Disinformation Detection: Deepfakes, Spoofing, and Forgeries in Scientific Publications

Prof. Anderson Rocha
Institute of Computing, Unicamp
arrocha@unicamp.br

Unicamp Professor for 15+ years
Expert in **Artificial Intelligence** and **Complex Data** (23+ years)
Research in both theoretical and applied aspects of Artificial Intelligence

Reasoning for Complex Data (**Recod.ai**) Lab. Coordinator
> Recod.ai counts with ~350 collaborators worldwide
> One of the largest and most productive in Latin America (LATAM)

IEEE Fellow
IEEE Biometrics Council Distinguished Lecturer
Microsoft, Google e Tan-Chin Tuan Foundation **Fellow**
Asia Pacific Association AI Fellow

Listed among the **TOP-2% Scientists** worldwide (According to Stanford/PlosOne Study)

Visiting Professor to multiple institutions over the years

Tag me

Synthetic Reality

AI-driven synthetic media

Context

Narratives

Number of AI-Created Images*

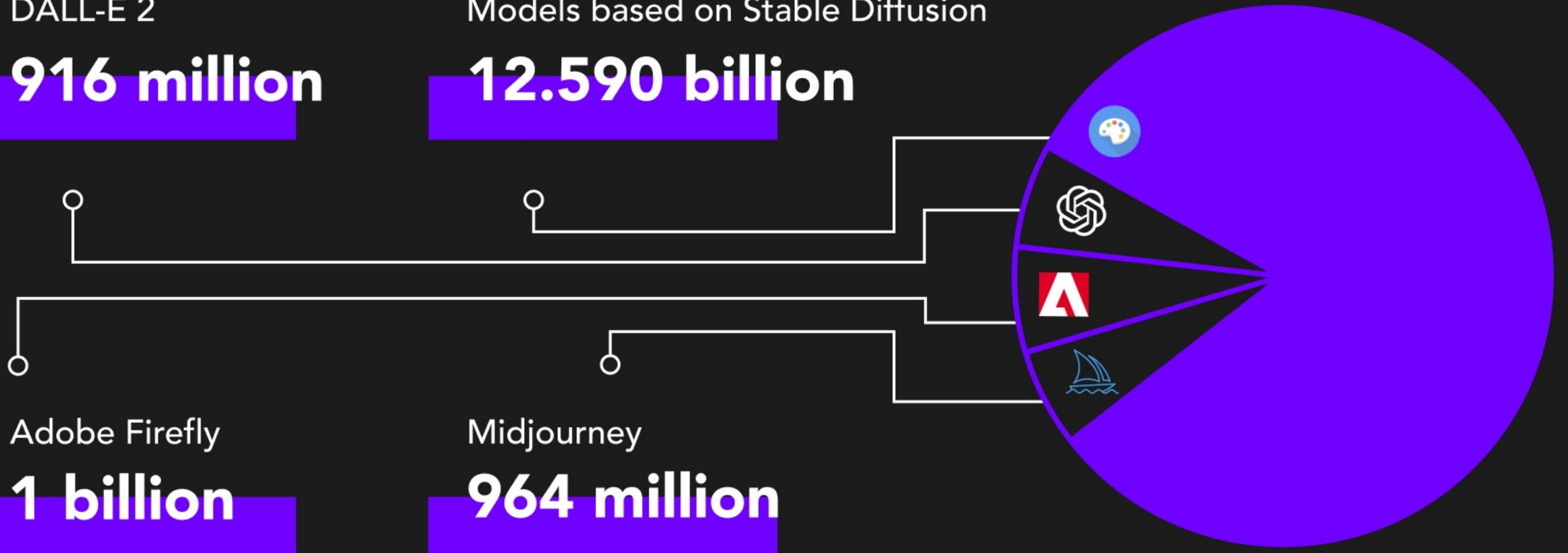
EVERYPIXEL

DALL-E 2

916 million

Models based on Stable Diffusion

12.590 billion



Adobe Firefly

1 billion

Midjourney

964 million

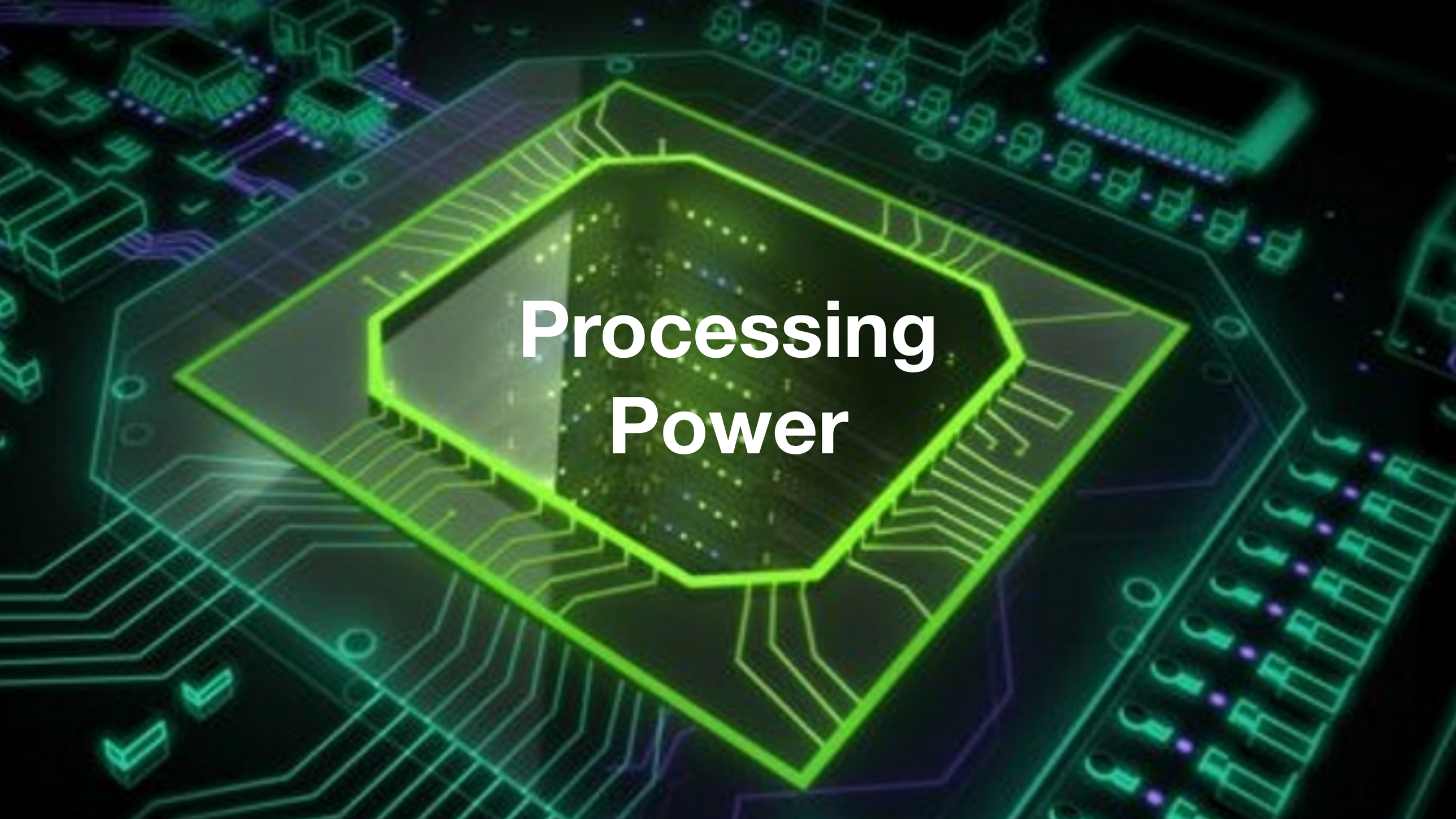
15.470 billion

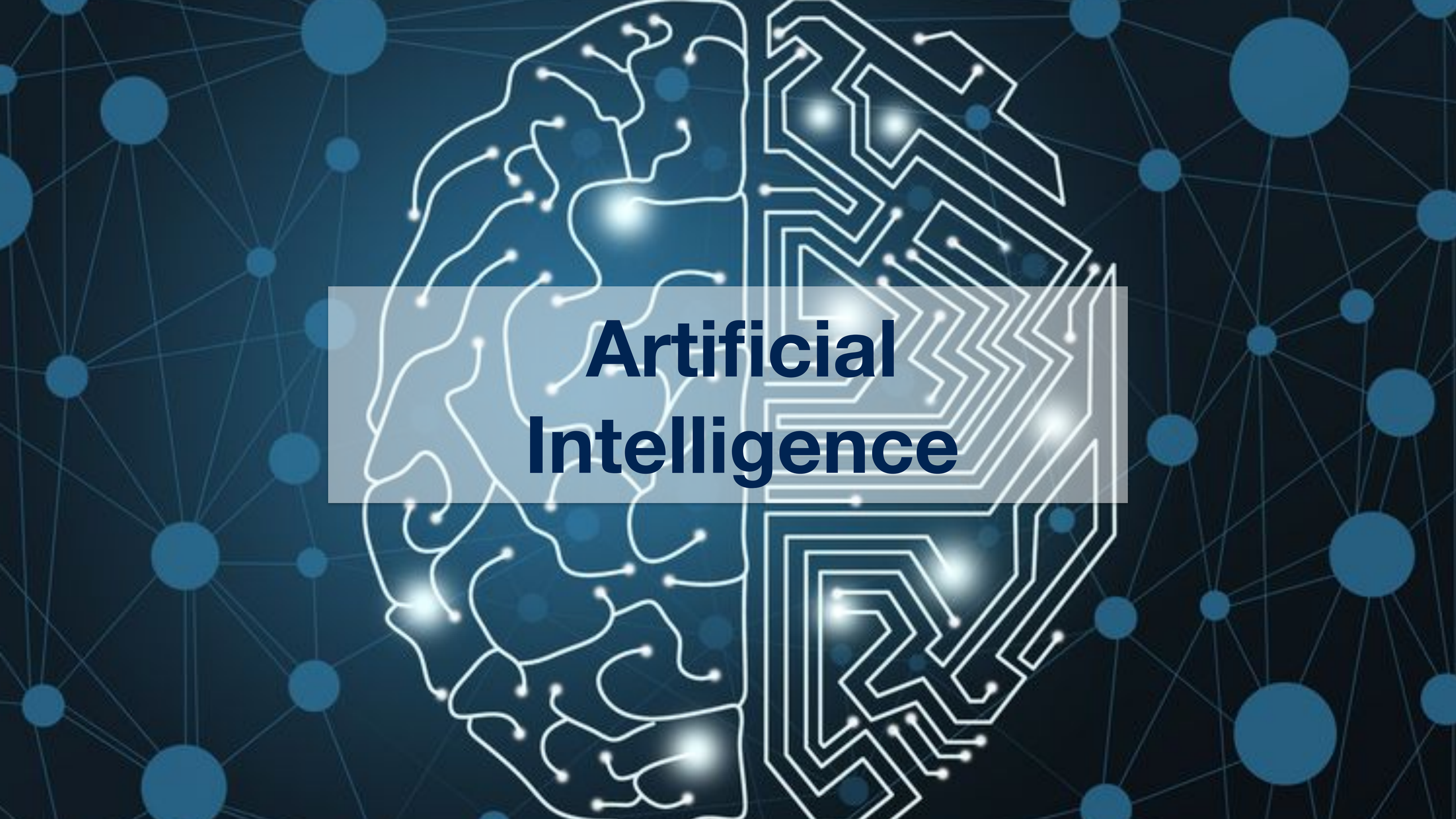
Sources: Adobe;
our estimates, based on Photutorial, OpenAI, Civitai

*As of August 2023

> Data and Image from EveryPixel


Data!
(as never before)


Processing
Power

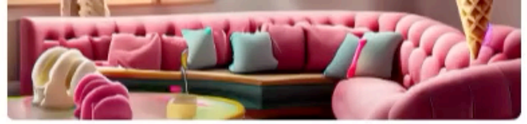

Artificial
Intelligence


Théâtre d'Opéra Spatial by Jason M. Allen

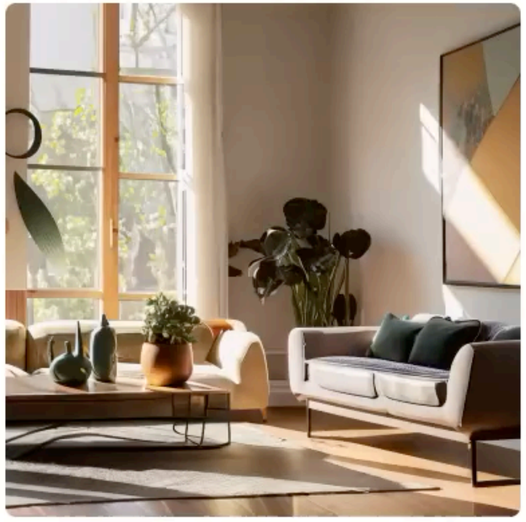
https://www.cpr.org/2023/09/06/jason-allens-ai-art-won-colorado-fair-feds-deny-copyright-protection/

Jason Allen at his Atelier, Colorado, U.S.













ChatGPT
deepseek


runway

All digital content has a history

In this new world of synthetic media and generative AI, the need for transparency has arrived. Using C2PA, Truepic provides publishers, creators, and consumers the ability to trace the origin of different types of media.

2.8B

people regularly use image editing apps

34.0M

images are generated with AI every day

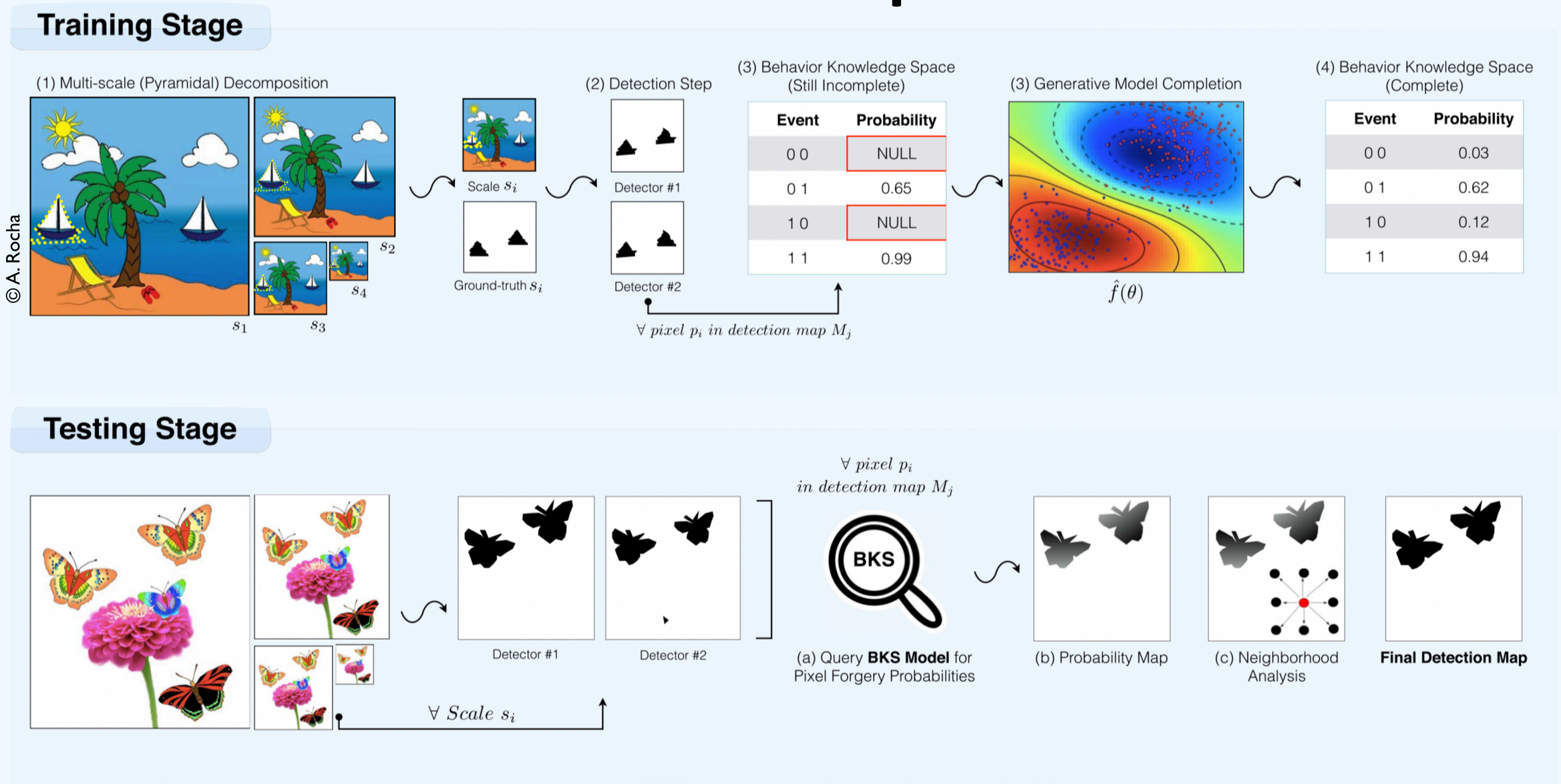
51.1%

of online misinformation comes from manipulated images

> Data and Image from TruePic.com

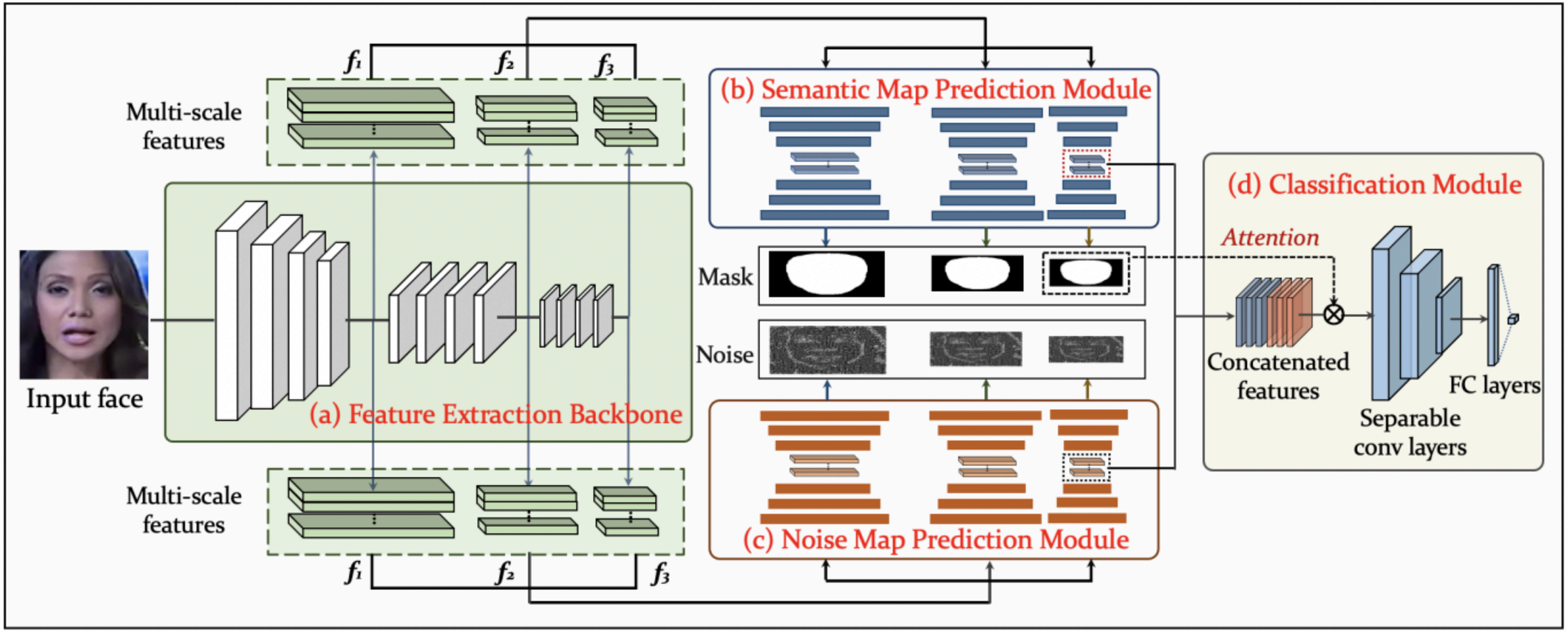
What can we **do**?

Empower detection methods



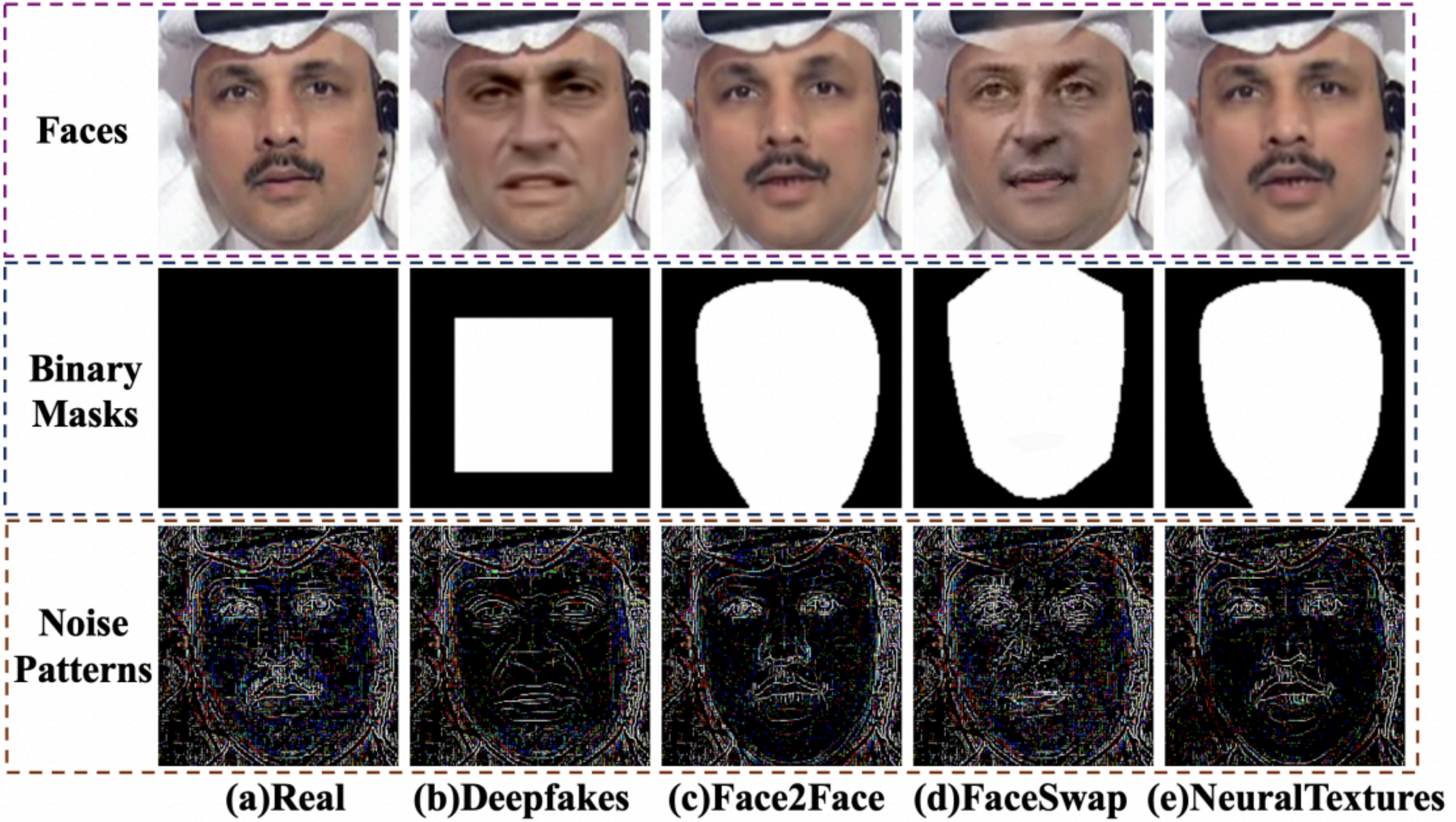
Ferreira, Anselmo, et al. "Behavior knowledge space-based fusion for copy–move forgery detection." IEEE Transactions on Image Processing 25.10 (2016): 4729-4742.

Explore **unseen** telltales



Kong, Chenqi, et al. "Detect and locate: Exposing face manipulation by semantic-and noise-level telltales." IEEE Transactions on Information Forensics and Security 17 (2022): 1741-1756.

Explore **unseen** telltales



Kong, Chenqi, et al. "Detect and locate: Exposing face manipulation by semantic-and noise-level telltales." IEEE Transactions on Information Forensics and Security 17 (2022): 1741-1756.

SYNTHETIC REALITIES: WHERE ARE WE?

Overview Paper

The Age of Synthetic Realities: Challenges and Opportunities

recod.ai

João Phillipe Cardenuto[1*], Jing Yang[1], Rafael Padilha[1], Renjie Wan[2], Daniel Moreira[3], Haoliang Li[4], Shiqi Wang[5], Fernanda Andaló[1], Sébastien Marcel[6,7] and Anderson Rocha[1]

[1] *Artificial Intelligence Lab., Recod.ai, Institute of Computing, Universidade Estadual de Campinas, Campinas, SP, Brazil*
[2] *Department of Computer Science, Hong Kong Baptist University, Hong Kong*
[3] *Department of Computer Science, Loyola University Chicago, USA*
[4] *Department of Electrical Engineering, City University of Hong Kong, Hong Kong*
[5] *Department of Computer Science, City University of Hong Kong, Hong Kong*
[6] *Idiap Research Institute, Martigny, Switzerland*
[7] *University of Lausanne, Lausanne, Switzerland*

Counteracting the contemporaneous proliferation of digital forgeries and fake news

ALEXANDRE FERREIRA[1], TIAGO CARVALHO[2], FERNANDA ANDALÓ[1] and ANDERSON ROCHA[1]

[1]Institute of Computing, University of Campinas (Unicamp), Av. Albert Einstein, 1251, 13083-852 Campinas, SP, Brazil
[2]Instituto Federal de São Paulo (IFSP), Av. Comendador Aladino Selmi, s/n, 13069-901 Campinas, SP, Brazil

Leveraging Ensembles and Self-Supervised Learning for Fully-Unsupervised Person Re-Identification and Text Authorship Attribution

Gabriel Bertocco, Antonio Theophilo, Fernanda Andaló, *Member, IEEE,* and Anderson Rocha, *Senior Member, IEEE*

EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR AUTHORSHIP ATTRIBUTION ON SOCIAL MEDIA

Antonio Theophilo†, Rafael Padilha*, Fernanda A. Andaló*, Anderson Rocha**

* Artificial Intelligence Lab. (Recod.ai) Institute of Computing, University of Campinas, Brazil
† Center for Information Technology Renato Archer, Campinas, Brazil

Content-Based Detection of Temporal Metadata Manipulation

Rafael Padilha[1] ✉, Tawfiq Salem[2], Scott Workman[3], Fernanda A. Andaló[1], Anderson Rocha[1], Nathan Jacobs[4]

[1] University of Campinas, Brazil
[2] Purdue University, USA
[3] DZYNE Technologies, USA
[4] University of Kentucky, USA

Forensic Event Analysis: From Seemingly Unrelated Data to Understanding

Rafael Padilha, Caroline Mazini Rodrigues, Fernanda Andaló, Gabriel Bertocco, Zanoni Dias, and Anderson Rocha

nature

How to stop AI deepfakes from sinking society — and science

Deceptive videos and images created using generative AI could sway elections, crash stock markets and ruin reputations. Researchers are developing methods to limit their harm.

By Nicola Jones

Illustration by Señor Salme

scientific reports

nature

Explore content | About the journal | Publish with us

nature > scientific reports > articles > article

Article | Open access | Published: 31 October 2022

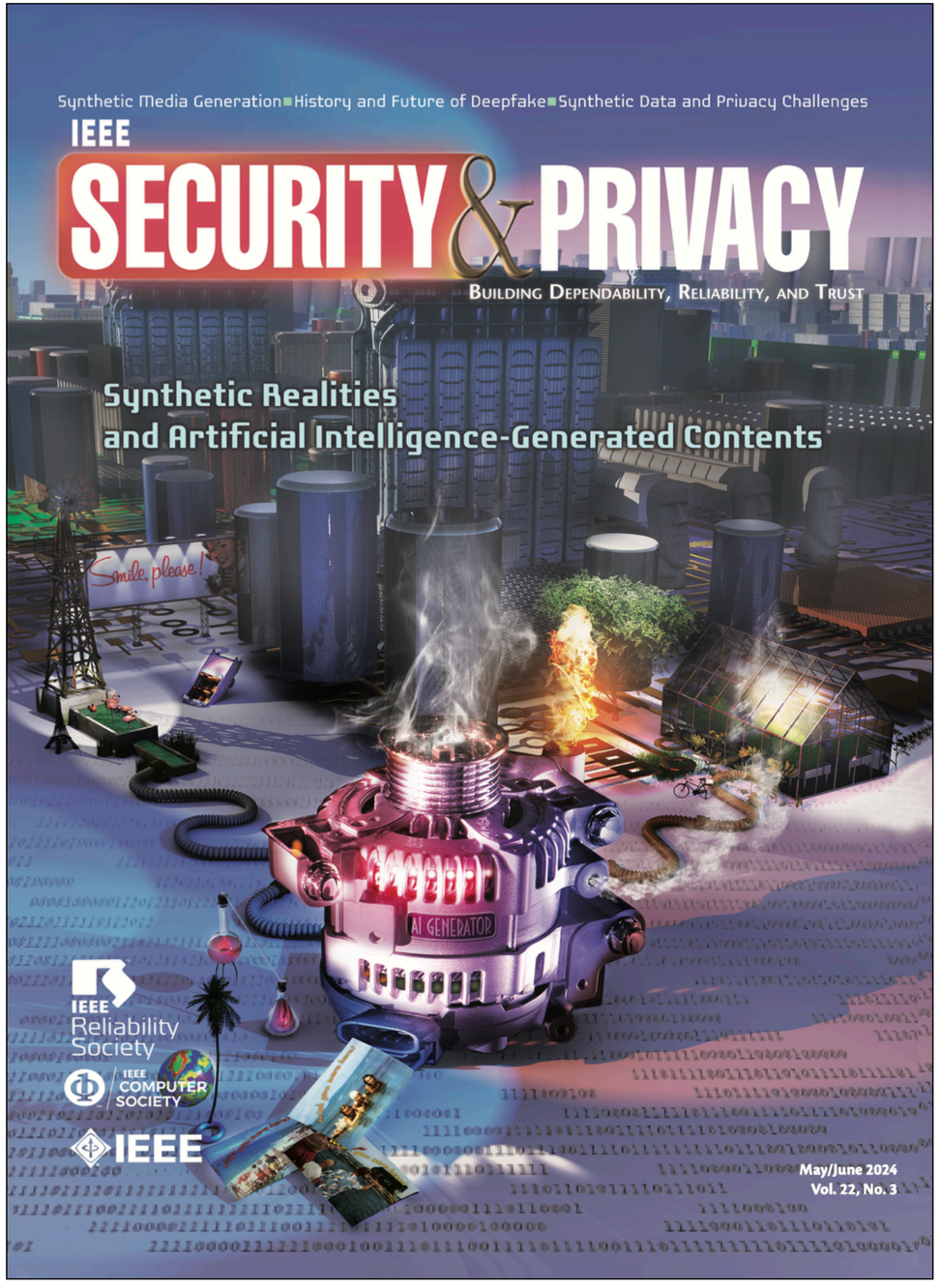
SILA: a system for scientific image analysis

Daniel Moreira, João Phillipe Cardenuto, Ruiting Shao, Sriram Baireddy, Davide Cozzolino, Diego Gragnaniello, Wael Abd-Almageed, Paolo Bestagini, Stefano Tubaro, Anderson Rocha, Walter Scheirer, Luisa Verdoliva & Edward Delp

Scientific Reports **12**, Article number: 18306 (2022) | Cite this article

Abstract

A great deal of the images found in scientific publications are retouched, reused, or composed to enhance the quality of the presentation. In most instances, these edits are benign and help the reader better understand the material in a paper. However, some edits are instances of scientific misconduct and undermine the integrity of the presented research. Determining the legitimacy of edits made to scientific images is an open problem that no current technology can perform satisfactorily in a fully automated fashion. It thus remains up to human experts to inspect images as part of the peer-review process. Nonetheless, image analysis technologies promise to become helpful to experts to perform such an essential yet arduous task. Therefore, we introduce SILA, a system that makes



Synthetic Realities and Artificial Intelligence-Generated Contents

Daniel Moreira | Loyola University Chicago
Sébastien Marcel | Idiap Research Institute
Anderson Rocha | University of Campinas

Welcome to the *IEEE Security & Privacy* special issue on synthetic realities and artificial intelligence-generated contents! In this edition, we delve into the topic of synthetic realities, where generative artificial intelligence (GAI) is revolutionizing the construction of narratives, blurring the boundaries between fact and fiction, for the good and the bad. Indeed, content created or enabled by GAI spans a wide spectrum of usage and intentions, from fostering positive experiences, such as entertainment, training, and education, to more questionable utilization, such as deception, propaganda, and manipulation.

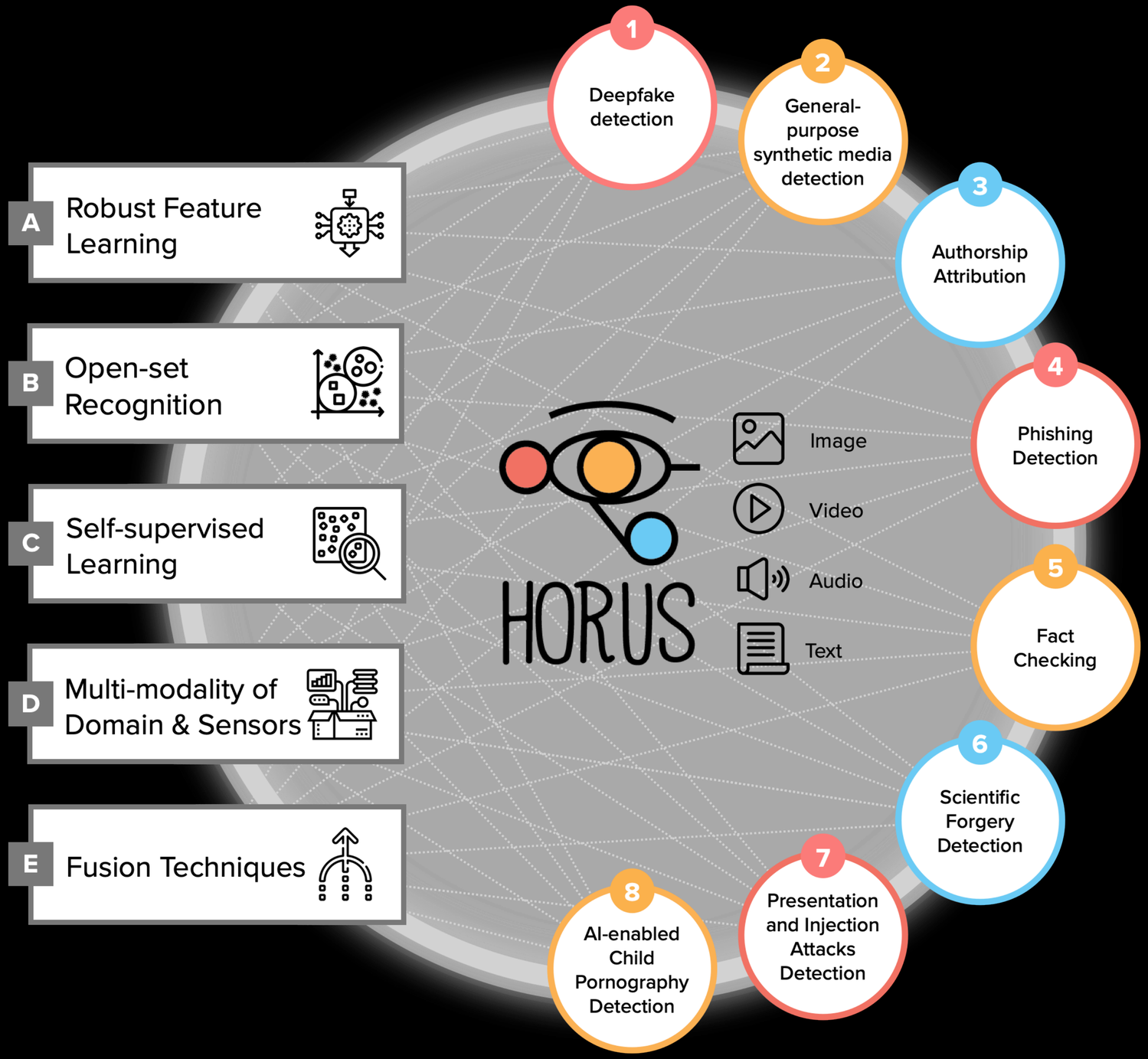
With the advent and maturity of GAI techniques, much has changed in forensics, security, and privacy. The way researchers and experts have been doing forensics and security over the past decades is continuously challenged with each new version of powerful AI content generators. The synthetic content ranges from audio, image, and video to text and their combinations, coming from prominent models, such as ChatGPT, LaMDA, ImageGen, StableDiffusion, Sora, and Gemini, among others.

This special issue seeks to understand the required changes in the way forensics, security, and privacy experts operate, including how to deal with autogenerated fake and synthetic data (e.g., text, images, videos, and 3D content), how much autogeneration methods are "shaping" new realities that do not exist, and what it means for our society. The call presented the following important questions: What are the possible new applications for forensics, security, and privacy? What are the threats and challenges? Forensic aspects should include any topics related to post hoc investigation practices after the occurrence of events regarding created content (eg, generated fake news or deepfakes and how to detect them). Security aspects should include topics related to how such contents might affect our lives in terms of document authenticity and deception. Privacy should

©SHUTTERSTOCK.COM/SUMMIT ART CREATIONS

Digital Object Identifier 10.1109/MSEC.2024.3388244
Date of current version: 10 May 2024


HORUS


A
Robust Feature Learning
B
Open-set Recognition
C
Self-supervised Learning
D
Multi-modality of Domain & Sensors
E
Fusion Techniques
HORUS
Image
Video
Audio
Text
1
Deepfake detection
2
General-purpose synthetic media detection
3
Authorship Attribution
4
Phishing Detection
5
Fact Checking
6
Scientific Forgery Detection
7
Presentation and Injection Attacks Detection
8
AI-enabled Child Pornography Detection

Phishing

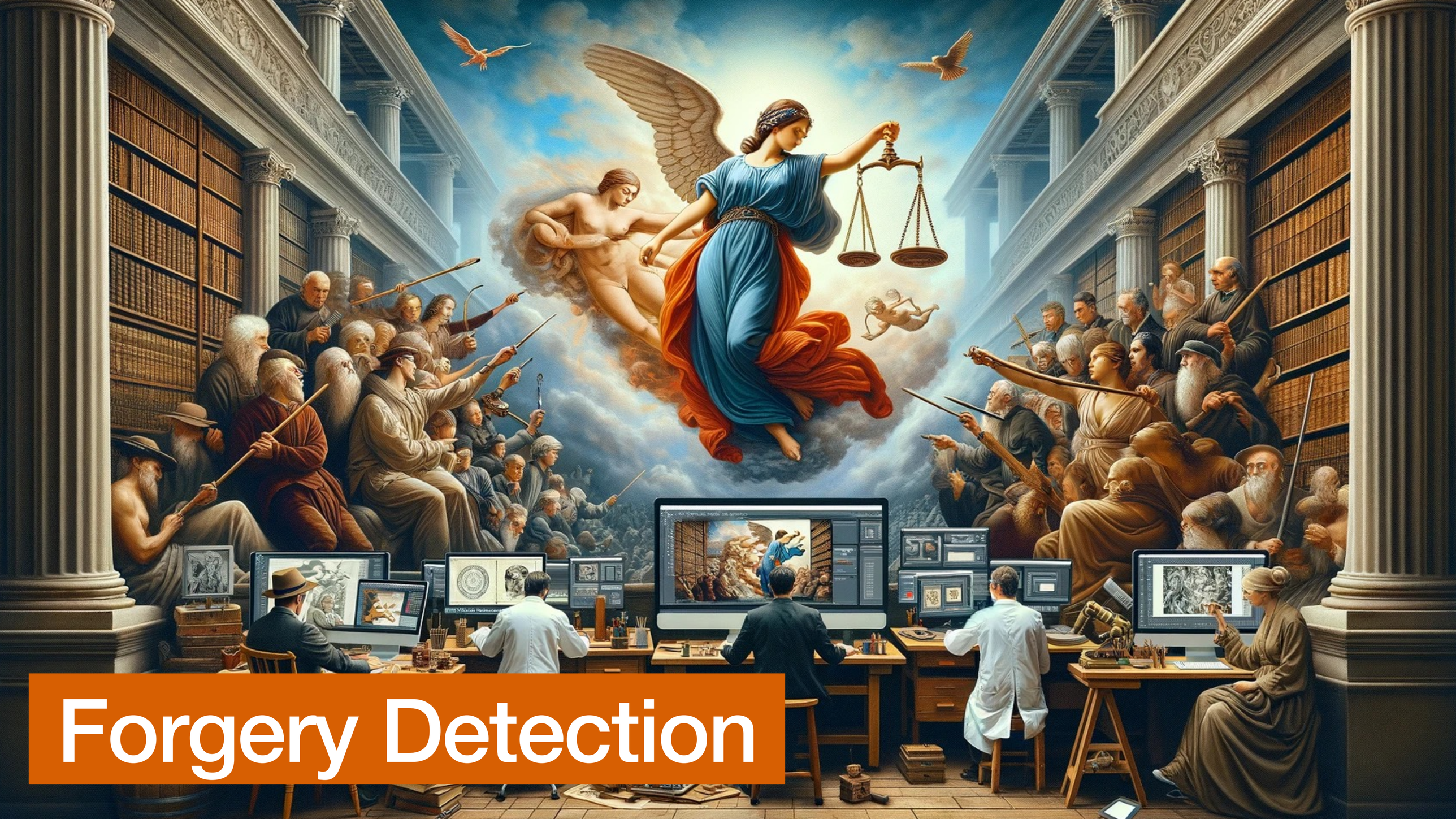
Deepfake

Communications

Parental Control

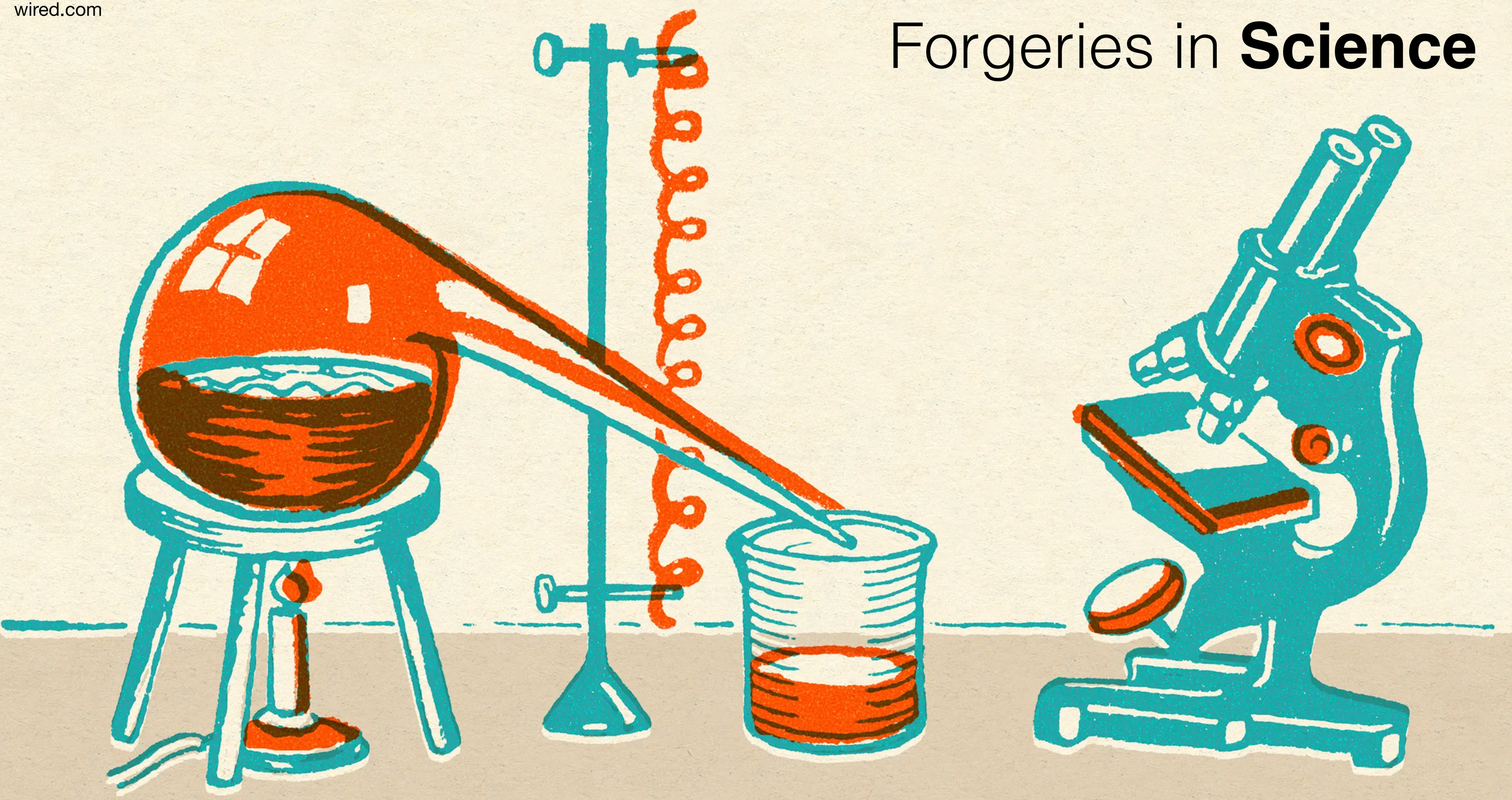
Liveness Proof

Content Protection


Forgery Detection

wired.com

Forgeries in **Science**



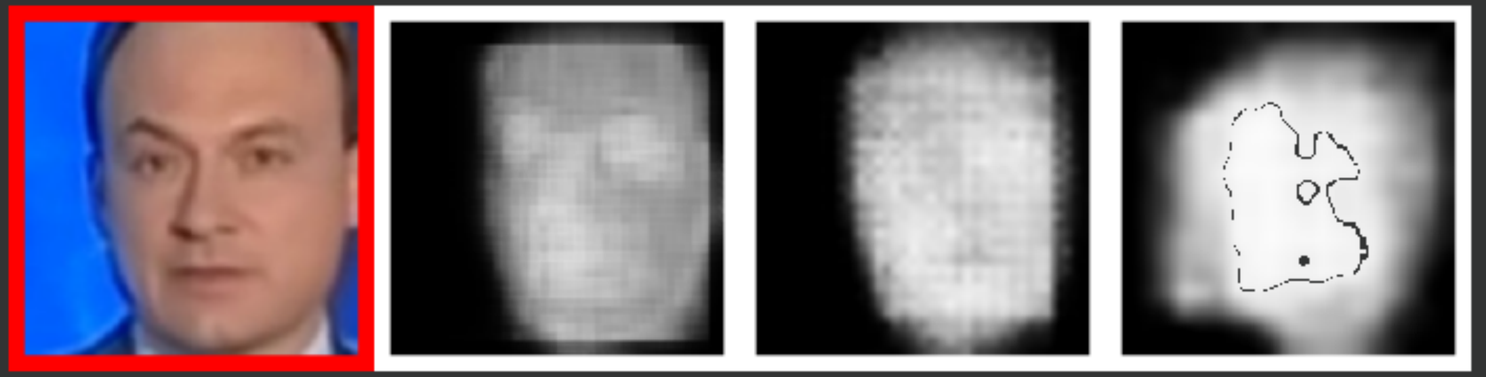
wired.com

Fighting AI-enabled **Child Pornography**

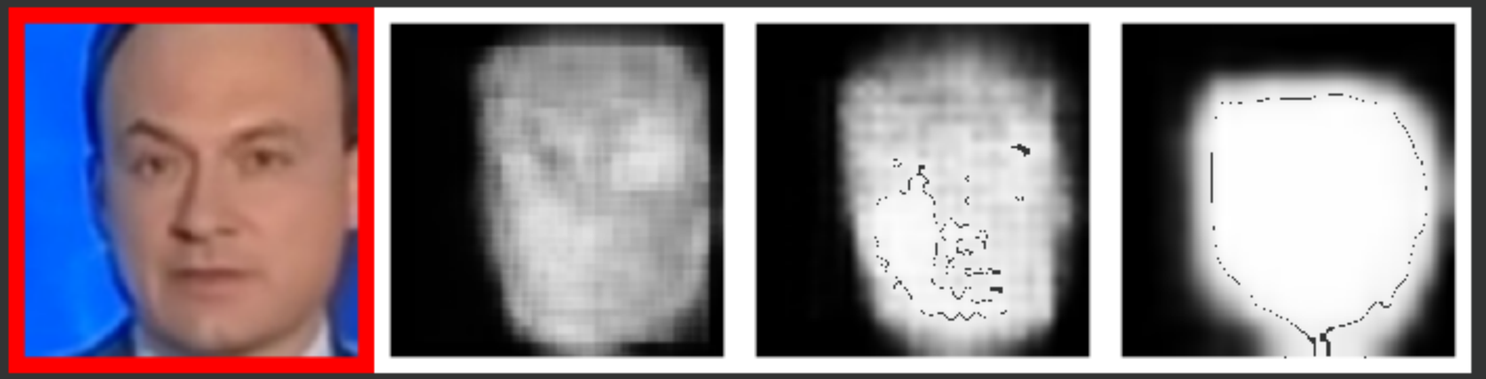
recod.ai

reasoning for complex data

DeepFake Detection System



Probability of Fake for Expert #1: 90.07%



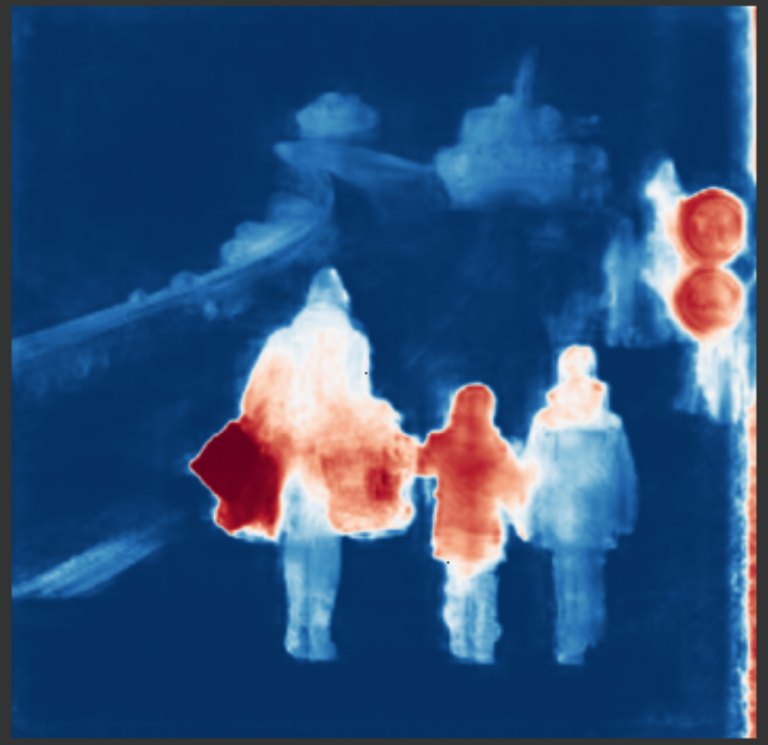
Probability of Fake for Expert #2: 99.81%

Upload Image

Detect

recod.ai
reasoning for complex data

Image Forgery Detection System



Prob. of Forgery for Expert #1: 96.06%

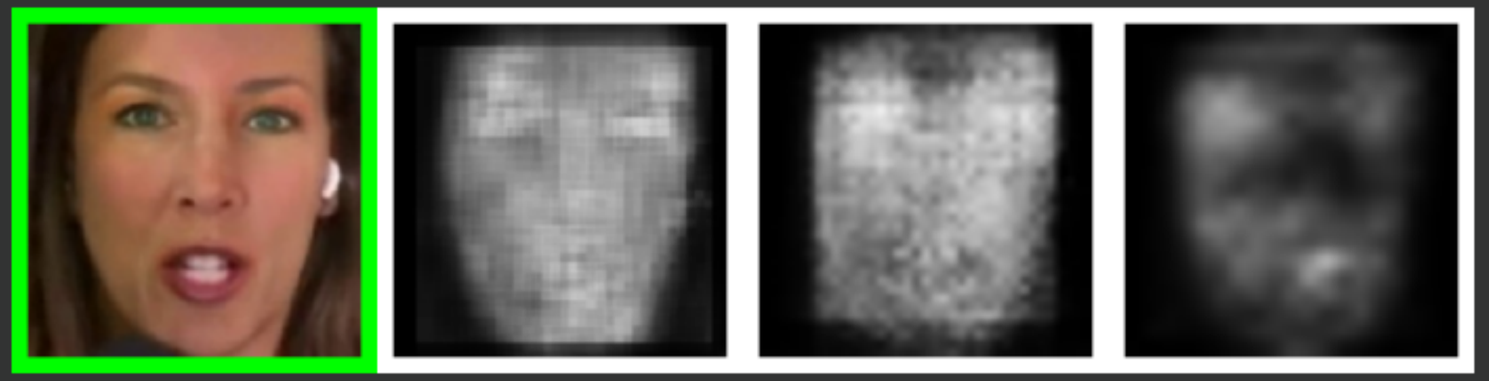
Prob. of of Forgery for Expert #2: 52.40%

Upload Image

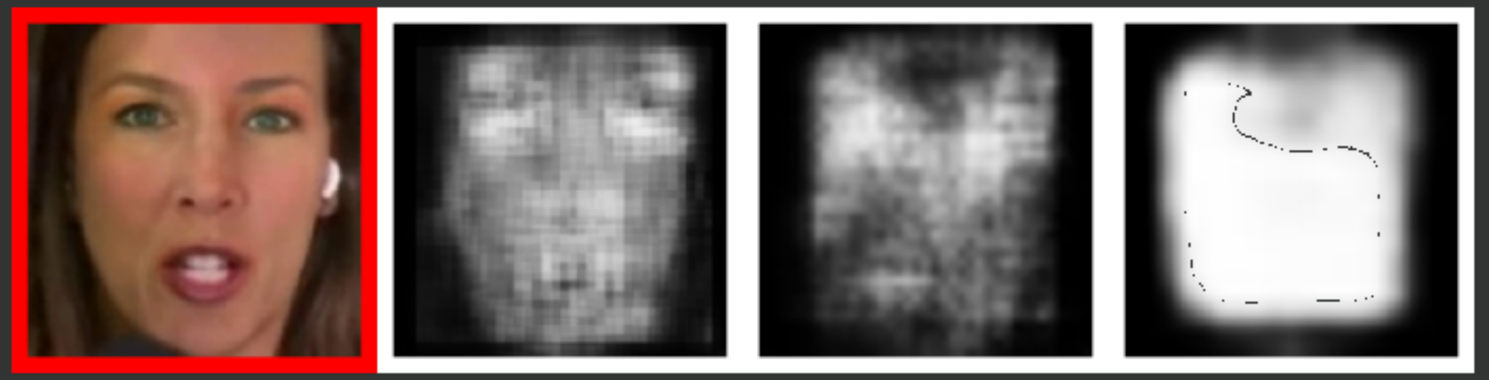
Detect

recod.ai DeepFake Detection System

reasoning for complex data



Probability of Fake for Expert #1: 31.31%



Probability of Fake for Expert #2: 95.91%

Upload Image

Detect

recod.ai
reasoning for complex data

Image Forgery Detection System

Prob. of Forgery for Expert #1: 99.99%

Prob. of of Forgery for Expert #2: 99.68%

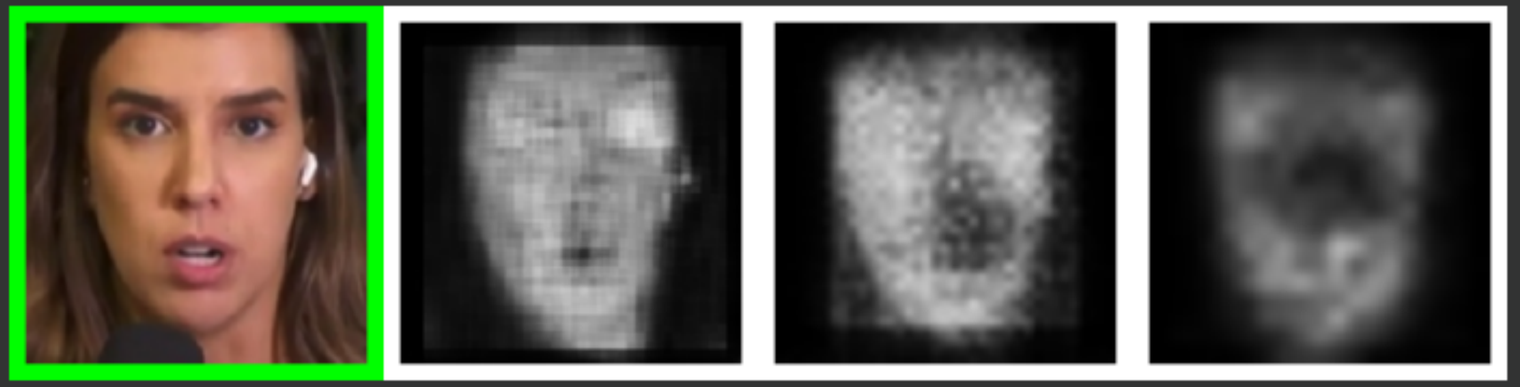
Upload Image

Detect

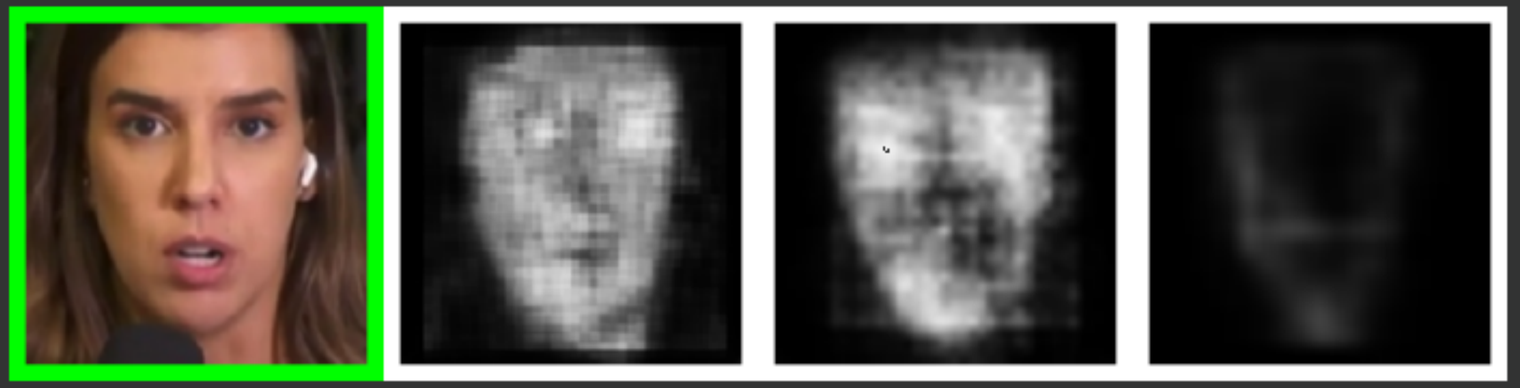
recod.ai

reasoning for complex data

DeepFake Detection System



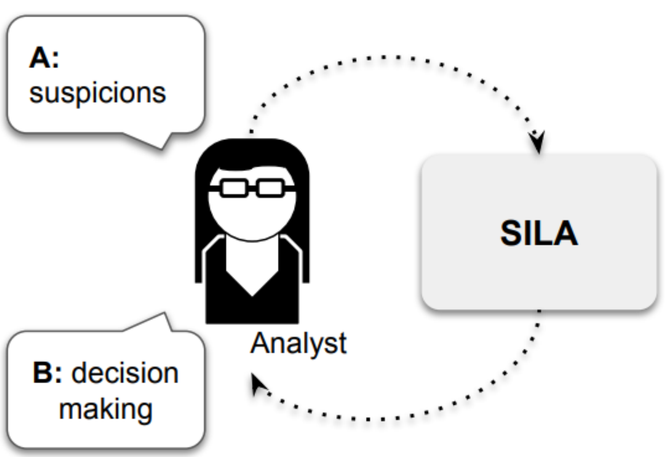
Probability of Fake for Expert #1: 18.37%



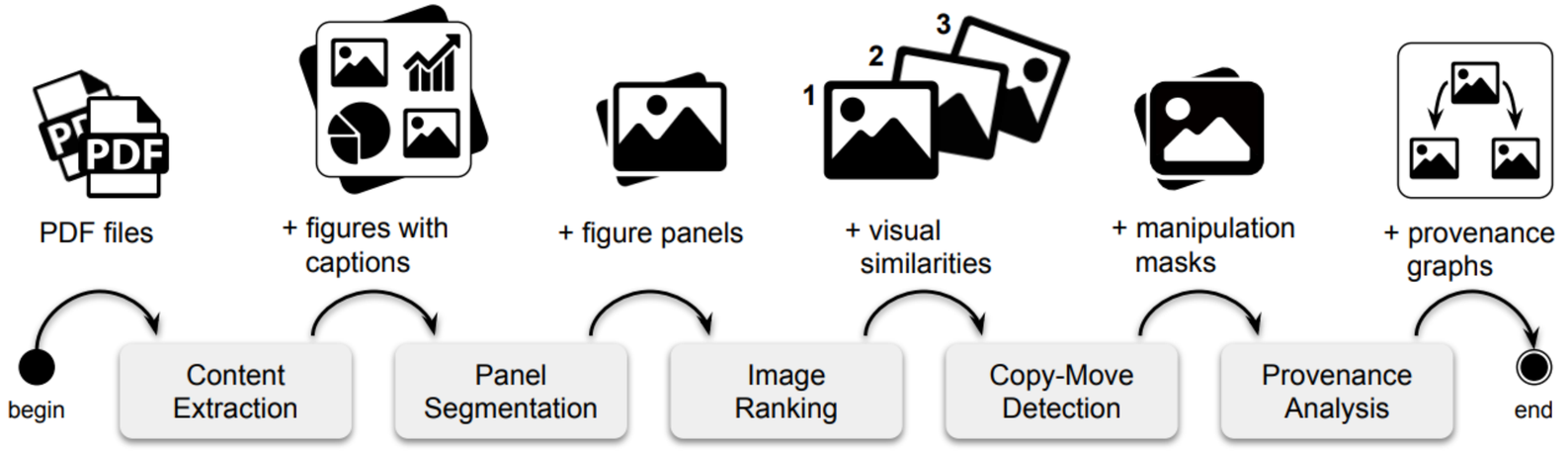
Probability of Fake for Expert #2: 7.54%

Upload Image

Detect



(a) SILA overview.



(b) Proposed system workflow.

www.nature.com/scientificreports

scientific reports

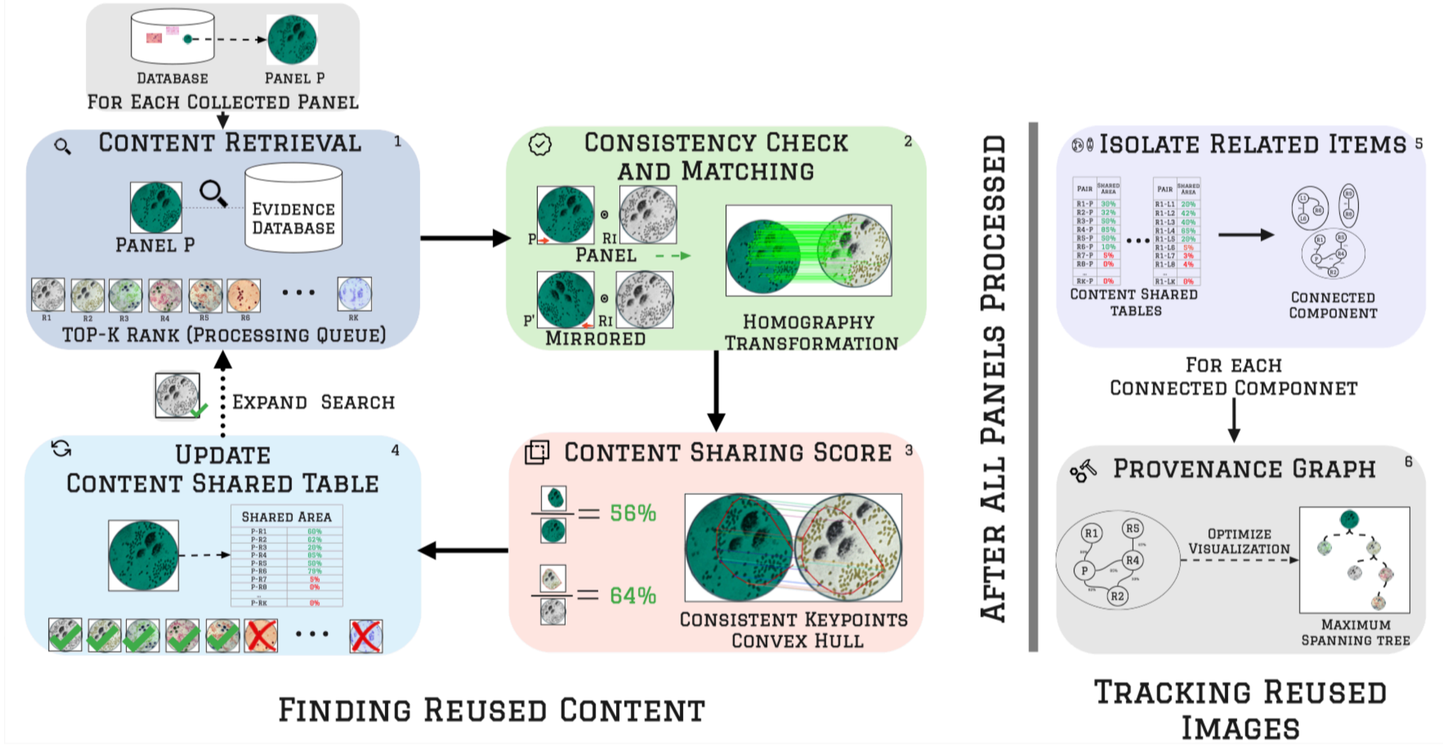
OPEN

SILA: a system for scientific image analysis

Daniel Moreira[1], João Phillipe Cardenuto[2], Ruiting Shao[3], Sriram Baireddy[3], Davide Cozzolino[4], Diego Gragnaniello[5], Wael Abd-Almageed[6], Paolo Bestagini[7], Stefano Tubaro[7], Anderson Rocha[2], Walter Scheirer[8], Luisa Verdoliva[9] & Edward Delp[3]✉

A great deal of the images found in scientific publications are retouched, reused, or composed to enhance the quality of the presentation. In most instances, these edits are benign and help the reader better understand the material in a paper. However, some edits are instances of scientific misconduct and undermine the integrity of the presented research. Determining the legitimacy of edits made to scientific images is an open problem that no current technology can perform satisfactorily in a fully automated fashion. It thus remains up to human experts to inspect images as part of the peer-review process. Nonetheless, image analysis technologies promise to become helpful to experts to perform such an essential yet arduous task. Therefore, we introduce SILA, a system that makes image analysis tools available to reviewers and editors in a principled way. Further, SILA is the first human-in-the-loop end-to-end system that starts by processing article PDF files, performs image manipulation detection on the automatically extracted figures, and ends with image provenance graphs expressing the relationships between the images in question, to explain potential problems. To assess its efficacy, we introduce a dataset of scientific papers from around the globe containing annotated image manipulations and inadvertent reuse, which can serve as a benchmark for the problem at hand. Qualitative and quantitative results of the system are described using this dataset.

Since the early days of photography, images have been used in scientific publications to illustrate the proposed methods, aid in explaining theories, and—most importantly—present the results of experiments. Photography itself became part of experimentation, producing key results such as *Photo 51*, an X-ray diffraction image clearly showing the structure of deoxyribonucleic acid (DNA) for the first time[1]. Later on, with the advent and popularization of digital photography, digital images were added to the scientific repertory, greatly enhancing the speed at which photographic content is produced. In some scientific fields such as biomedicine, images captured by dedicated apparatus are accepted as the results themselves, constituting the elements to be scrutinized while

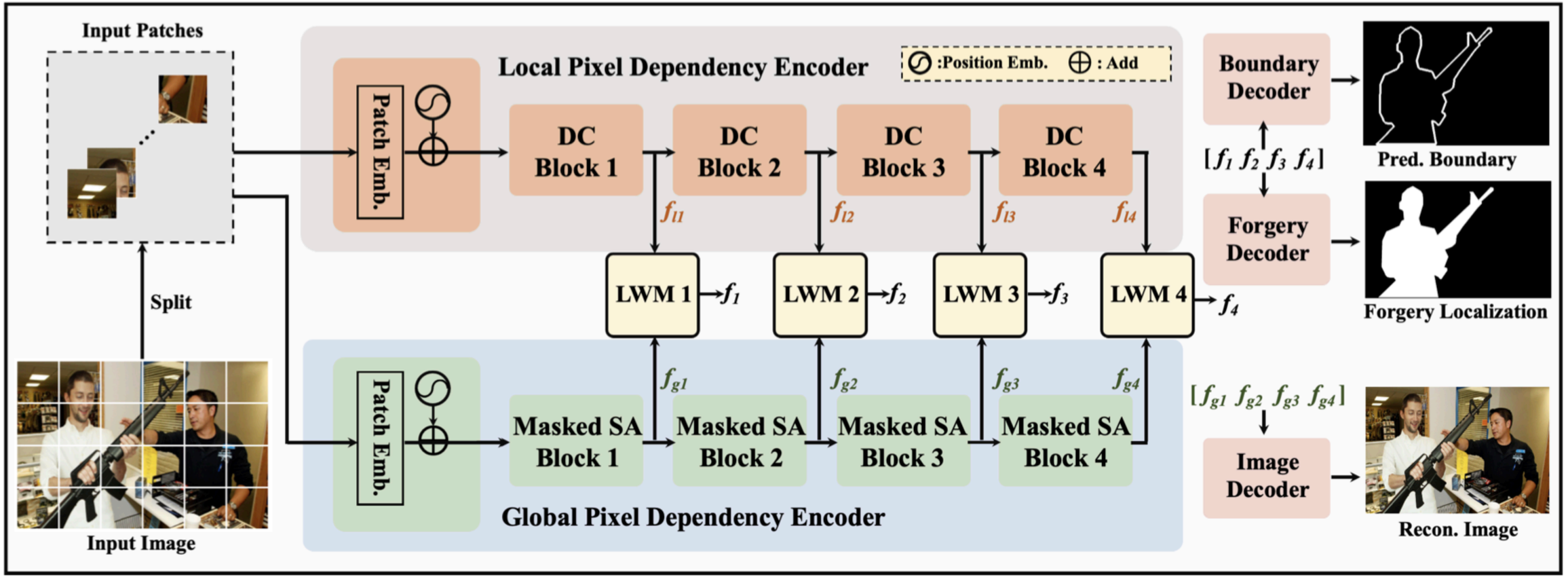


This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2025.3541028

Pixel-Inconsistency Modeling for Image Manipulation Localization

Chenqi Kong, *Member, IEEE,* Anwei Luo, Shiqi Wang, *Senior Member, IEEE,* Haoliang Li, *Member, IEEE,* Anderson Rocha, *Fellow, IEEE,* and Alex C. Kot, *Life Fellow, IEEE*

IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND MACHINE INTELLIGENCE

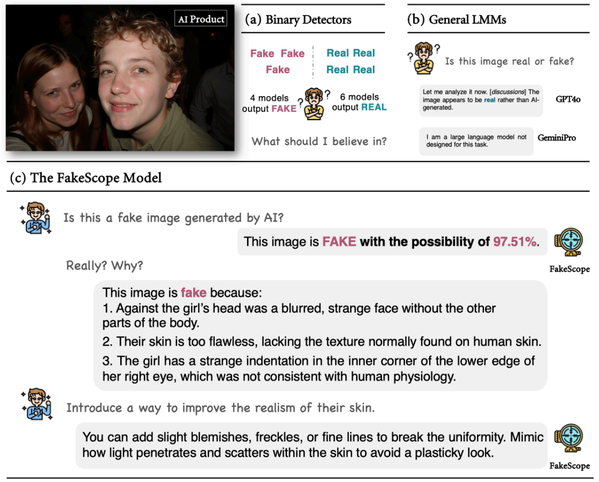


FakeScope: Large Multimodal Expert Model for Transparent AI-Generated Image Forensics

Yixuan Li, Yu Tian, Yipo Huang, Wei Lu, *Member, IEEE*, Shiqi Wang†, *Senior Member, IEEE*, Weisi Lin, *Fellow, IEEE* and Anderson Rocha, *Fellow, IEEE*

IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND MACHINE INTELLIGENCE

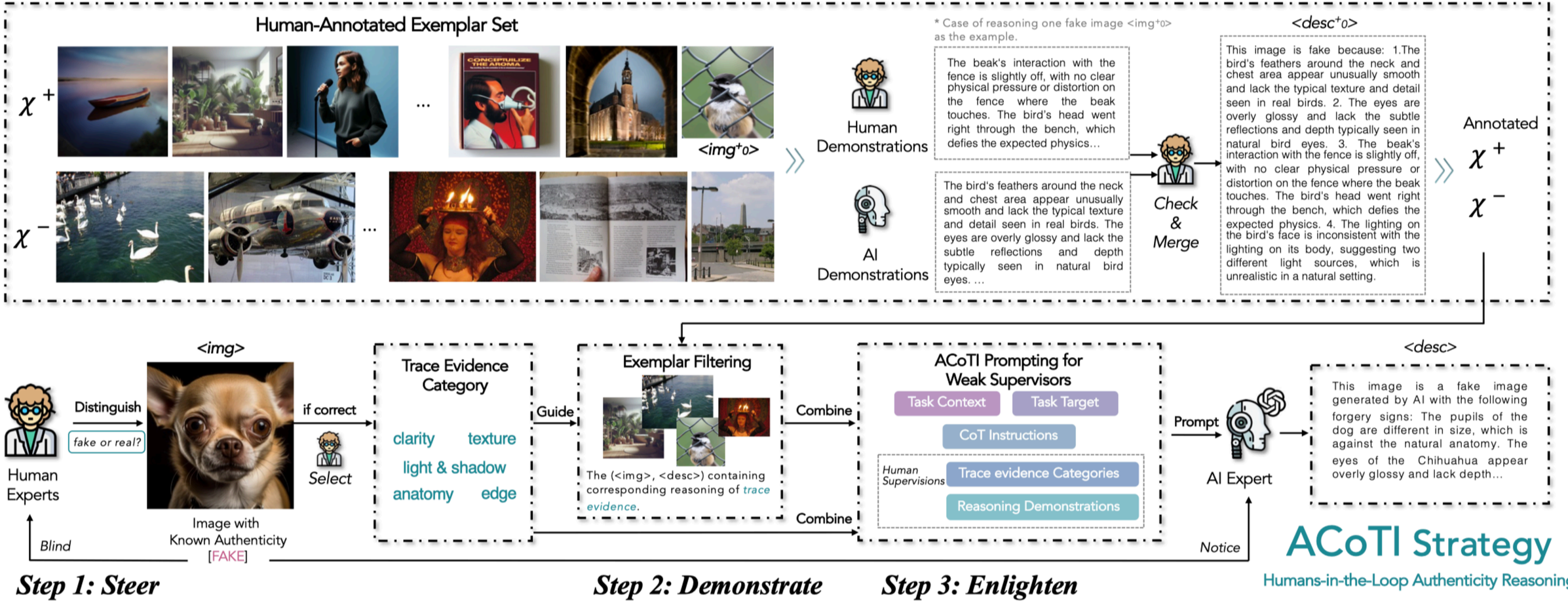
Fig. 2. Contributions of this work. (a) FakeChain dataset, containing long-form reasoning on image authenticity, constructed via the proposed *ACoTI* strategy (Sec. 3); (b) FakeInstruct, containing 2 million visual instructions of image forensic knowledge (Sec. 4); (c) FakeScope model, the expert model for transparent AI-generated image forensics, capable of multi-dimensional forensic capabilities (Sec. 5).



FakeScope: Large Multimodal Expert Model for Transparent AI-Generated Image Forensics

Yixuan Li, Yu Tian, Yipo Huang, Wei Lu, *Member, IEEE*, Shiqi Wang†, *Senior Member, IEEE*, Weisi Lin, *Fellow, IEEE* and Anderson Rocha, *Fellow, IEEE*

IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND MACHINE INTELLIGENCE



FakeScope: Large Multimodal Expert Model for Transparent AI-Generated Image Forensics

Yixuan Li, Yu Tian, Yipo Huang, Wei Lu, *Member, IEEE*, Shiqi Wang†, *Senior Member, IEEE*, Weisi Lin, *Fellow, IEEE* and Anderson Rocha, *Fellow, IEEE*

IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND MACHINE INTELLIGENCE

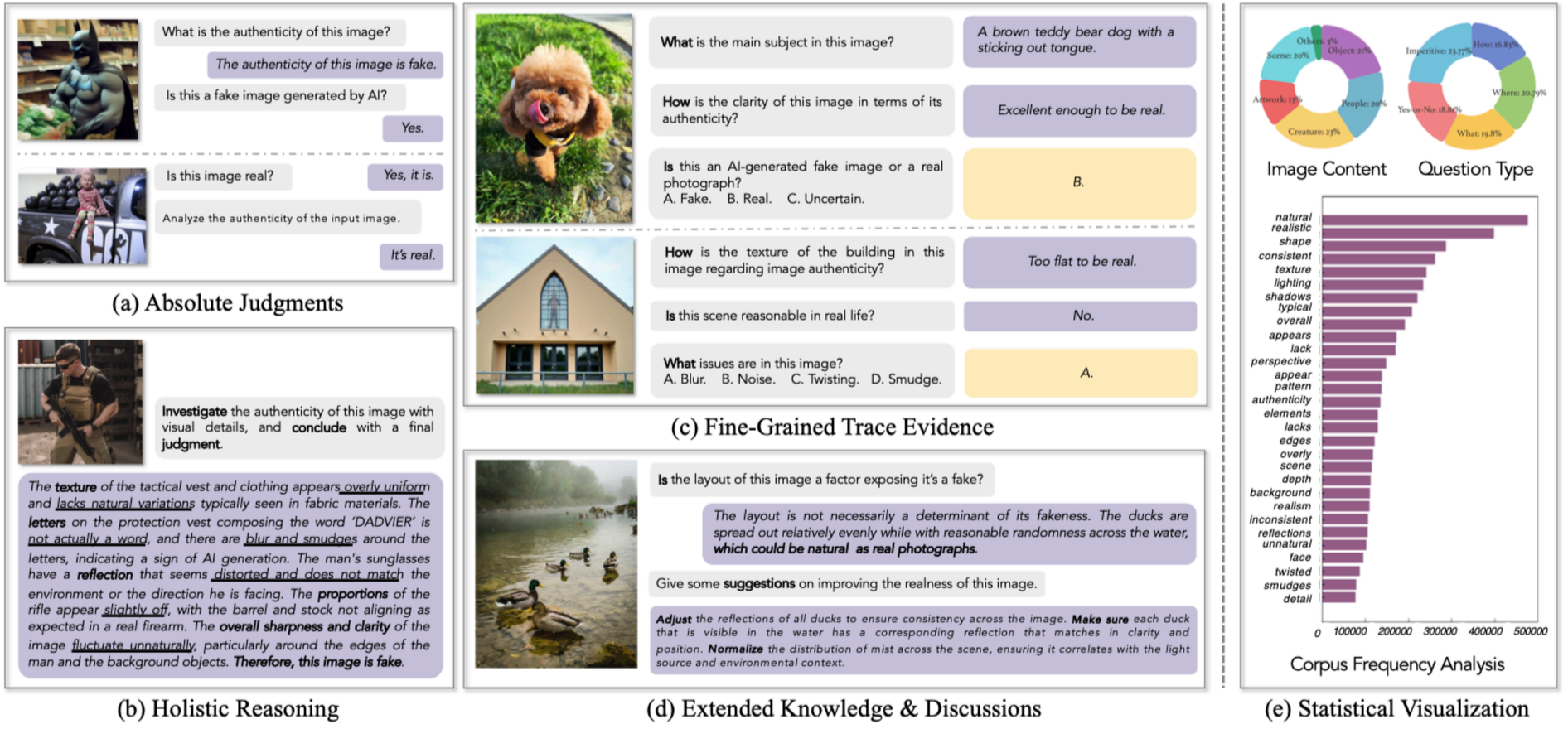


Fig. 4. The composition of the **FakeInstruct**, which is derived from the FakeChain dataset, containing 47K visual instructions on absolute authenticity judgments, 95K instructions on holistic reasoning, 715K on fine-grained visual trace evidence, and 1190K on extended knowledge and discussions. The million-scale diversified visual instructions of FakeInstruct enable LMMs with a broad and fine-grained understanding of image authenticity, ensuring LMMs are equipped to handle diverse forensic tasks with both interpretability and accuracy.

recod.ai
reasoning for complex data

Buscar

UNICAMP

INÍCIO | SOBRE | OPORTUNIDADES | NOVIDADES | EQUIPE | PUBLICAÇÕES | PROJETOS | RECOD.AI NA MÍDIA | OUTROS

RECOD.AI NA ONU

RECOD.AI NA IEEE TRANSACTIONS...

OPORTUNIDADE: PESQUISADOR E...

POS-DOUTORADO: APRENDIZADO ...

POS-DOUTORADO: ENGENHARIA D...

Recod.ai na ONU

Nesta quinta-feira (13), Anderson Rocha, coordenador do Recod.ai e professor do Instituto de Computação –...

VEJA MAIS

Mais notícias

VER TODAS

NOTÍCIA

Recod.ai na ONU

Nesta quinta-feira (13), Anderson Rocha, coordenador do Recod.ai e professor do Instituto...

ARTIGOS

Recod.ai na IEEE Transactions on Pattern Analysis and Machine Intelligence

Pesquisa do Recod.ai,

OPORTUNIDADE

Oportunidade: Pesquisador em IA

Sobre a vagaO Recod.ai abre chamada para...

OPORTUNIDADE

Pós-Doutorado: Aprendizado de Máquina

Sobre a vagaO Recod.ai abre

OPORTUNIDADE

Pós-Doutorado: Engenharia de Petróleo

Sobre a vagaO Recod.ai abre


Obrigado!
https://br.pinterest.com/pin/473582335650000265/