IBC2024

#ACCELERATORS2024

DESIGN YOUR WEAPONS IN THE FIGHT AGAINST DISINFORMATION

15 SEPTEMBER 2024

Associate sponsors: AMD × hp together we advance_

IBC2024

#ACCELERATORS2024

ACCELERATOR PROJECT:
THE FIGHT AGAINST DISINFORMATION

Champions

ALJAZEERA

AP
The Associated Press
Advancing the Power of Facts

BBC

4

CBC Radio-Canada

globo

CBS NEWS AND STATIONS

EBU
OPERATING EUROVISION AND EURORADIO

IET The Institution of Engineering and Technology

etc ENTERTAINMENT TECHNOLOGY CENTER

yle

2G DIGITAL

ITN

Medieklyngen
Media Cluster Norway

Paramount

Reuters
Imagen

TRANSMIXR

TV2 Danmark

Participants

ELUV.IO

HUMAN DIGITAL

NAGRA
KUDELSKI

OpenOrigins

VIDENTIFIER

TrueMedia.org

Design Your Weapons in the Fight Against Disinformation

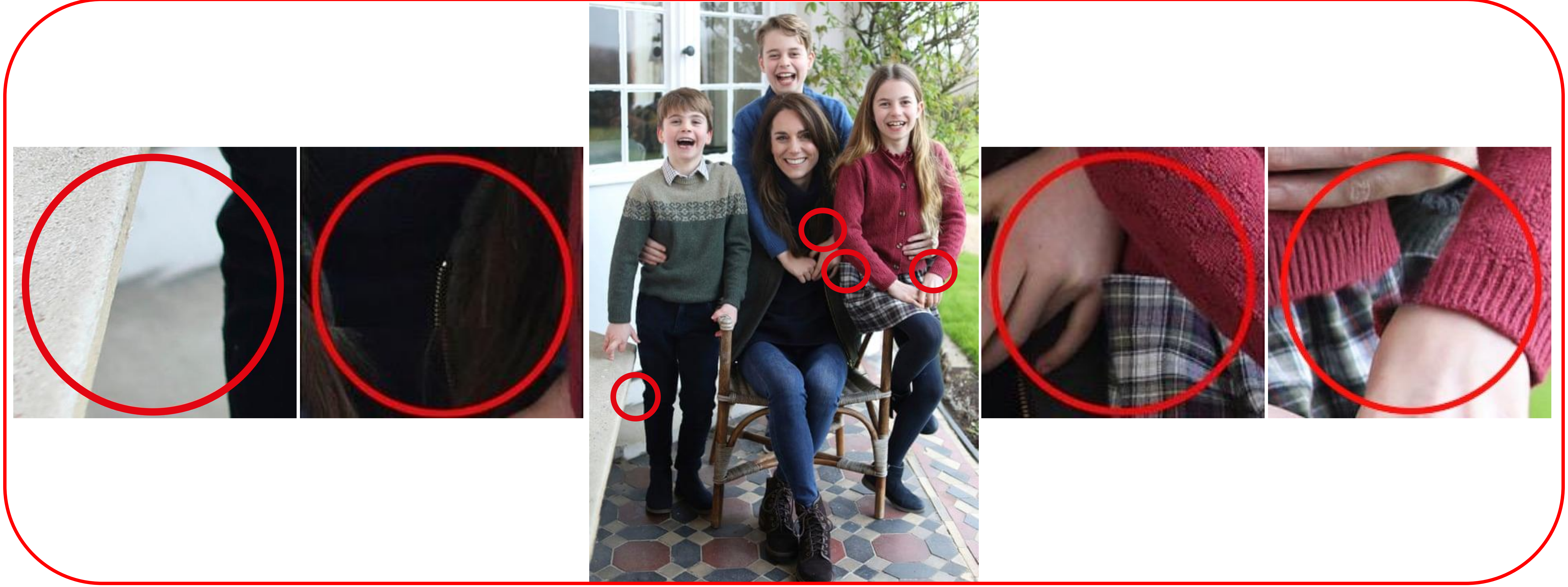
The Problem

Is it true?

Is it real?

Manipulation is everywhere

The Problem

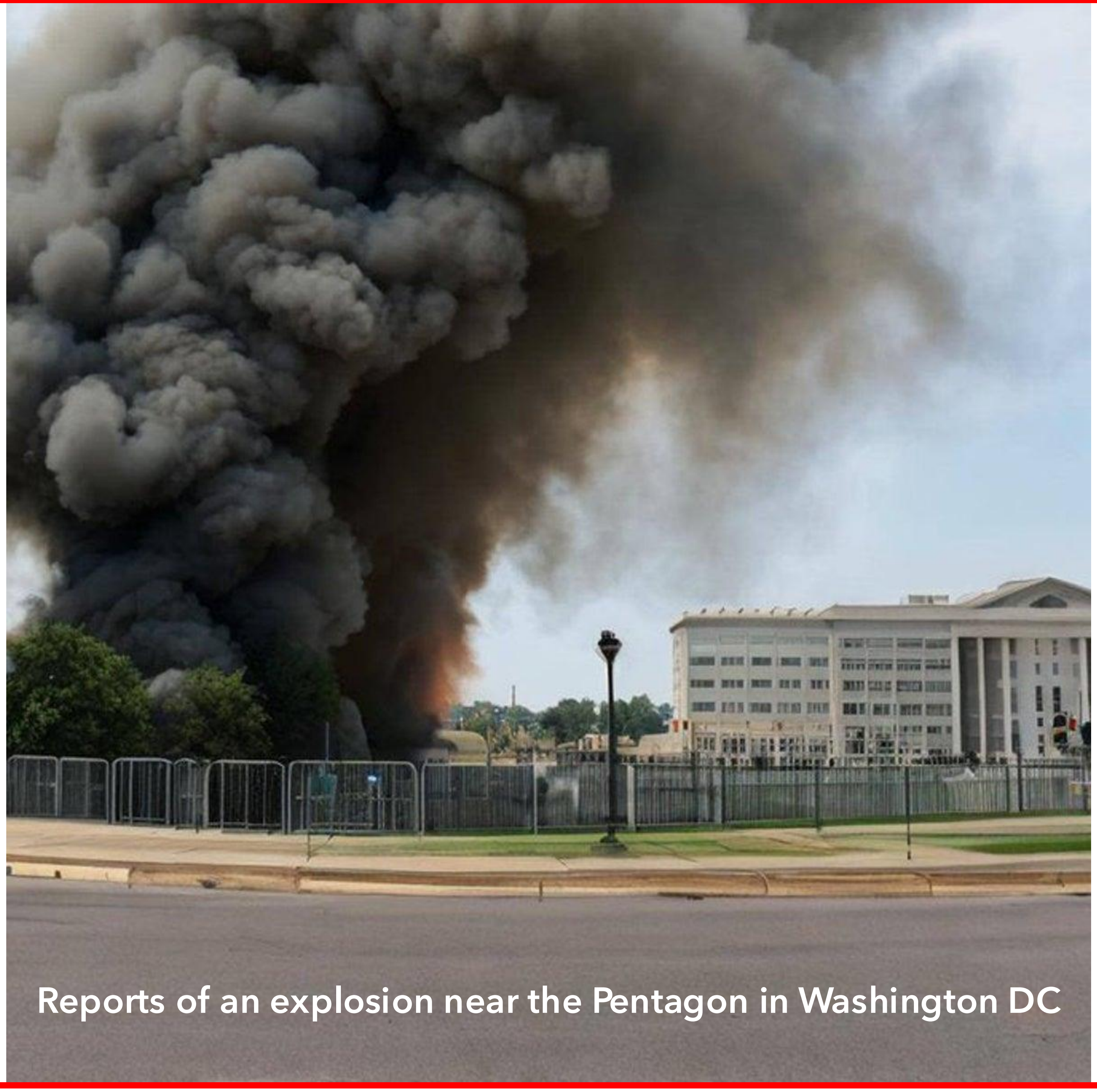


Fakes can be believed

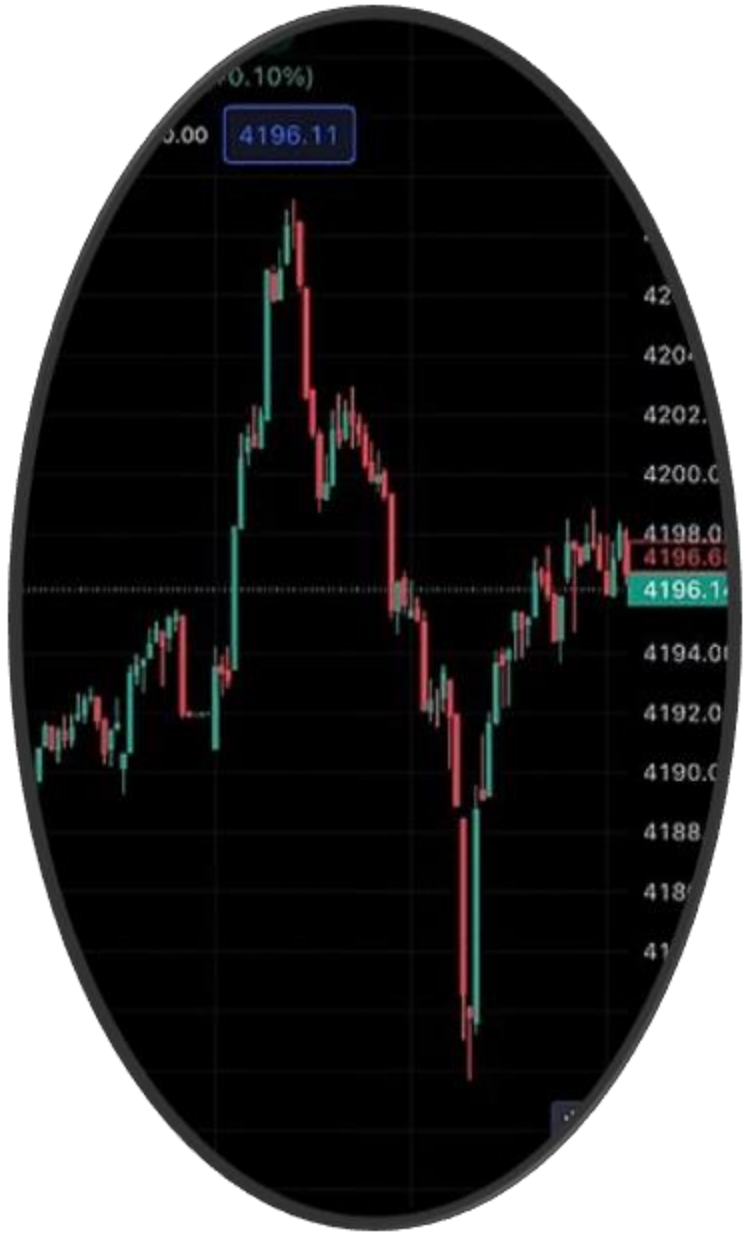
The Problem

“The S&P 500 briefly dropped… as social media accounts… repeated the false claims.”

AP, May 23 2023



Reports of an explosion near the Pentagon in Washington DC



Fakes can be believed

The Problem

WORLD

Deepfake targets Ukraine's first lady Olena Zelenksa with false claim she bought Bugatti

CBS News, July 2024

Participants could only verbally identify deepfakes 37% of the time

Source: University of Sydney

34 million AI-generated images are made each day Source: Everypixel

'Deepfakes' of Michael Mosley and Hilary Jones being used to promote scams on social media

Sky News, July 2024

AI and deepfakes blur reality in India elections

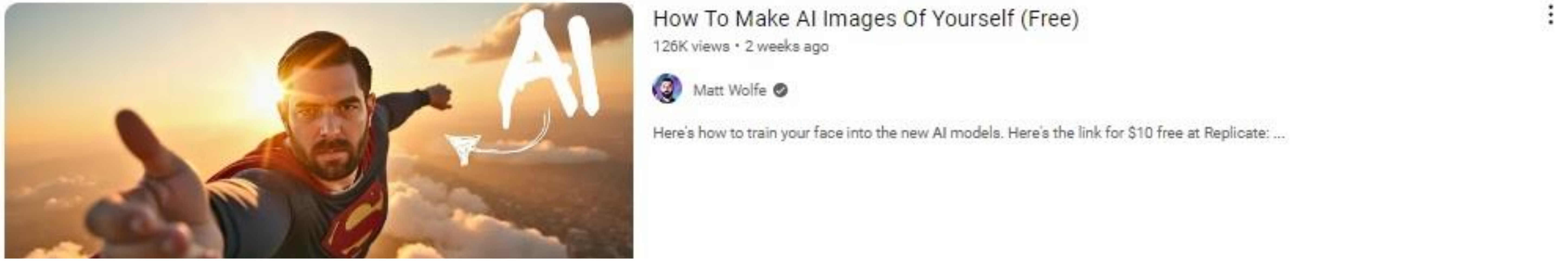
BBC, May 2024

Deepfake clips of Gareth Southgate swearing after England match go viral

The Guardian, July 2024

The technology is getting better and easier

The Problem



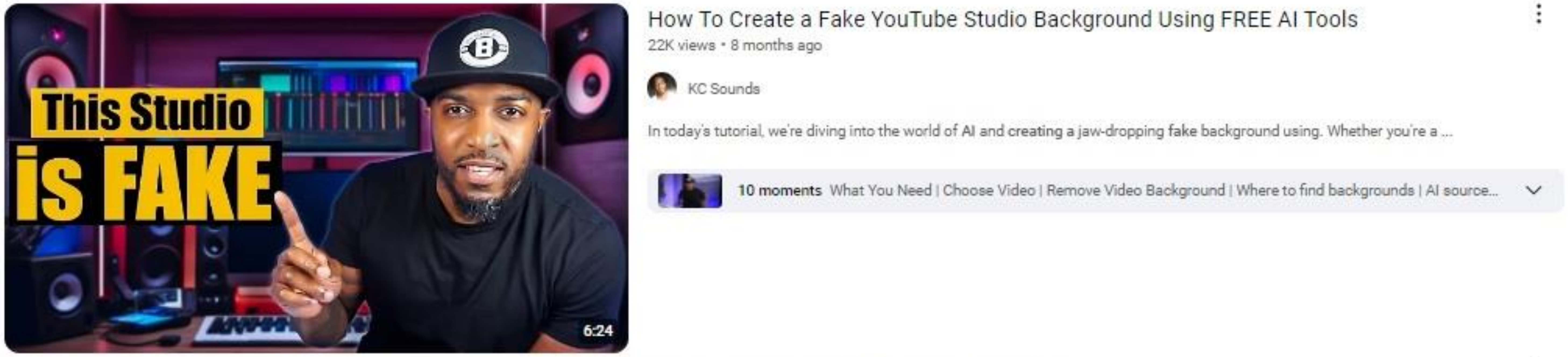
How To Make AI Images Of Yourself (Free)

126K views • 2 weeks ago

Matt Wolfe

Here's how to train your face into the new AI models. Here's the link for $10 free at Replicate: ...

News needs to separate fact from fiction



How To Create a Fake YouTube Studio Background Using FREE AI Tools

22K views • 8 months ago

KC Sounds

In today's tutorial, we're diving into the world of AI and creating a jaw-dropping fake background using. Whether you're a ...

10 moments What You Need | Choose Video | Remove Video Background | Where to find backgrounds | AI source...

IBC2024

Design Your Weapons in the Fight Against Disinformation

#ACCELERATORS2024

ALJAZEERA

AP

BBC

The Associated Press
Advancing the Power of Facts

4

CBC Radio-Canada

yle

CBS NEWS AND STATIONS

etc
ENTERTAINMENT
TECHNOLOGY
CENTER

TV2 Danmark

EBU
OPERATING EUROVISION AND EURORADIO

IET The Institution of Engineering and Technology

globo

2G DIGITAL

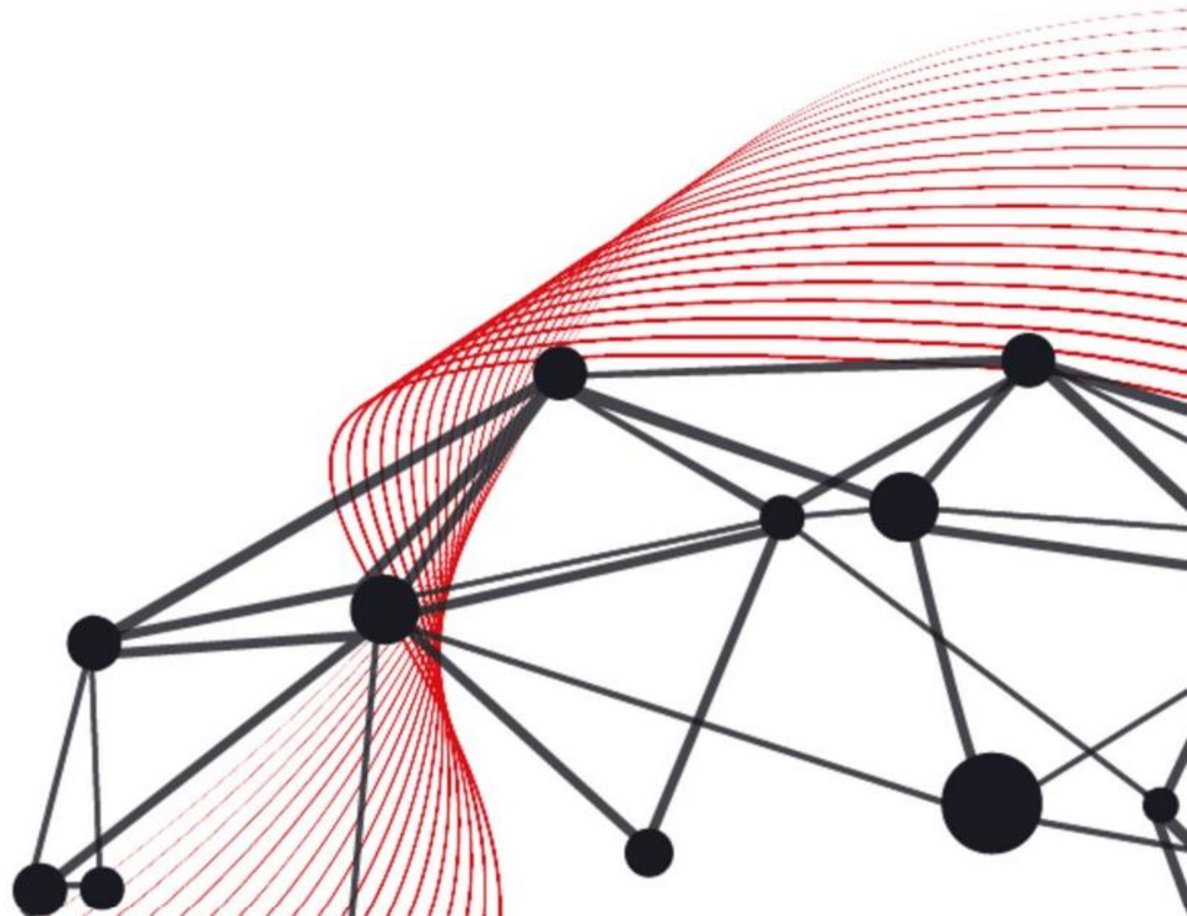
ITN

Medieklyngen
Media Cluster Norway

Paramount

Reuters Imagen

TRANSMIXR



Design Your Weapons in the Fight Against Disinformation

ELUV.IO

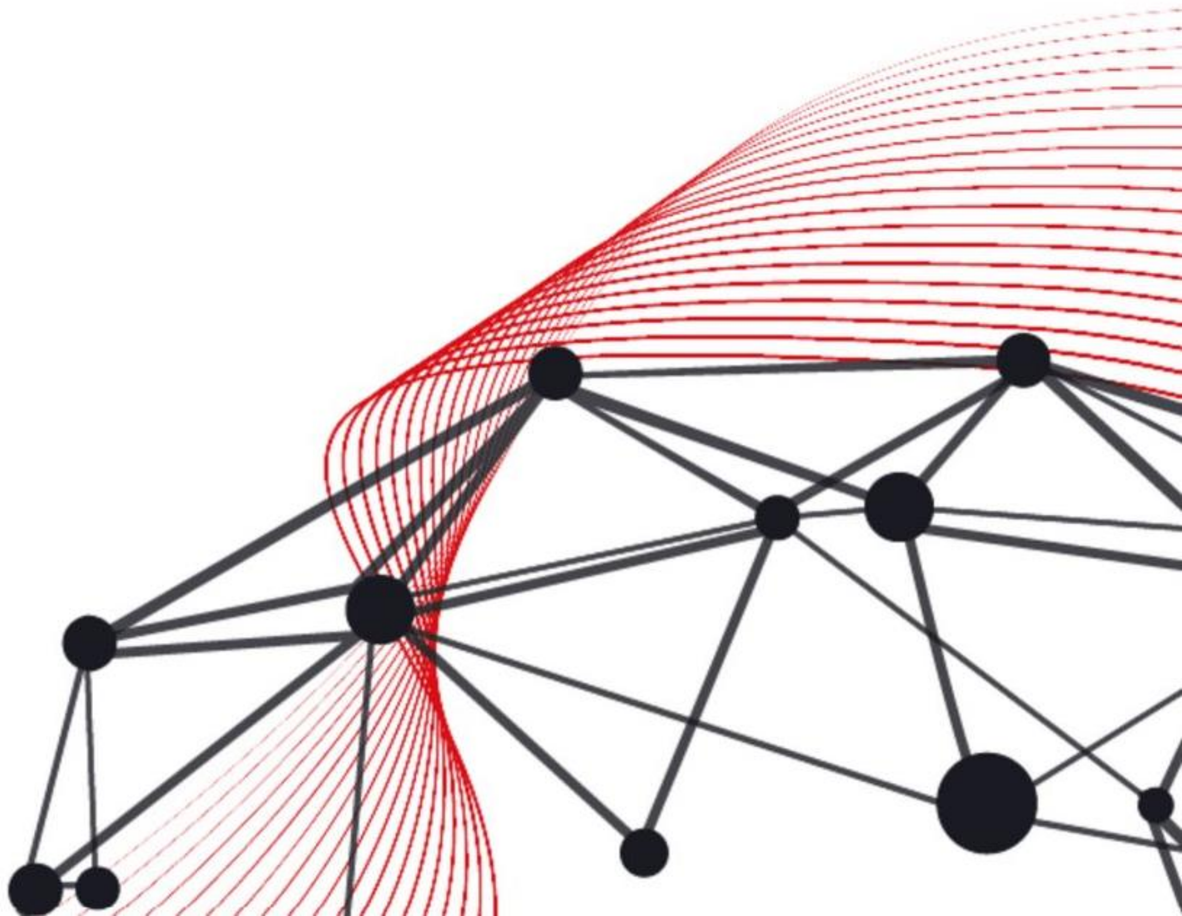
HUMAN DIGITAL

NAGRA KUDELSKI

OpenOrigins

TrueMedia.org

VIDENTIFIER



Design Your Weapons in the Fight Against Disinformation

Provenance

Detection

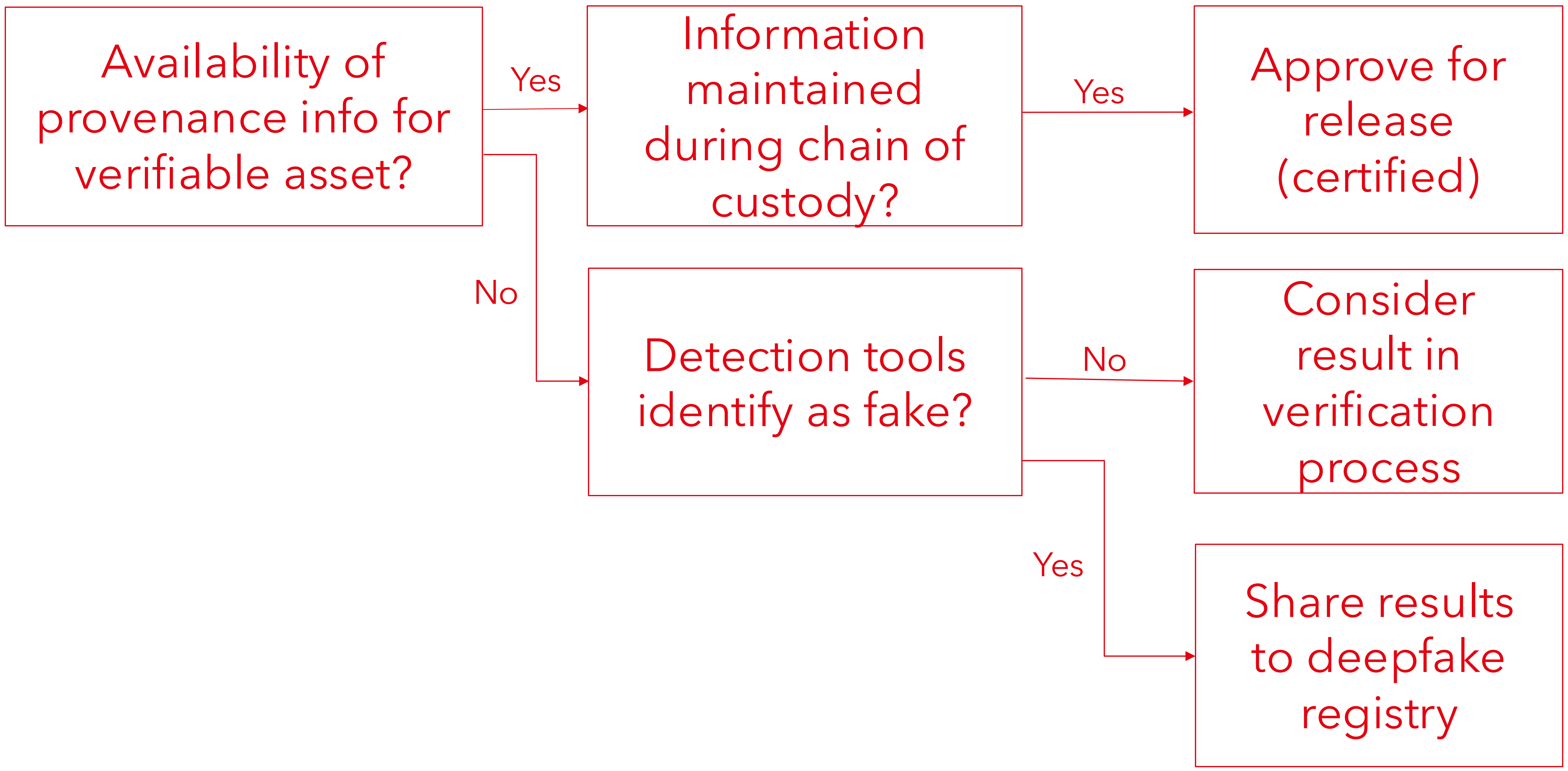
Collaboration

When to use different approaches?

Provenance

Detection

Collaboration



Availability of provenance info for verifiable asset?

Yes → Information maintained during chain of custody?

Yes → Approve for release (certified)

No → Detection tools identify as fake?

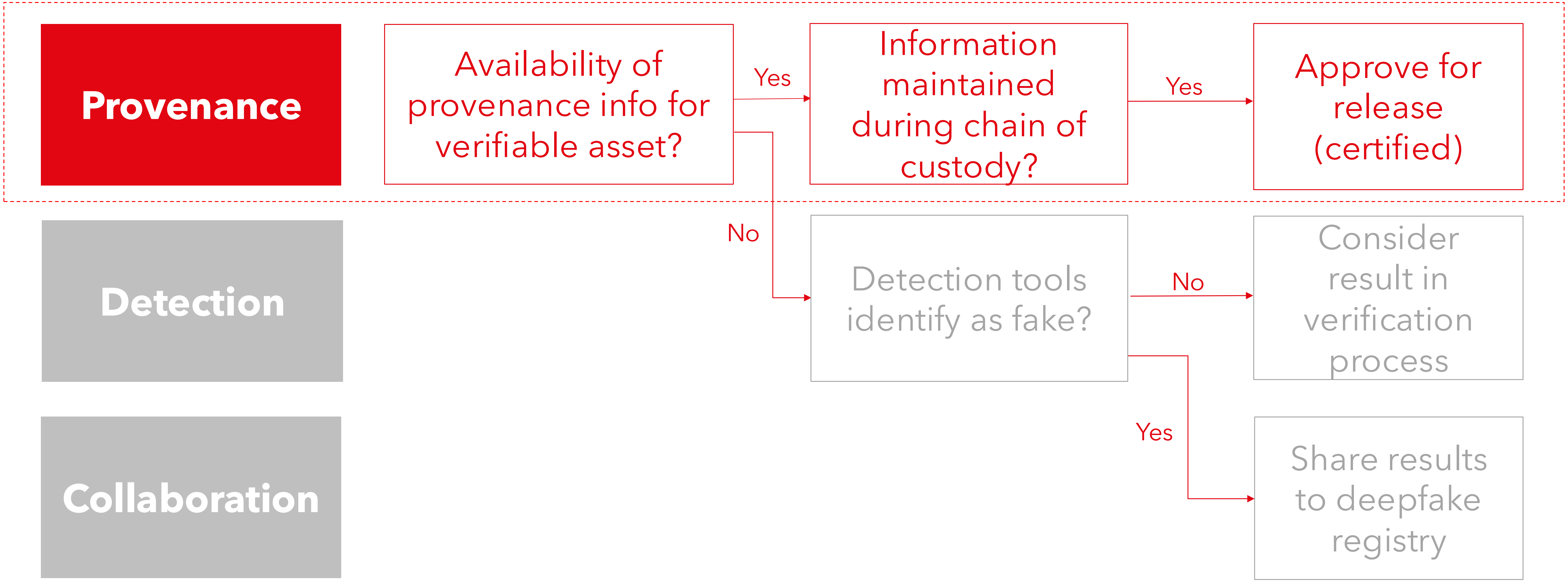
No → Consider result in verification process

Yes → Share results to deepfake registry

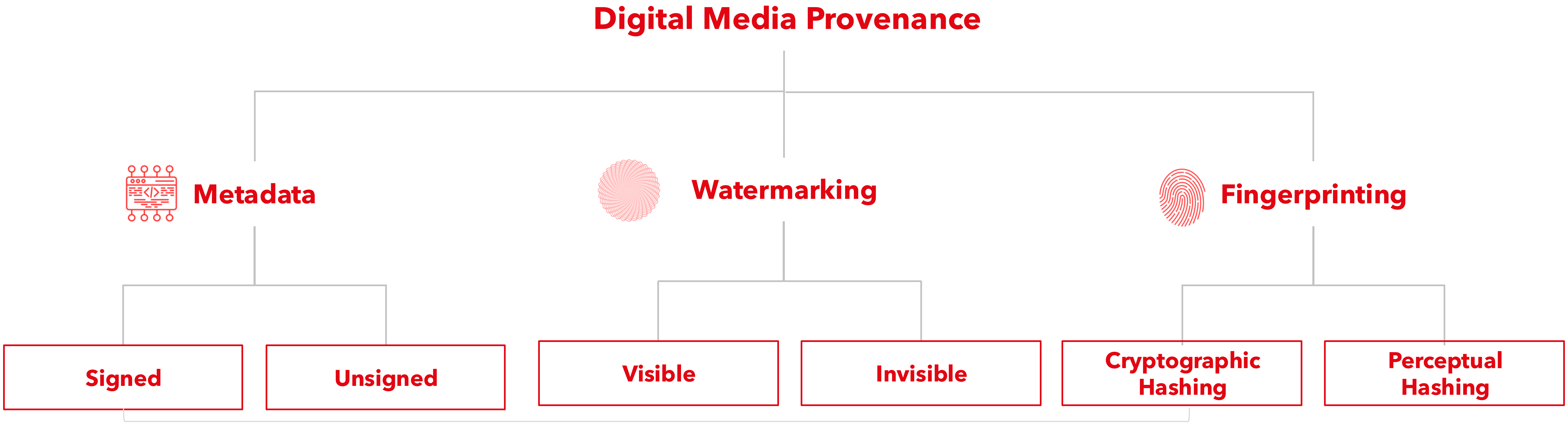
IBC2024

#ACCELERATORS2024

When to use provenance?


Provenance
Availability of provenance info for verifiable asset?
Yes
Information maintained during chain of custody?
Yes
Approve for release (certified)
No
Detection
Detection tools identify as fake?
No
Consider result in verification process
Yes
Collaboration
Share results to deepfake registry

What is provenance, and why is it becoming so significant?



Digital Media Provenance

Metadata

- Signed
- Unsigned

Watermarking

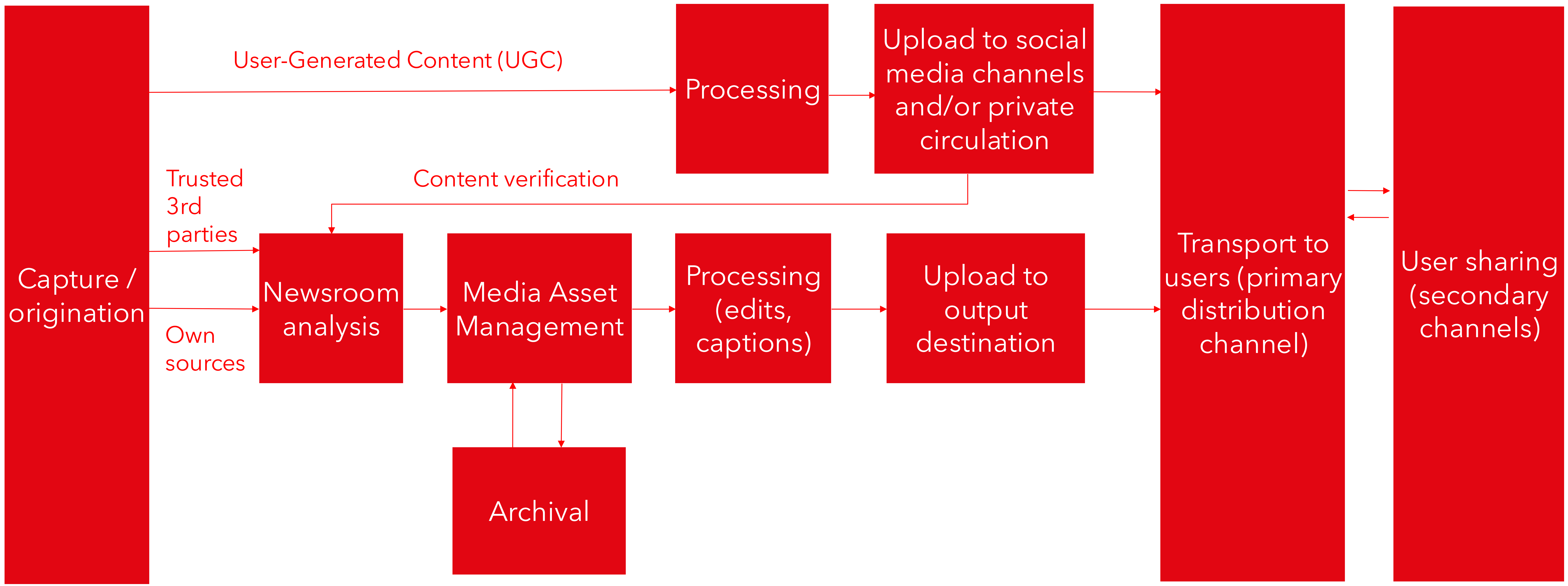
- Visible
- Invisible

Fingerprinting

- Cryptographic Hashing
- Perceptual Hashing

Provenance has impact throughout the media lifecycle

Timeframe: *Seconds to days* *May continue for years*



Capture / origination

User-Generated Content (UGC)

Processing

Upload to social media channels and/or private circulation

Content verification

Trusted 3rd parties

Own sources

Newsroom analysis

Media Asset Management

Archival

Processing (edits, captions)

Upload to output destination

Transport to users (primary distribution channel)

User sharing (secondary channels)

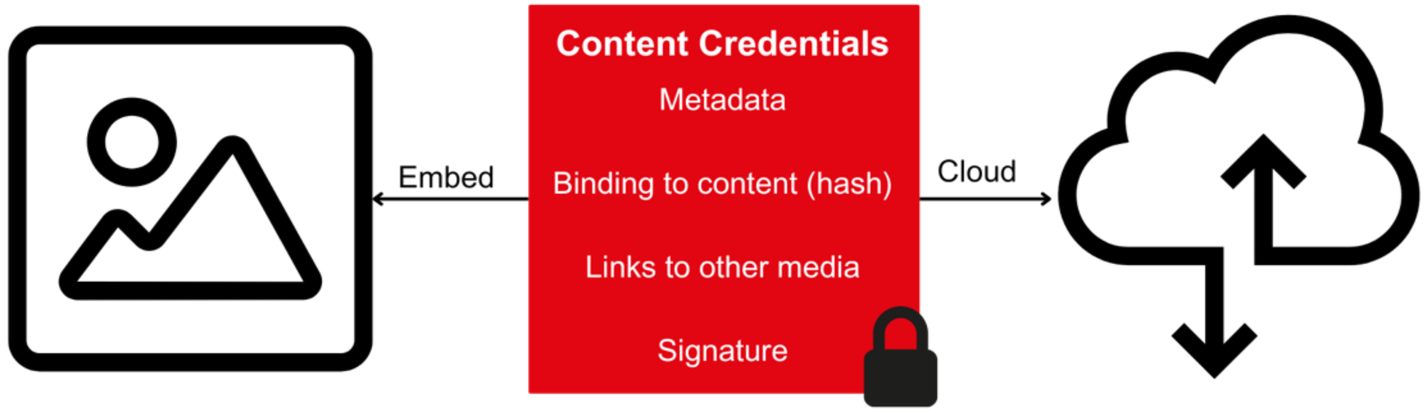
Who does provenance benefit?

		...and how?
Content Creators	e.g. trusted photojournalists, UGC	• Proving authenticity of their work • Tracking content usage (copyright)
Journalists	e.g. newsrooms, editors, archivists	• Rapid & accurate verification by newsrooms • Safeguarding information on origins (archival content) • Tracking and legitimising edits and captions
Viewers	e.g. audiences, copyright owners	• Audience understanding of origin point • Tracking content ownership & authorship

Standards and solutions are complementary

Provenance Standards (e.g. C2PA)

- Protocols and guidelines that define how provenance data should be captured
- C2PA is a widely adopted standard that uses cryptographic hashing to securely bind metadata to the content, ensuring authenticity and integrity.



Content Credentials
Metadata
Binding to content (hash)
Links to other media
Signature

Provenance Solutions

- The technical systems and tools that actually implement provenance tracking, ranging from proprietary systems to open-source frameworks
- A range of tools and solutions exist in the market, tailored to specific needs and workflows, offering diverse approaches to ensuring the authenticity and integrity of digital content.

ELUV.IO
TRANSMIXR
OpenOrigins
HUMAN DIGITAL
VIDENTIFIER
NAGRA KUDELSKI

Solutions ecosystem consists of a range of contributing technologies

ELUV.IO

Next-generation content management and distribution solution, utilizing blockchain technology and cryptographic hashing to allow users to control access and ensure monetization for their content.

OpenOrigins

End-to-end provenance solution, providing tools for authenticated media capture, archival media protection, and verification of online content. Secures provenance information in a tamper-proof blockchain to provide a decentralised trust infrastructure.

VIDENTIFIER

Automated visual matching and linking of content, using advanced visual fingerprinting for accuracy and speed. Enables users to verify content matching trusted sources and to detect changesmatching made.

TRANSMIXR

Consortium developing the foundation for the future of visual content, including virtual and augmented reality, with focus on the future of newsroom content creation and distribution.

HUMAN DIGITAL

Interoperable talent identifier, enabling ID resolution and provenance verification of notable real (human) public figures, their connected digital replicas, and fictional characters in the media supply chain. A DOI Foundation registration agency (ISO 26324.2022). Member: C2PA, IPTC.

NAGRA KUDELSKI

Content provenance and authenticity solutions based on forensic watermarking and fingerprinting, enabling users to distribute, protect and monetize their content.

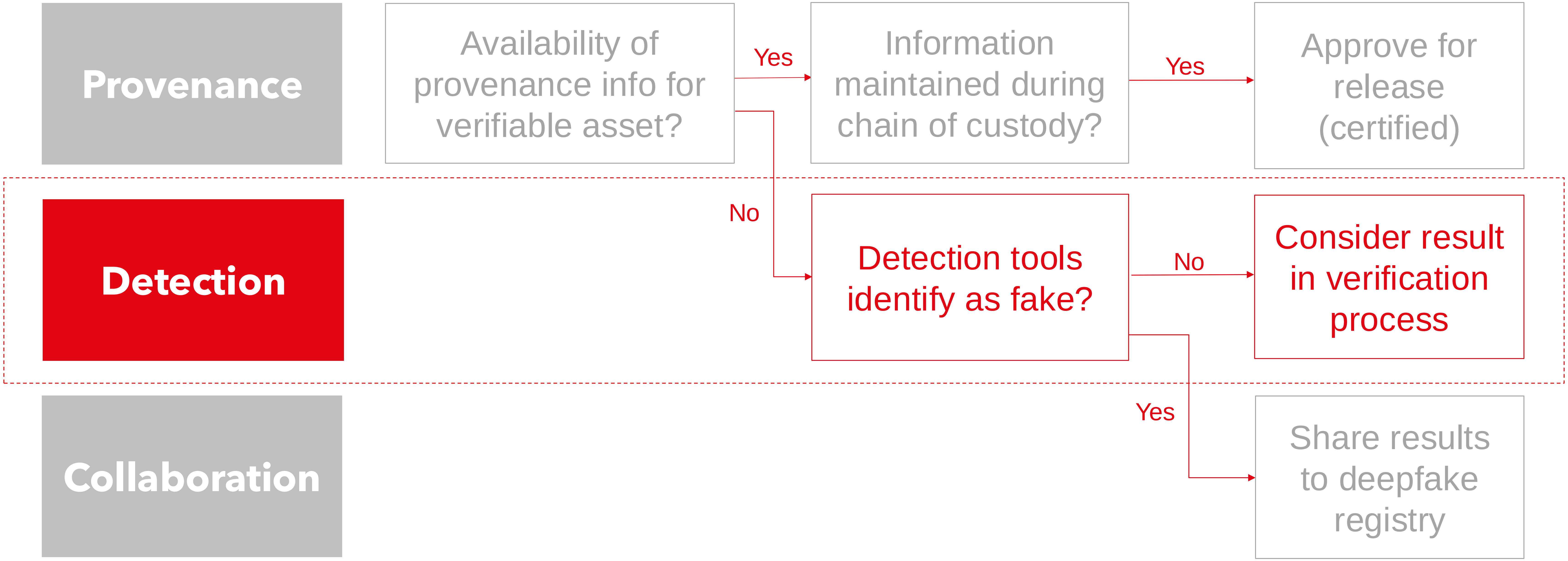
What would an ideal scenario look like?

- **Verifiable provenance metadata** embedded into every asset
- **Clear display of provenance information** (or lack of) at every stage of content distribution
- **An educated audience** with a zero-trust mindset
- **Provenance is accessible** with no barriers to entry

IBC2024

#ACCELERATORS2024

When to use provenance and/or detection?



Provenance
Availability of provenance info for verifiable asset?
Yes
Information maintained during chain of custody?
Yes
Approve for release (certified)
No
Detection
Detection tools identify as fake?
No
Consider result in verification process
Yes
Collaboration
Share results to deepfake registry

Detection may be leveraged when provenance data is unavailable

- **Detection** is the leveraging of visual analysis and machine learning to **automate detection of manipulated visuals, narrative or wholly synthetically-generated content**.
- Newsrooms and end users can benefit from the ability to detect Generative AI across various media formats, especially when provenance information can't be accessed.

Challenges in detection today:

- Generative AI is developing at a faster rate than detection modules.
- Multi-layered editing of AI-generated content can impact the efficacy of detection modules.
- No single solution available to address all issues, with a risk of false results and inaccuracies.

A combinations of detection tools can inform various elements of the verification process, acting as an aid rather than an alternative to human-led content validation.

Unverified Source

Scan against existing authenticated database

- Identify duplicates of authenticated original content and assess potential IP implications
- Identify manipulated content based off authenticated original media

Scan against collaborative database of problematic content

- Identify common disinformation topics or bad actors
- Collaboration or Machine led curation can improve reaction speed and help with allocation of resources

Visual analysis to identify manipulated or wholly AI-generated content

- Newsrooms are currently mostly reliant on manual checks and contextual verification
- Automated solutions exist but can struggle to keep up with the advances in Generative-AI

See Project Demo Video

(The Principles of Accelerator Programme for Provenance and Detection)

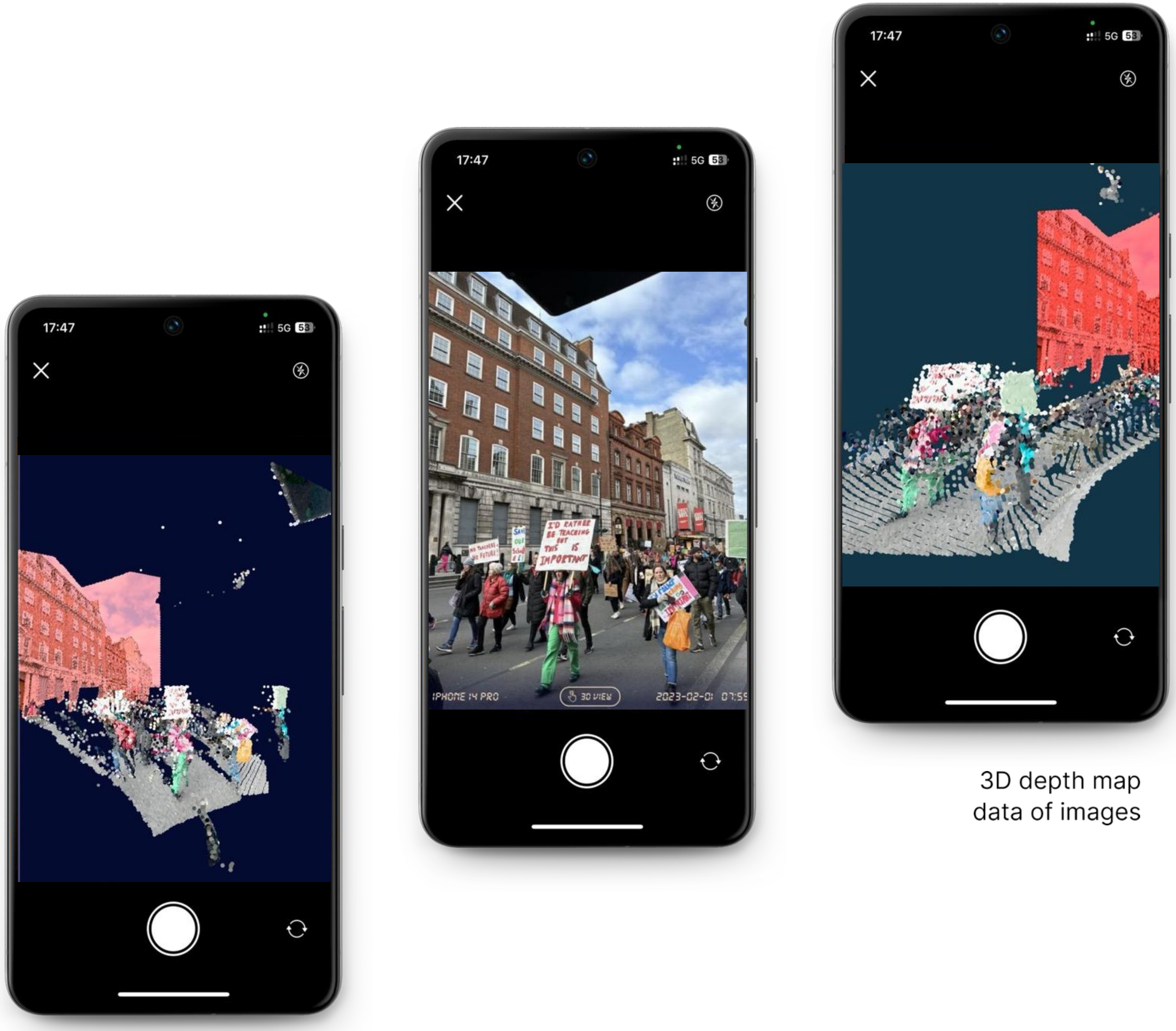
Provenance and Detection in action

OpenOrigins

Secure capture of new media

EXAMPLE: USER-GENERATED CONTENT

- **Metadata & unique fingerprint** captured
- Hardware verification ensures that the **capture device is trusted**
- **LiDAR & sensor data** captured along with the image
- Stored on a **tamper-proof blockchain**



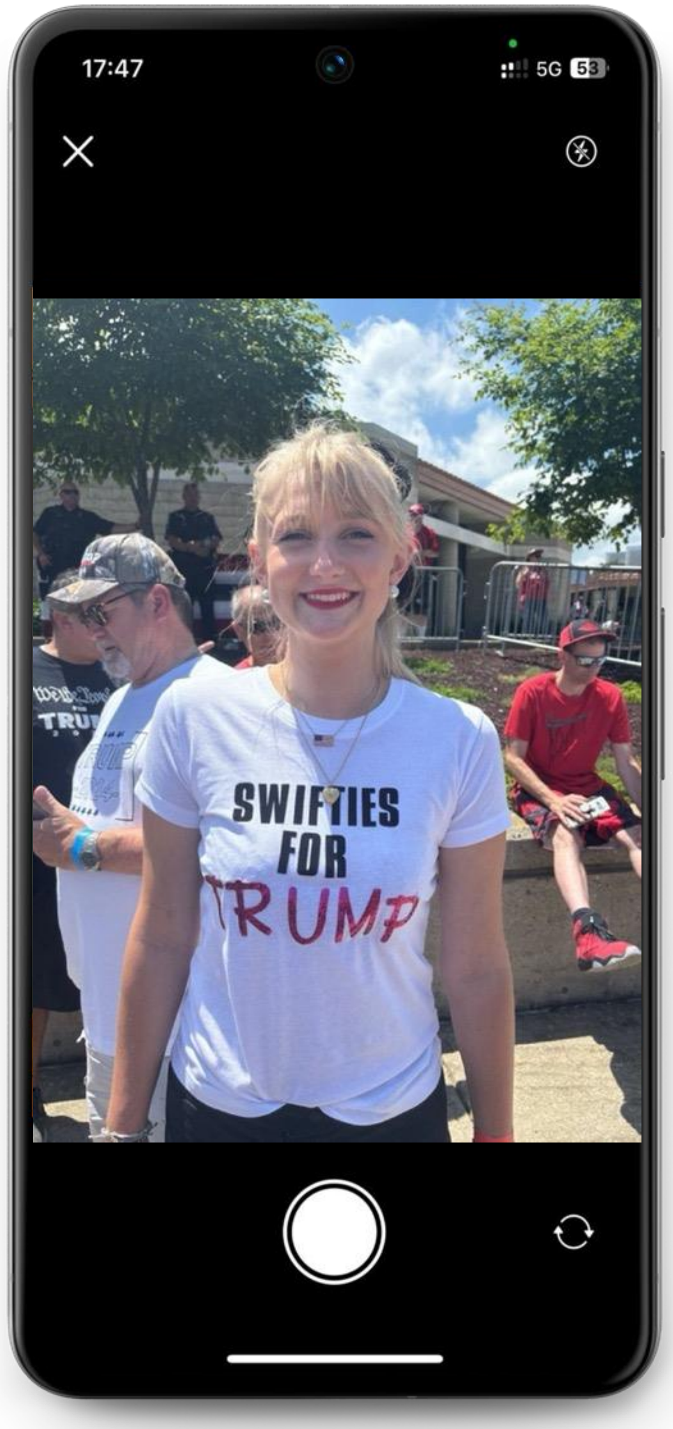
3D depth map data of images

OpenOrigins

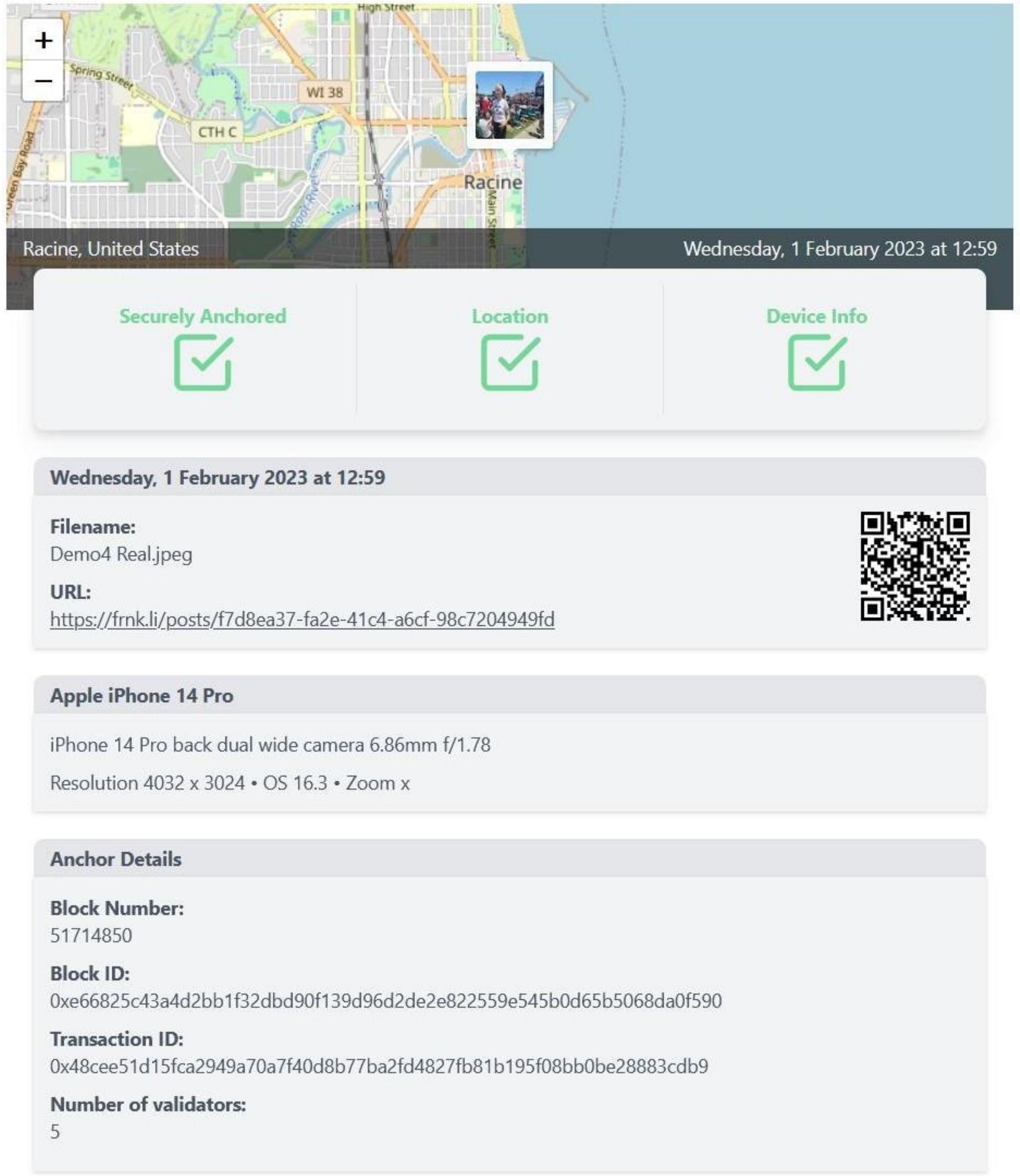
Newsroom analysis

EXAMPLE: USER-GENERATED CONTENT

- Photo editor can look up proof information and understand:
  - Image has **verifiable origins**
  - Taken on a **trusted device**
  - Taken at **correct time & location**



Everyday photographer shares photo with journalist



Racine, United States

Wednesday, 1 February 2023 at 12:59

Securely Anchored ☑

Location ☑

Device Info ☑

Wednesday, 1 February 2023 at 12:59

Filename:
Demo4 Real.jpeg

URL:
https://frnk.li/posts/f7d8ea37-fa2e-41c4-a6cf-98c7204949fd

Apple iPhone 14 Pro

iPhone 14 Pro back dual wide camera 6.86mm f/1.78

Resolution 4032 x 3024 • OS 16.3 • Zoom x

Anchor Details

Block Number:
51714850

Block ID:
0xe66825c43a4d2bb1f32dbd90f139d96d2de2e822559e545b0d65b5068da0f590

Transaction ID:
0x48cee51d15fca2949a70a7f40d8b77ba2fd4827fb81b195f08bb0be28883cdb9

Number of validators:
5

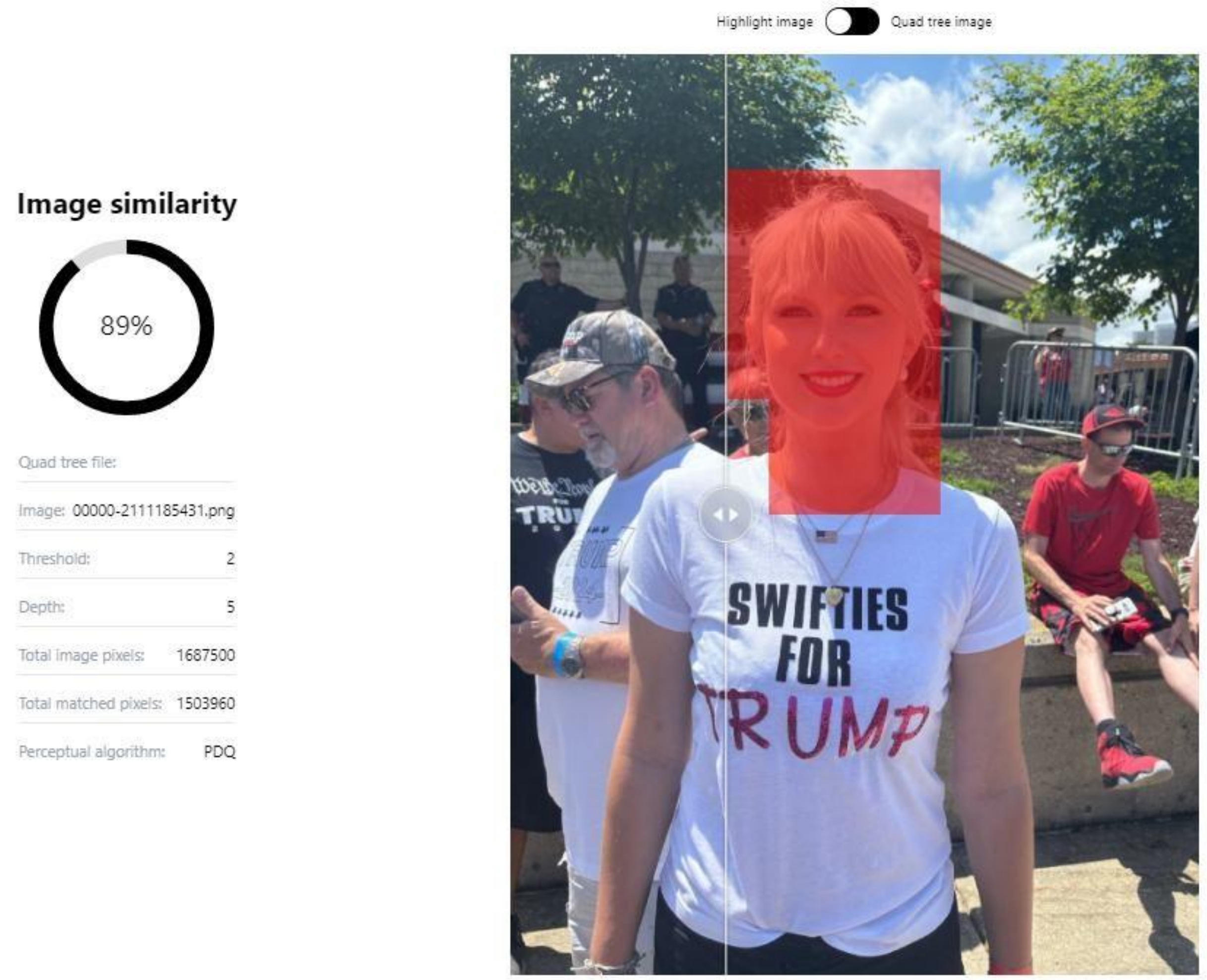
Photo editor looks up origin information

OpenOrigins

Transport to viewer

EXAMPLE: USER-GENERATED CONTENT

- Chrome plugin displays **green boxes around known verified content**
- Comparison tools allows viewers to check if **an image they see has been modified**



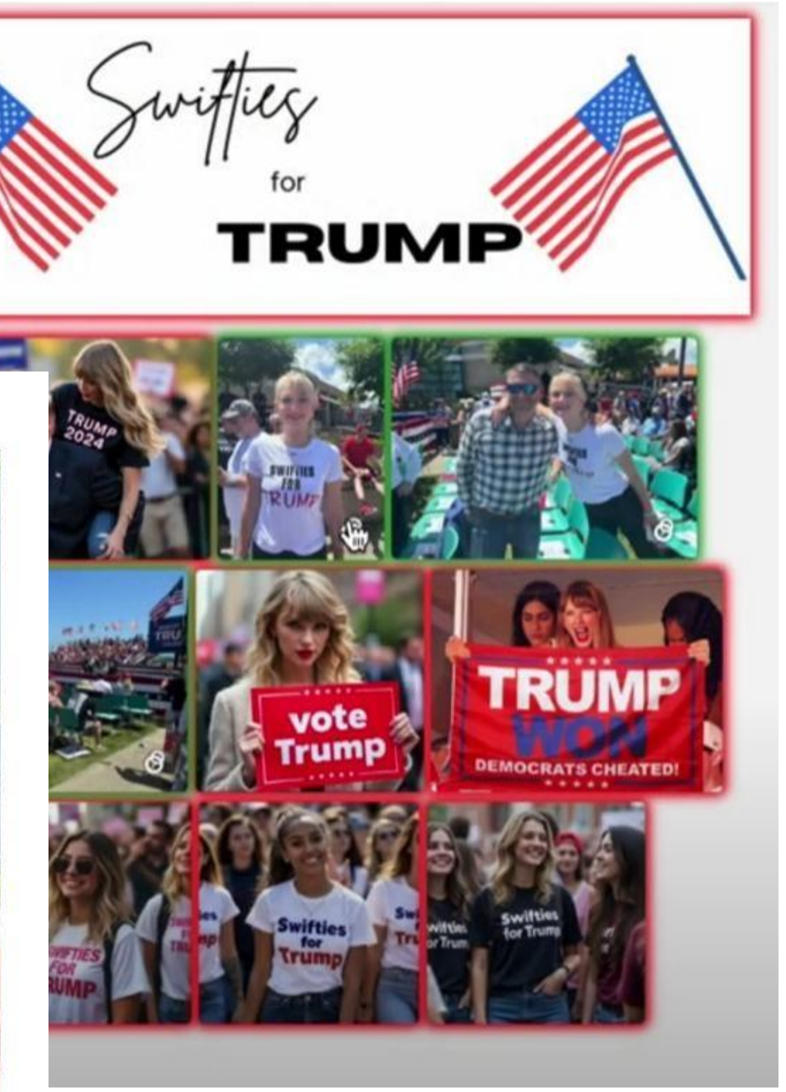
Highlight image ● Quad tree image

Image similarity

89%

Quad tree file:	
Image:	00000-2111185431.png
Threshold:	2
Depth:	5
Total image pixels:	1687500
Total matched pixels:	1503960
Perceptual algorithm:	PDQ

Comparsion tool for edited images

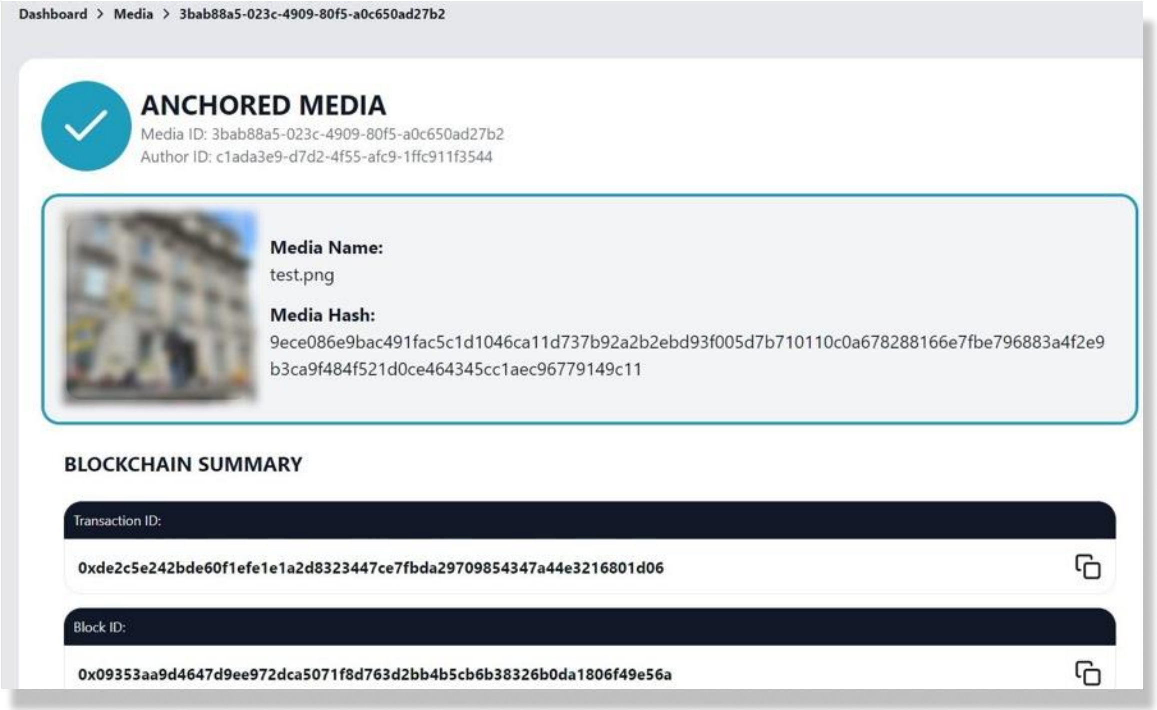


Mock-up of a disinformation site with verified and non-verified content

Archive Anchoring

SAFEGUARDING ARCHIVES WITH IMMUTABLE PROVENANCE, PRESERVING OUR WORLD'S SHARED HISTORY

- Deep database analytics to ensure the media is **validated and authentic.**
- Scalable architecture allows for the **securing of billions of media items** that already exist in archives.
- Without proper authentication, **real and fake become indistinguishable**



Dashboard > Media > 3bab88a5-023c-4909-80f5-a0c650ad27b2

ANCHORED MEDIA

Media ID: 3bab88a5-023c-4909-80f5-a0c650ad27b2

Author ID: c1ada3e9-d7d2-4f55-afc9-1ffc911f3544

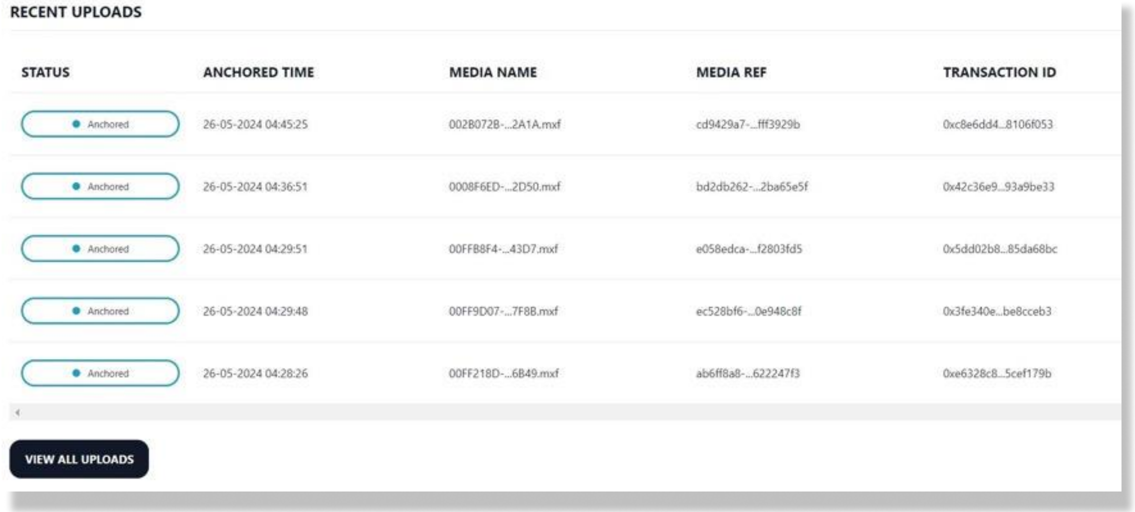
Media Name:
test.png

Media Hash:
9ece086e9bac491fac5c1d1046ca11d737b92a2b2ebd93f005d7b710110c0a678288166e7fbe796883a4f2e9b3ca9f484f521d0ce464345cc1aec96779149c11

BLOCKCHAIN SUMMARY

Transaction ID:
0xde2c5e242bde60f1efe1e1a2d8323447ce7fbda29709854347a44e3216801d06

Block ID:
0x09353aa9d4647d9ee972dca5071f8d763d2bb4b5cb6b38326b0da1806f49e56a

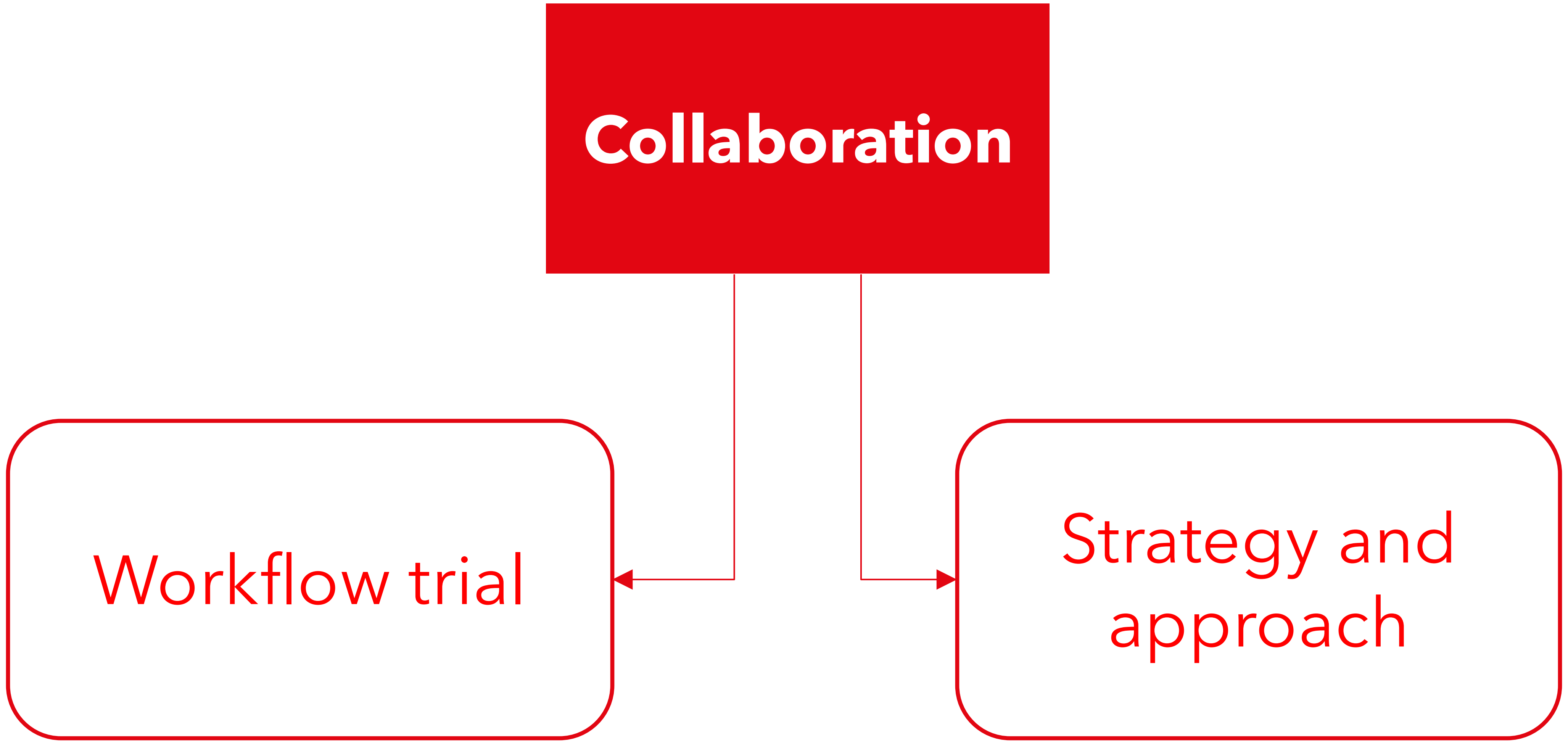


RECENT UPLOADS

STATUS	ANCHORED TIME	MEDIA NAME	MEDIA REF	TRANSACTION ID
● Anchored	26-05-2024 04:45:25	002B072B-...2A1A.mxf	cd9429a7-...fff3929b	0xc8e6dd4...8106f053
● Anchored	26-05-2024 04:36:51	0008F6ED-...2D50.mxf	bd2db262-...2ba65e5f	0x42c36e9...93a9be33
● Anchored	26-05-2024 04:29:51	00FFB8F4-...43D7.mxf	e058edca-...f2803fd5	0x5dd02b8...85da68bc
● Anchored	26-05-2024 04:29:48	00FF9D07-...7F8B.mxf	ec528bf6-...0e948c8f	0x3fe340e...be8cceb3
● Anchored	26-05-2024 04:28:26	00FF218D-...6B49.mxf	ab6ff8a8-...622247f3	0xe6328c8...5cef179b

Media organisation's view of secured media proofs for archival content

What does Collaboration look like?

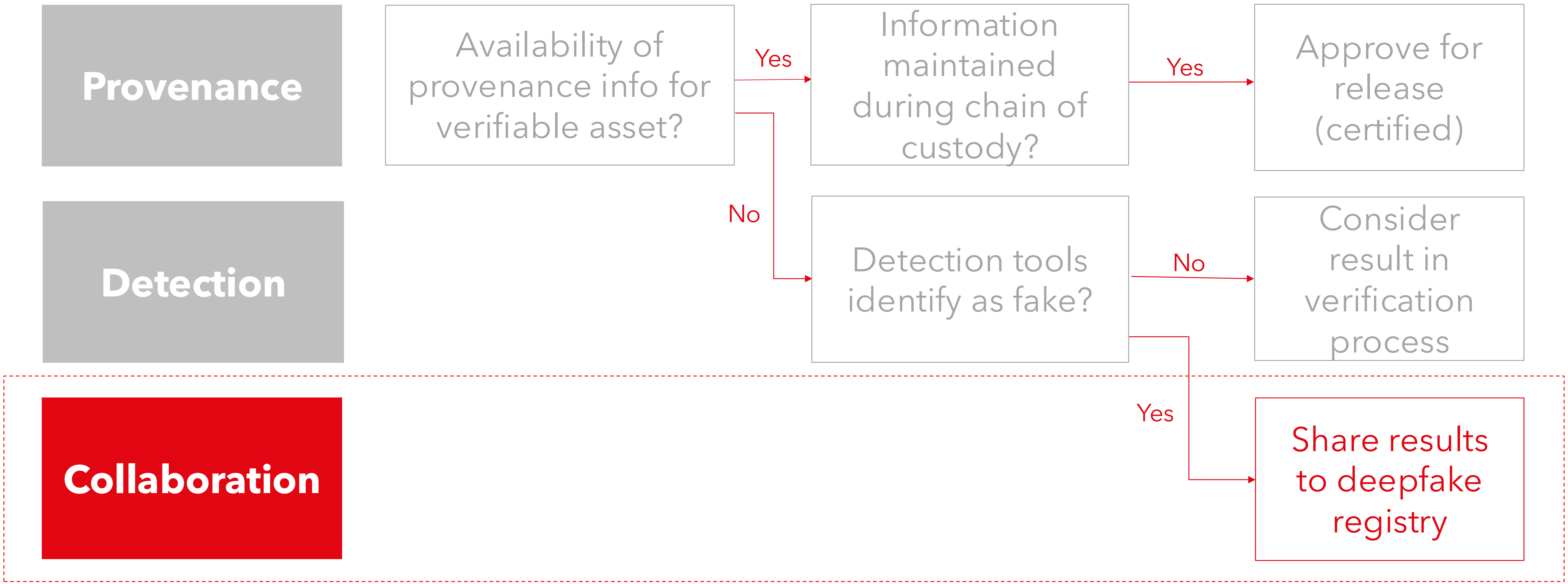


Collaboration

Workflow trial

Strategy and approach

Collaboration – Products and Workflow



Provenance

Detection

Collaboration

Availability of provenance info for verifiable asset?

Information maintained during chain of custody?

Approve for release (certified)

Detection tools identify as fake?

Consider result in verification process

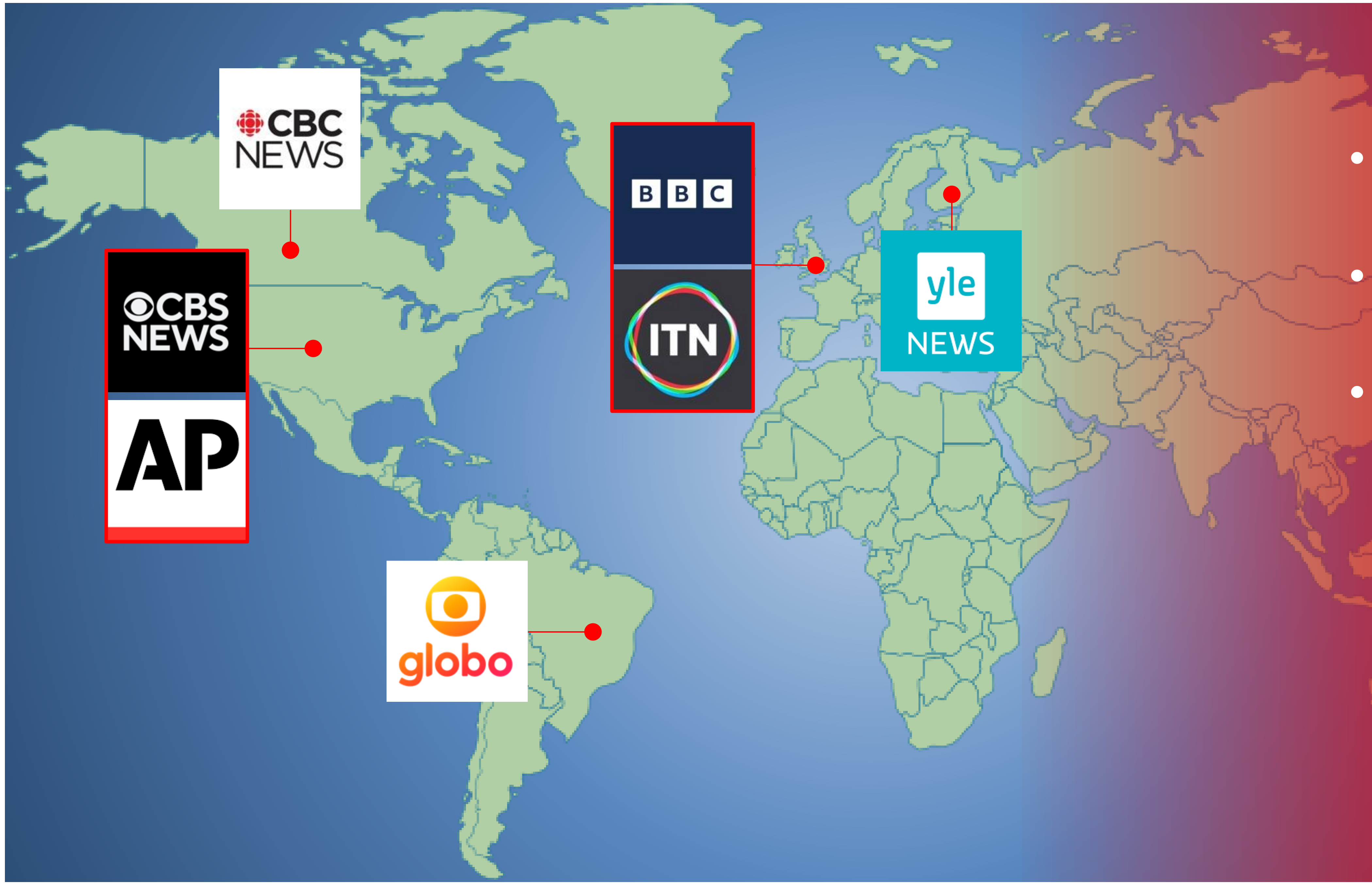
Share results to deepfake registry

IBC2024

#ACCELERATORS2024

Collaboration – Workflow Trial

Are newsrooms prepared to share knowledge that benefits everyone to produce faster and more informed reporting?

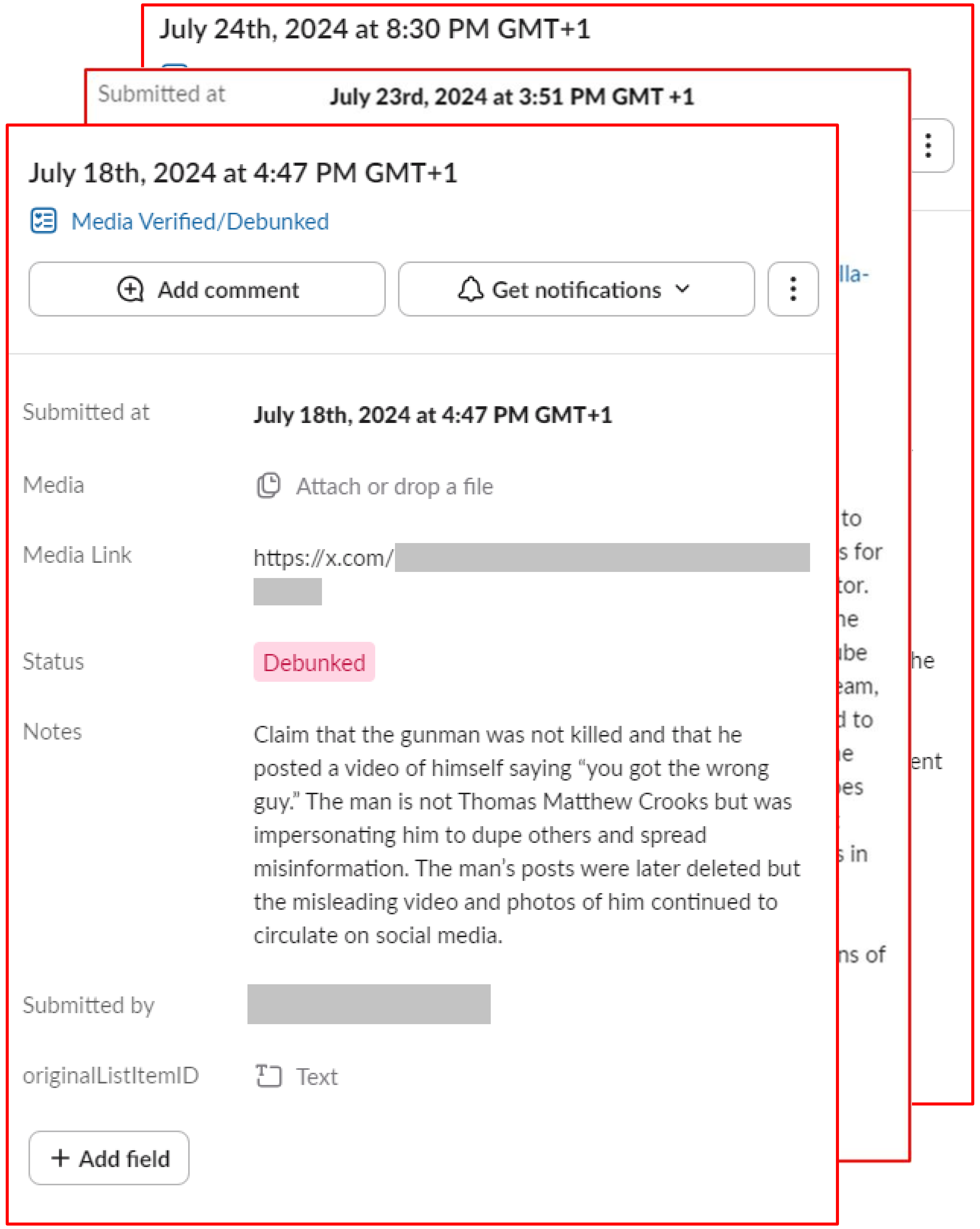


- 1 month
- 7 news organisations
- Shared information on suspect content

Tech support: Google

System development: TRANSMIXR

Collaboration – Workflow Trial



July 24th, 2024 at 8:30 PM GMT+1

Submitted at July 23rd, 2024 at 3:51 PM GMT +1

lla-

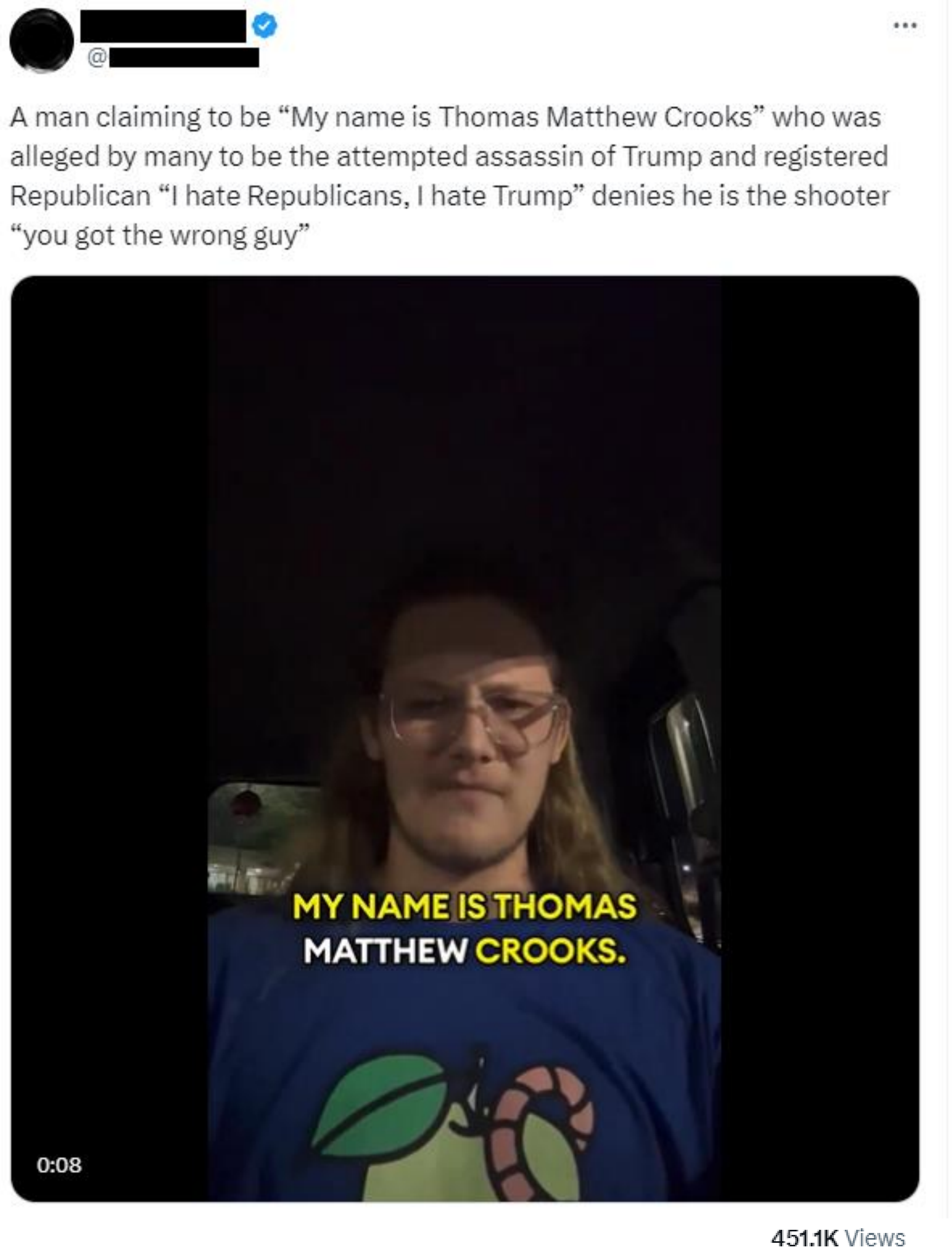
to
s for
tor.
he
ube
eam,
d to
e
es
s in
ns of

he
ent

July 18th, 2024 at 4:47 PM GMT+1

Media Verified/Debunked

Submitted at	July 18th, 2024 at 4:47 PM GMT+1
Media	Attach or drop a file
Media Link	https://x.com/[redacted]
Status	Debunked
Notes	Claim that the gunman was not killed and that he posted a video of himself saying "you got the wrong guy." The man is not Thomas Matthew Crooks but was impersonating him to dupe others and spread misinformation. The man's posts were later deleted but the misleading video and photos of him continued to circulate on social media.
Submitted by	[redacted]
originalListItemID	Text



[redacted]
@[redacted]

A man claiming to be "My name is Thomas Matthew Crooks" who was alleged by many to be the attempted assassin of Trump and registered Republican "I hate Republicans, I hate Trump" denies he is the shooter "you got the wrong guy"

451.1K Views

Collaboration – Workflow Trial

Two major stories during trial period

US Politics

UK Riots

and yet…

	Content shared
Week 1	10 stories
Week 2	5 stories
Week 3	3 stories
Week 4	6 stories

Collaboration – Workflow Trial

	Content shared
Week 1	10 stories
Week 2	5 stories
Week 3	3 stories
Week 4	6 stories

- **Fake content and disinformation spreads on social media** where it can gain huge traction without being debunked or taken down.

- **News organisations ignore most fake content.** There is too much to debunk it all.

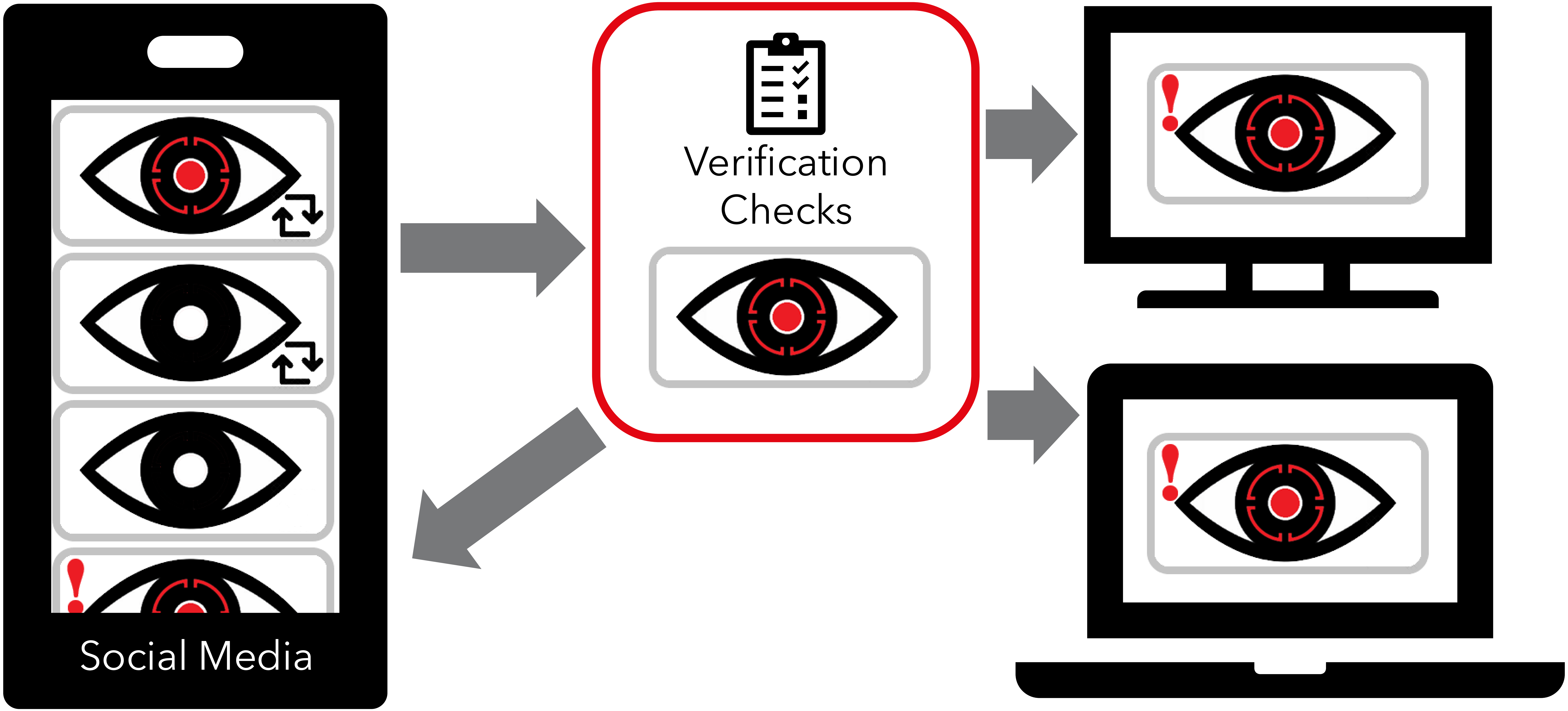
- **Journalists operate with caution.** Fake and genuine content is checked, taking time and allowing disinformation to spread through unverified means.

- **That means mis- and dis-information can spread without authoritative challenge,** causing confusion and affecting narratives.

IBC2024

#ACCELERATORS2024

Collaboration – Approach & Strategy



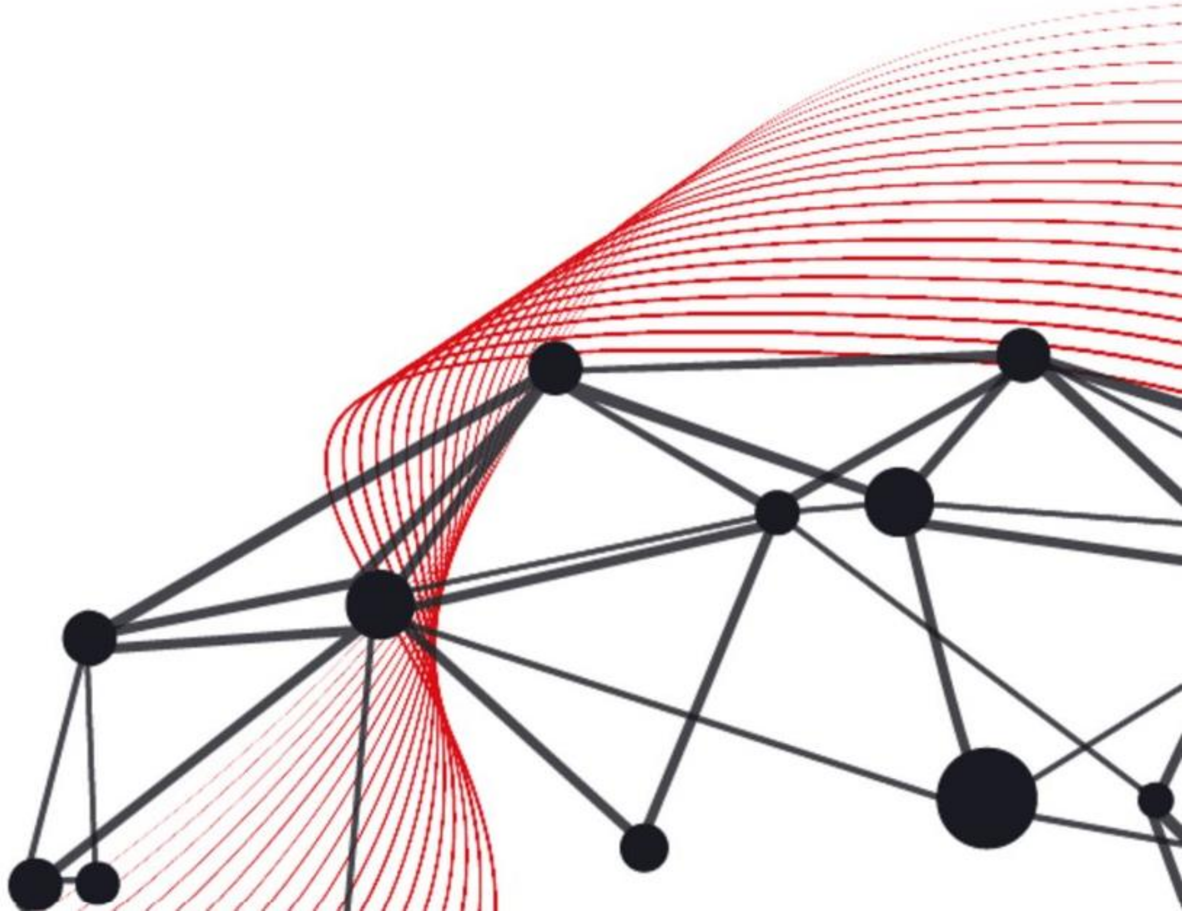
CALL TO ARMS!

- We all need to work together closely to make sure everyone has access to information they can trust
- We all need to develop a deeper understanding of the products and systems available to fight disinformation and fake content
- AI generated fake content is getting increasingly hard to detect and easier to produce.

CALL TO ARMS!

Join us on this journey

accelerators@ibc.org



IBC2024

#ACCELERATORS2024

DESIGN YOUR WEAPONS IN THE FIGHT AGAINST DISINFORMATION

15 SEPTEMBER 2024

Associate sponsors: AMD × hp together we advance_