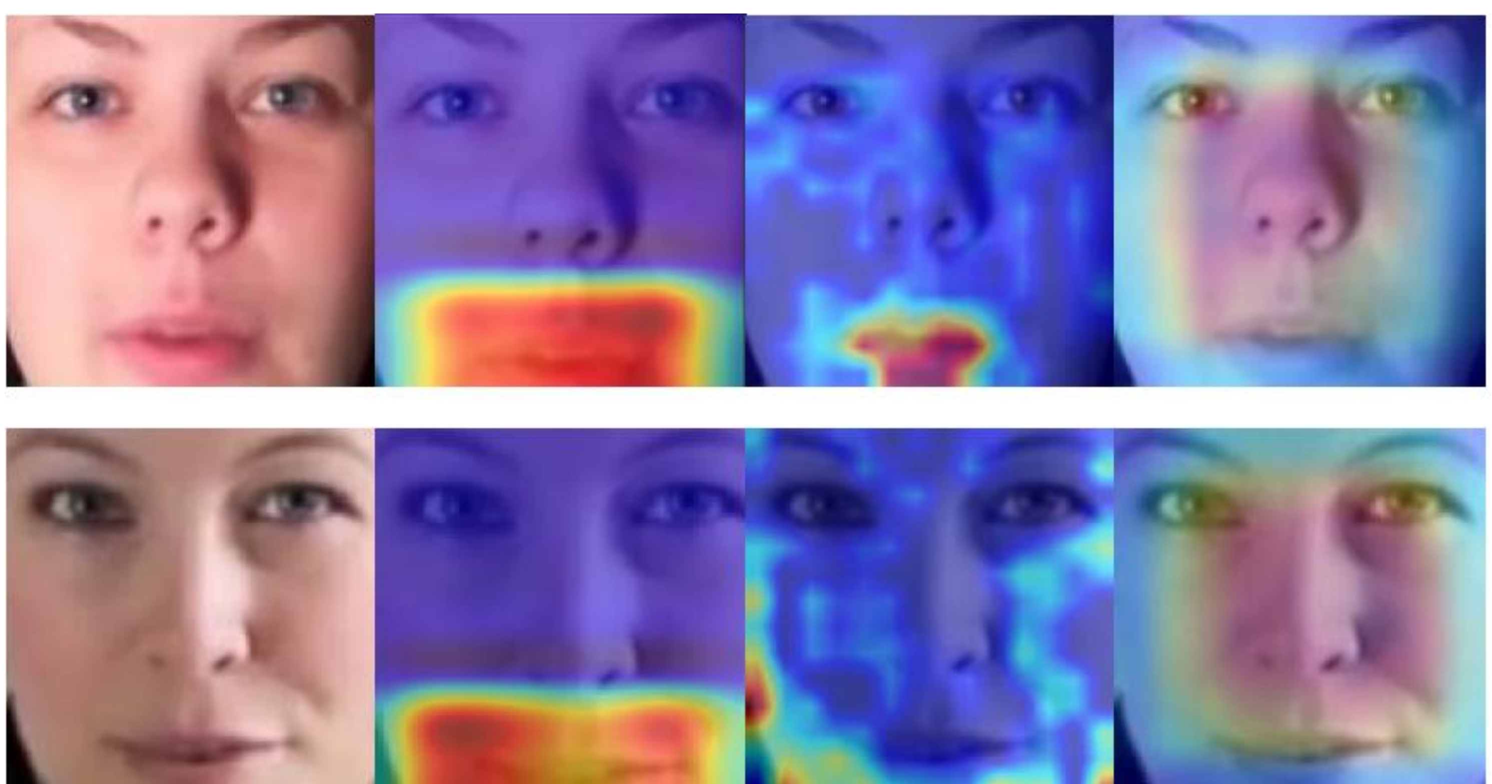
The proposed HCiT method for deepfake video detection. In HCiT (CNN-ViT), the features extracted by the CNN module are used as inputs by ViT for the binary classification task.

[Deepfake Detection Using Spatiotemporal Transformer]

B Kaddar, SA Fezza, Z Akhtar, W Hamidouche, A Hadid, J Serra-Sagrist,

ACM Transactions on Multimedia Computing, Communications and Applications, 2023

Method	Training on				Testing on	
	DF	FS	F2F	NT	DFDC-p	Celeb
Xception	✓				55.01	53.75
		✓			37.74	14.80
			✓		16.31	21.39
				✓	35.52	37.88
ViT	✓				39.96	49.23
		✓			21.71	17.63
			✓		25.67	19.47
				✓	53.76	65.10
HCiT	✓				57.02	54.03
		✓			48.45	21.47
			✓		28.33	24.26
				✓	55.04	68.81



(a) Fake image (b) Xception (c) ViT (d) HCiT

Deep Fake (DF)
FaceSwap (FS)
Face2Face (F2F)
NeuralTexture (NT)

Architectures and features of **state-of-the-art AI-synthesized images detection models**

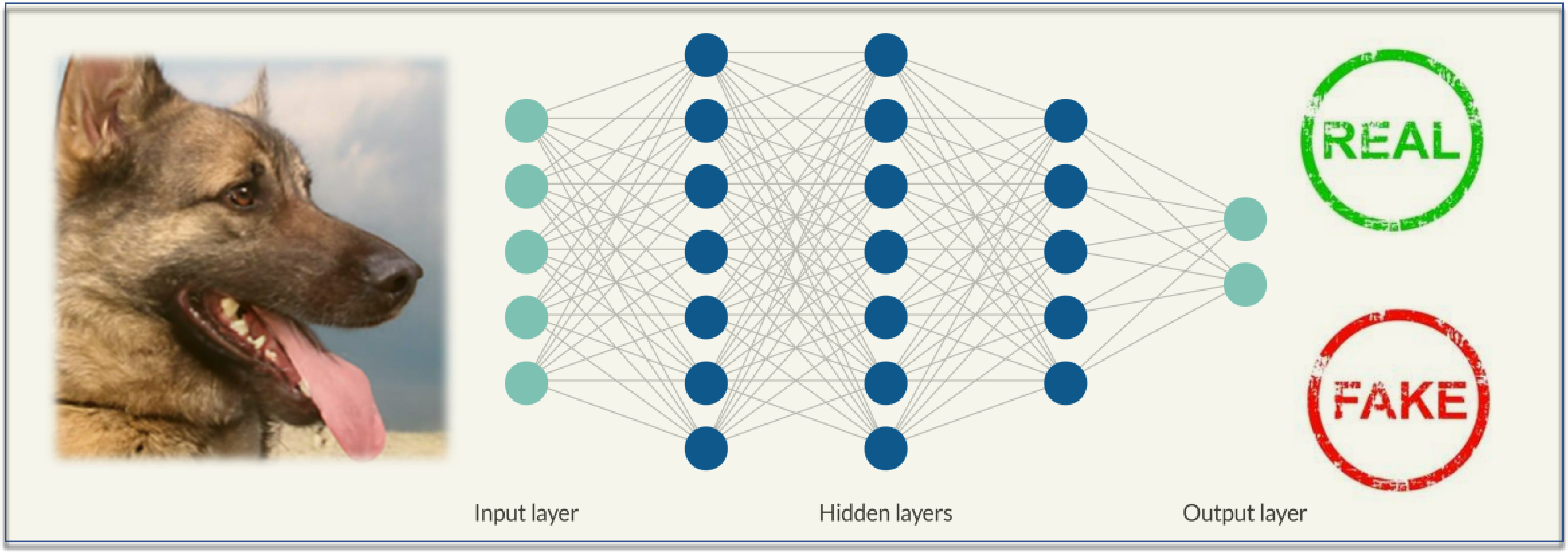
Authors	Architecture	Task
Sha *et al.*♣ (2022) [58]	CNN, CLIP + MLP	binary classification
Coccomini *et al.*♣ (2023) [60]	CNN, CLIP + MLP	binary classification
Guarnera *et al.*♣ (2023) [61]	CNN	binary classification
Amoroso *et al.*♣ (2023) [62]	CNN	binary classification
Wu *et al.*♣ (2023) [63]	CLIP + MLP	binary classification
Xi *et al.*♣ (2023) [64]	CNN	binary classification
Lorenz *et al.*♣ (2023) [65]	CNN + RF	binary classification
Ju *et al.*♣ (2023) [66]	CNN	binary classification
Sinitsa *et al.*♣ (2023) [67]	rule-based method	binary classification
Guo *et al.*♣ (2023) [68]	CNN	binary classification
Cozzolino *et al.*♣ (2023) [69]	CLIP + SVM	binary classification
Wang *et al.*• (2023) [70]	CNN	binary classification
Ma *et al.*• (2023) [71]	statistical-based approach, CNN	binary classification

♣ Deep feature. • Diffusion models unique attribute.

The datasets used in the evaluations of state-of-the-art AI-synthesized images detection models

Authors	Dataset
Sha *et al.*♣ (2022) [58]	DE-FAKE
Coccomini *et al.*♣ (2023) [60]	Diffusers
Guarnera *et al.*♣ (2023) [61]	Level up the DeepFake detection
Amoroso *et al.*♣ (2023) [62]	COCOFake
Wu *et al.*♣ (2023) [63]	LSUN, Danbooru, ProGAN, SD, BigGAN, GauGAN, styleGAN, DALLE, GLIDE, Guided Diffusion, Latent Diffusion, ImageNet, VISION, Artist, DreamBooth, Midjourney, NightCafe, StableAI, YiJian
Xi *et al.*♣ (2023) [64]	AI-Gen Image
Lorenz *et al.*♣ (2023) [65]	CiFAKE, ArtiFact, DIffusionDB, LAION-5B, SAC, SD-v2.1, LSUN-Bedroom
Ju *et al.*♣ (2023) [66]	LSUN, ProGAN, DF^3
Sinitsa *et al.*♣ (2023) [67]	Laion-5B, SD v-1.4, SD v-2.1, DALL-E-Mini, GLIDE [20], DALL-E-2, MidJourney, CycleGAN, ProGAN_e, ProGAN_t, BigGAN, StyleGAN, StyleGAN2, GauGAN, StarGAN
Guo *et al.*♣ (2023) [68]	HiFi-IFDL
Cozzolino *et al.*♣ (2023) [69]	ProGAN, StyleGAN2, StyleGAN3, StyleGAN-T, GigaGAN, (Score-SDE, ADM, GLIDE, eDiff-I, Latent and Stable Diffusion, DiT, DeepFloyd-IF, Stable Diffusion XL, DALL-E 2, DALL-E 3, Midjourney v5, Adobe Firefly, LSUN, FFHQ, ImageNet, COCO, LAION, RAISE
Wang *et al.*• (2023) [70]	DiffusionForensics
Ma *et al.*• (2023) [71]	CIFAR10, TinyImageNet, CelebA

The Detection of AI-Generated Content



Poor Generalization to Diffusion Models

Approaches based on binary classification are shown a serious lack of efficiency in generalizing to accurately discriminate new diffusion-generated images that have never been encountered before.

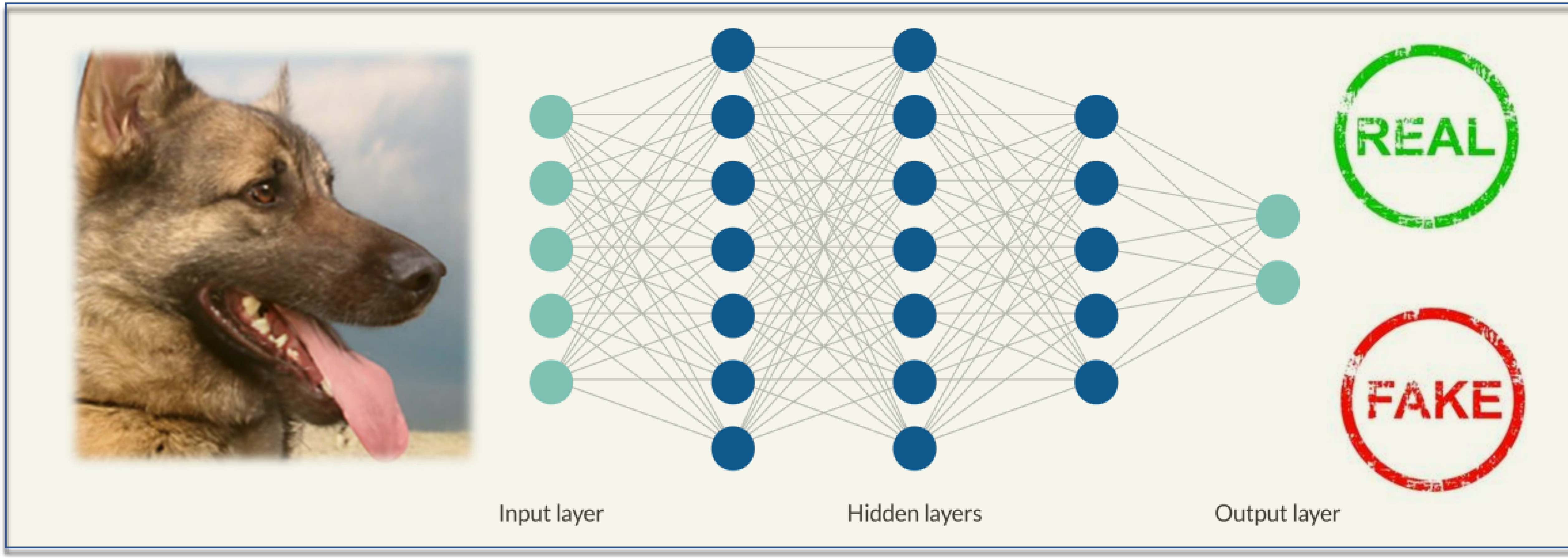
J. Ricker, S. Damm, T. Holz, and A. Fischer, "Towards the detection of diffusion model deepfakes," arXiv preprint arXiv:2210.14571, 2022. https://arxiv.org/pdf/2210.14571.pdf

AI-Generated Content: Opportunities and Challenges

1. ~~About my Background~~
2. ~~Introduction to AIGC~~
3. ~~Detection of AIGC: SOTA~~
4. **Our Proposed Methods**
5. Discussion

SCAI
SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ ABU DHABI



Frame deepfake detection as a image captioning task:

I

$\mathcal{M}$ VLM with parameters θ

$\hat{C} = \mathcal{M}_\theta(I)$

$\hat{C}$ **This image is...**

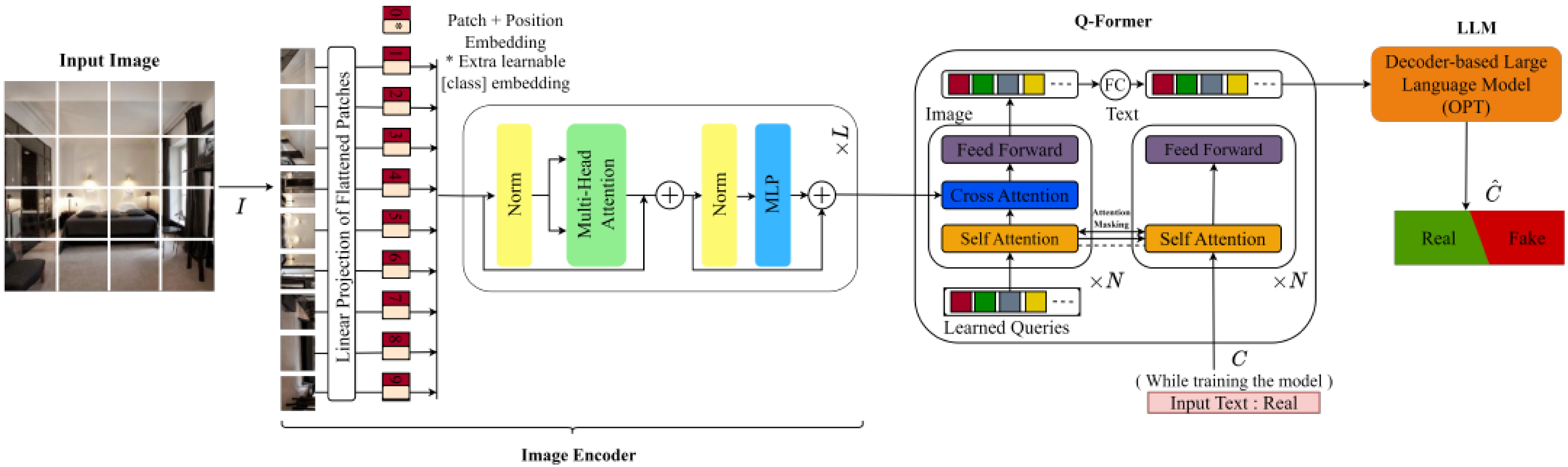
minimize a suitable loss function $\mathcal{L}$

$$\theta^* = \arg\min_\theta \sum_{i=1}^{n} \mathcal{L}(\mathcal{M}_\theta(I_i), C_i)$$

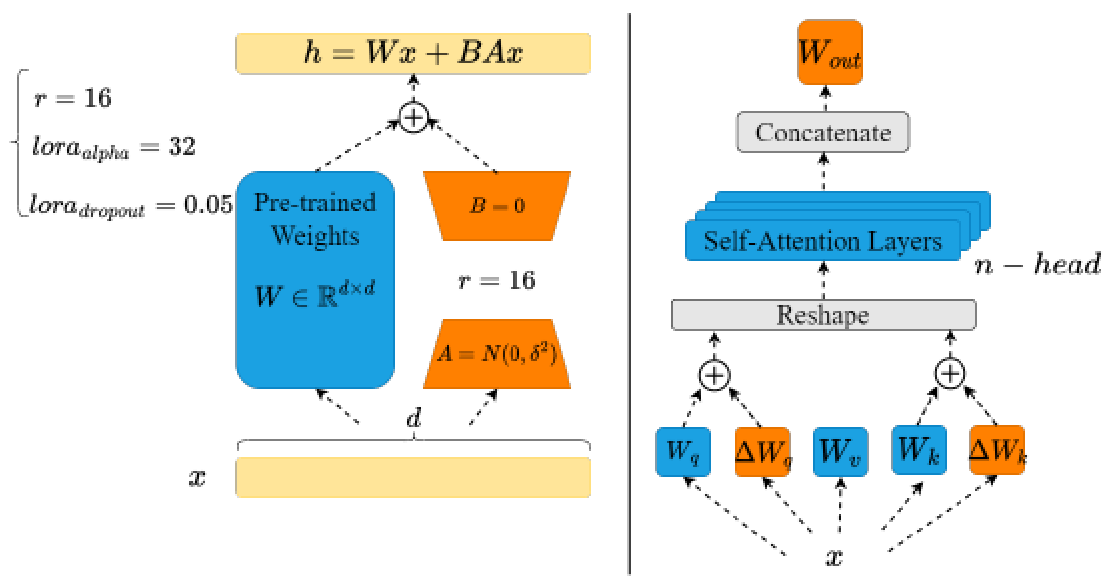
Real | Fake

$$D = \{(I_i, C_i) | 1 \leq i \leq n\}$$

[Bi-LORA: A Vision-Language Approach for Synthetic Image Detection](https://arxiv.org/abs/2404.01959) M Keita, W Hamidouche, HB Eutamene, A Hadid, A Taleb-Ahmed, https://arxiv.org/abs/2404.01959



Proposed methodology for synthetic image detection based on the BLIP-2 architecture.



BLIP-2 fine-tuning with LoRA.

Bi-LORA: A Vision-Language Approach for Synthetic Image Detection M Keita, W Hamidouche, HB Eutamene, A Hadid, A Taleb-Ahmed, https://arxiv.org/abs/2404.01959

Table 1: Results of different methods trained on LDM and evaluated on different testing subsets. We report ACC (%) / F1-Score (%).

Method	Testing Subset							Avg (%)
	LDM*	ADM⊕	DDPM⊕	IDDPM⊕	PNDM⊕	SD v1.4*	GLIDE*	
ResNet50	99.92 / 99.92	72.33 / 61.83	75.26 / 67.21	88.96 / 87.61	77.20 / 70.52	75.47 / 67.57	73.10 / 63.28	80.32 / 73.99
Xception	**99.96 / 99.96**	52.05 / 7.98	58.60 / 29.41	54.62 / 16.99	60.01 / 33.43	63.84 / 43.41	58.92 / 30.35	64.00 / 37.36
DeiT	99.83 / 99.83	50.40 / 2.01	50.18 / 1.17	50.14 / 1.01	56.25 / 22.54	96.02 / 95.86	98.15 / 98.11	71.56 / 45.79
ViTGPT2	99.40 / 99.40	70.84 / 59.21	69.60 / 56.72	84.08 / 81.20	95.40 / 95.22	**99.54 / 99.55**	**99.27 / 99.27**	88.30 / 84.37
BLIP-2	99.12 / 99.13	**85.24 / 82.97**	**98.47 / 98.47**	**97.02 / 96.97**	**99.22 / 99.23**	77.68 / 71.79	97.09 / 97.05	**93.41 / 92.23**

* Text-To-Image diffusion-based model. ⊕ Unconditional diffusion-based model.

Method	Avg (%)
ResNet50	80.32 / 73.99
Xception	64.00 / 37.36
DeiT	71.56 / 45.79
ViTGPT2	88.30 / 84.37
BLIP-2	**93.41 / 92.23**

Bi-LORA: A Vision-Language Approach for Synthetic Image Detection M Keita, W Hamidouche, HB Eutamene, A Hadid, A Taleb-Ahmed, https://arxiv.org/abs/2404.01959


(a) Real - 0|0|0|0|0


(b) Fake (ADM) - 0|0|0|0|1

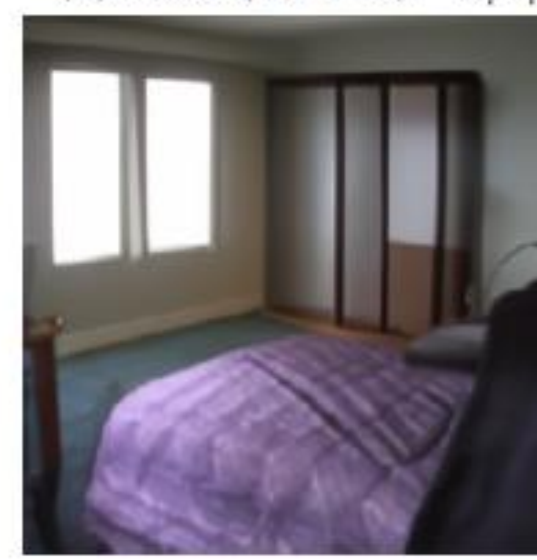

(c) Fake (LDM) - 1|1|1|1|1


(d) Fake (DDPM) - 0|0|0|1|1


(e) Fake (IDDPM) - 1|0|0|0|1

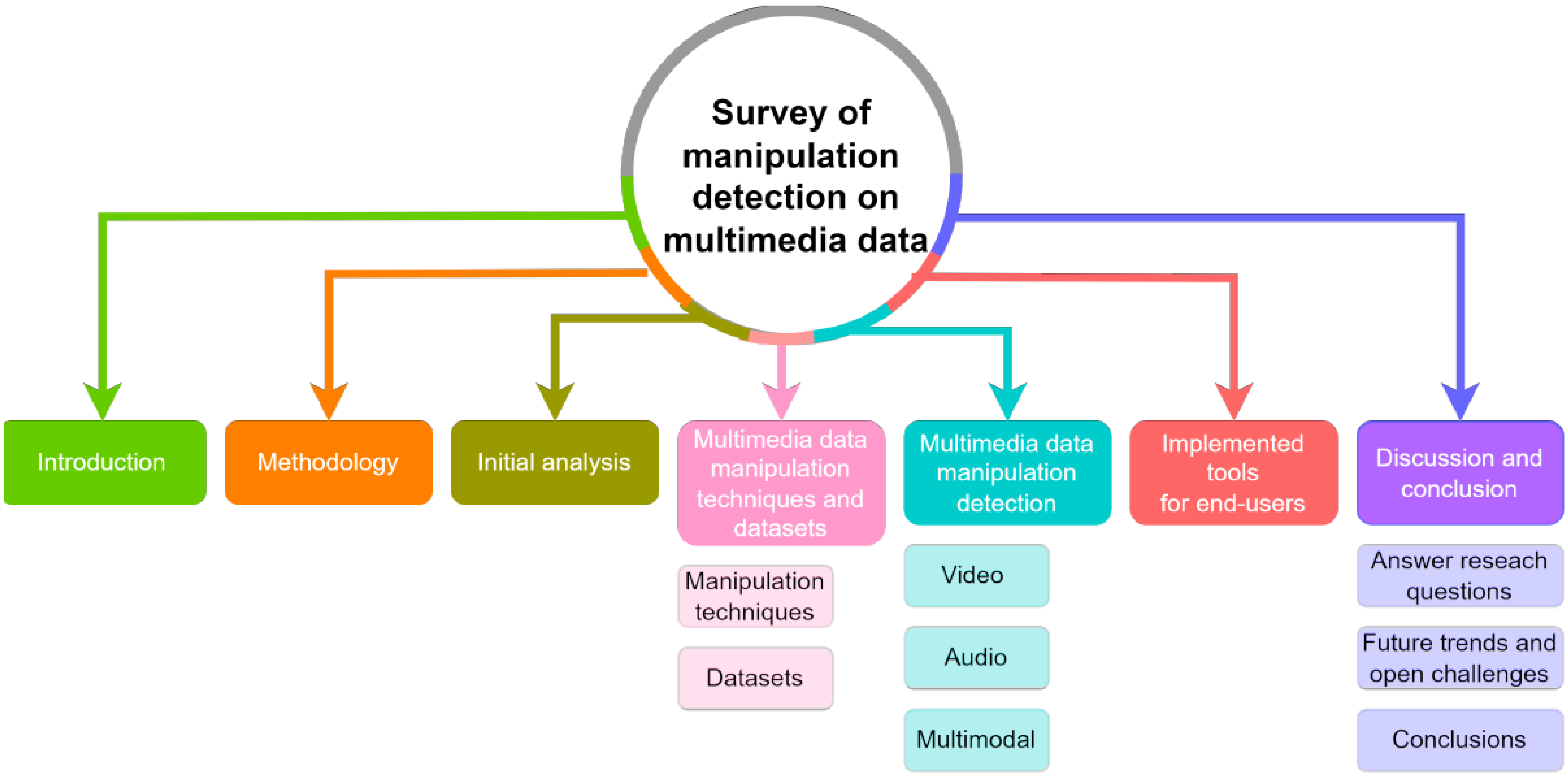

(f) Fake (SD) - 0|1|0|1|1


(g) Fake (PNDM) - 0|0|0|1|1


(h) Fake (GLIDE) - 0|0|1|1|1

The 5-digit binary code shows results from ResNet, Xception, DeiT, ViTGPT2, and BLIP2 models, where '0' means real and '1' means fake.

Multimodal Analysis



[Generation and detection of manipulated multimodal audiovisual content: Advances, trends and open challenges](https://www.sciencedirect.com/science/article/abs/pii/S1566253523004190)

H Liz-Lopez, M Keita, A Taleb-Ahmed, A Hadid, J Huertas-Tato, Information Fusion, 2024

https://www.sciencedirect.com/science/article/abs/pii/S1566253523004190

AI-Generated Content: Opportunities and Challenges

1. ~~About my Background~~
2. ~~Introduction to AIGC~~
3. ~~Detection of AIGC: SOTA~~
4. ~~Our Proposed Methods~~
5. Discussion

Take home messages

1. Nowadays, it is extremely easy to generate text, images, short videos, music - most recent models being based on diffusion-like models.

2. Impersonation, fraud and miss-use are serious threat. So, robust detection models are needed.

3. Perfect generalization is never garranted. Updates are always needed (horse race). Universal detectors?

4. Models-based on VLMs sound appealing.

SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ ABU DHABI

AI is a risky game that humanity should play !

Merci

Prof. Abdenour Hadid

abdenour.hadid@Sorbonne.ae

abdenour.hadid@ieee.org

SORBONNE CENTER FOR ARTIFICIAL INTELLIGENCE

SORBONNE UNIVERSITÉ ABU DHABI