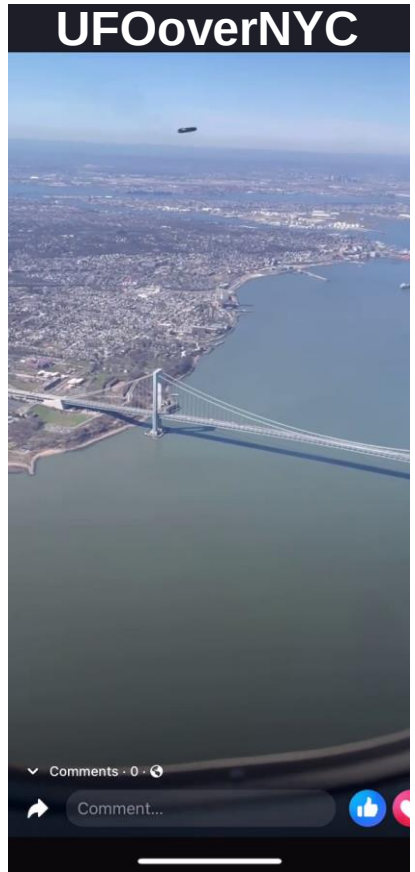
Data Integrity and Chain of Custody Architectures for Scientific UAP Studies

Dr. Doug Buettner
Stevens Institute of Technology
and the University of Utah

SOCIETY FOR UAP STUDIES

The Problem Statement: BIG Data Integrity Problem

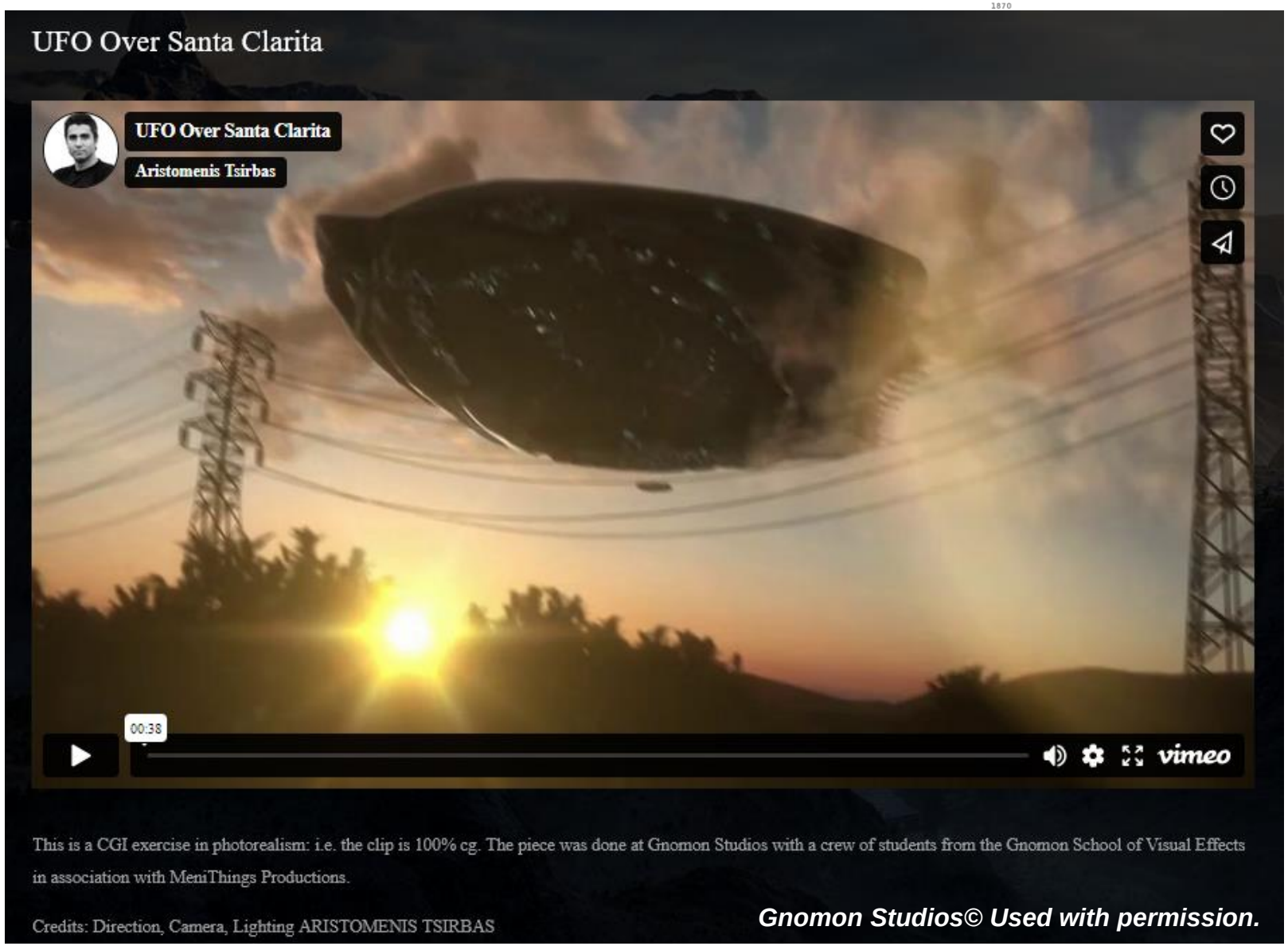
How is anyone supposed to know what images or videos of UAPs, aliens, cryptids, ghosts and other "Hi Strangeness" phenomena on the internet, social media or the plethora of "Reality TV-shows" is real or fake?




Faked: A Known CGI Example – Showing How Good CGI IS!
Gnomon Studios© Used with permission.

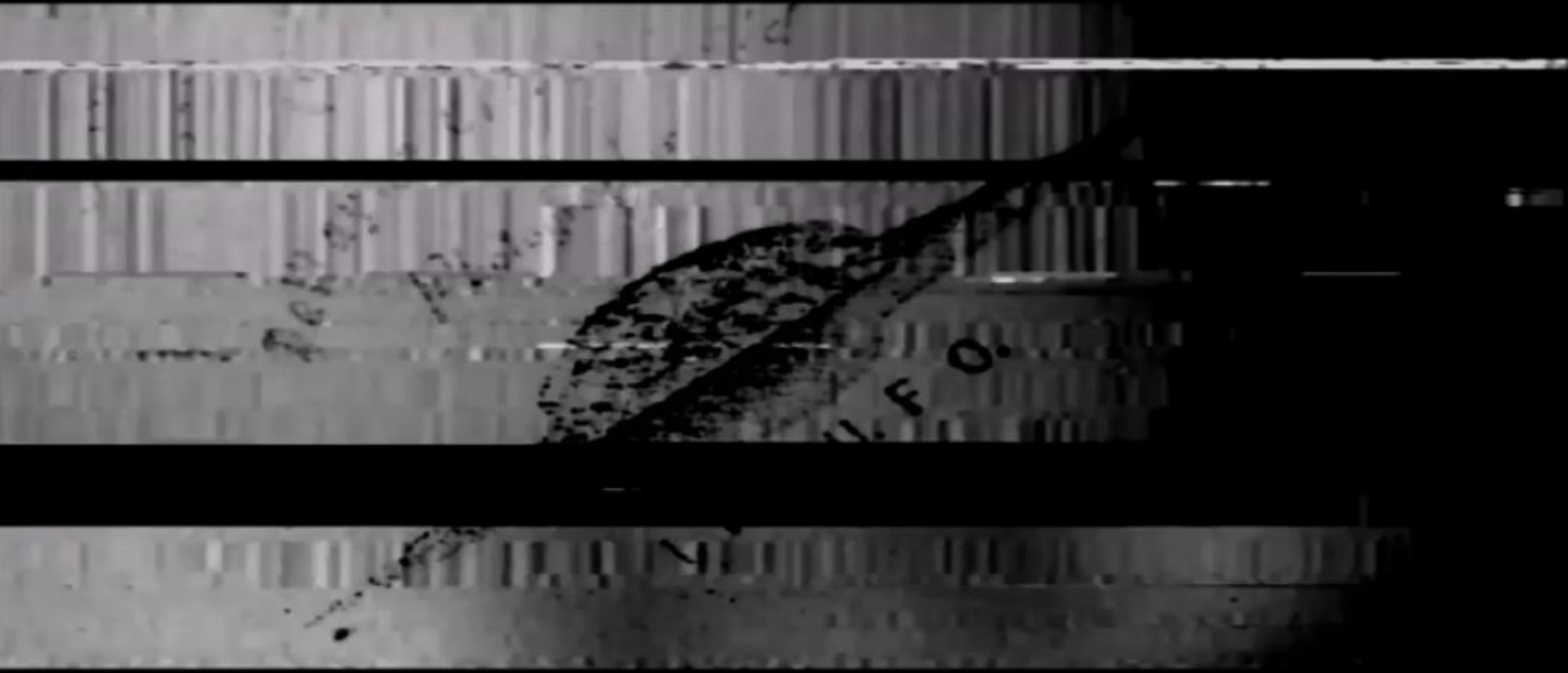
Faked: A Known CGI Example – Showing How Good CGI IS!

This is a CGI exercise in photorealism: i.e. the clip is 100% cg. The piece was done at Gnomon Studios with a crew of students from the Gnomon School of Visual Effects in association with MeniThings Productions.
Credits: Direction, Camera, Lighting ARISTOMENIS TSIRBAS
Produced by ALEX ALVAREZ, DARRIN KRUMWEIDE & ARISTOMENIS TSIRBAS
Visual Effects WAYNE HOLLINGSWORTH
Environment Model, Texture & Setup STEFANO FARCI & ADAD MORALES
Car & Mothership Models GUILHERME RAMBELLI UFO
Scout Model JOSE TRUJILLO
Hand Model, Texture & Rig TIFFANY YUNG
Smoke VFX, Flock model & Animation BERK HAKGUDER
Mothership Texture JEFF LUND & TIFFANY YUNG UFO
Scout Texture CHIKA SAITO
Additional Modeling BEN SARTORY
Mothership Concept LEONARDO KRAIDEN
UFO Scout Concept YURI BLAGOVESCHENSKIY
Created at GNOMON STUDIOS



Gnomon Studios© Used with permission.

Faked: A Known VFX Addition Example



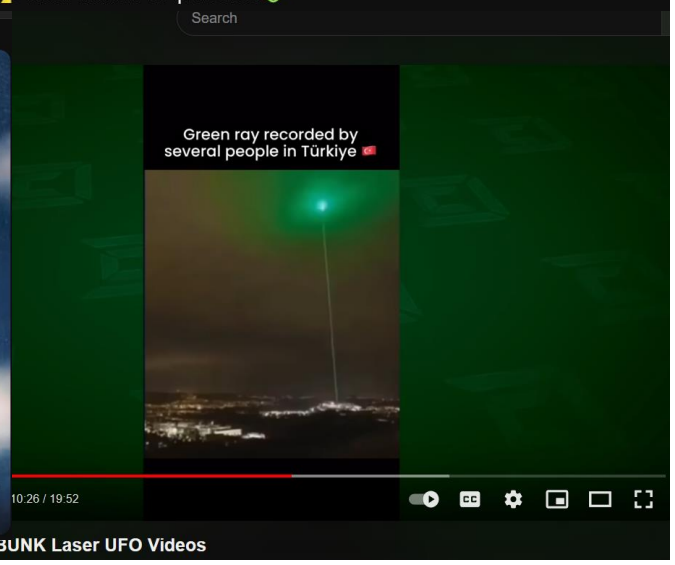
Corridor Digital© Used with permission.

A LOT OF FAKE UAPs/UFOs!

ael Shermer on How to Fake UFO Photographs

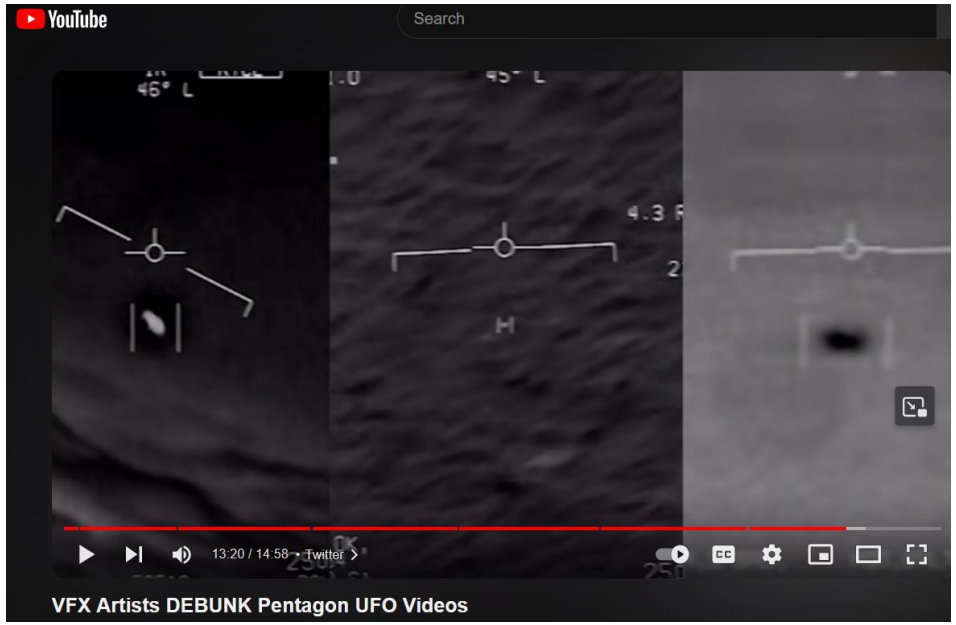
Faking a UFO Sighting 🛸 How hard is it? | Area 51 👽

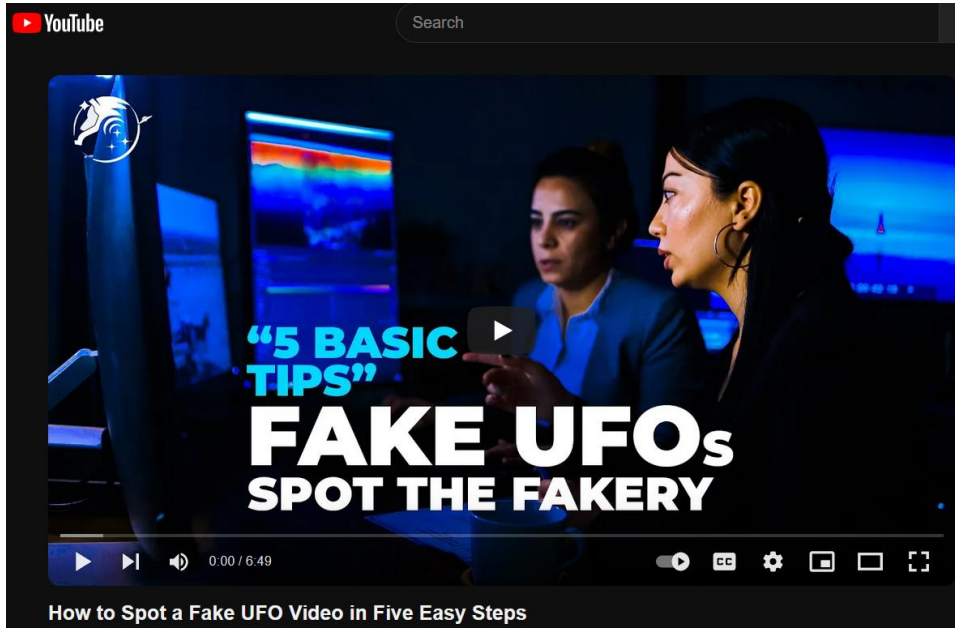
How easy it is to fake a UFO photo

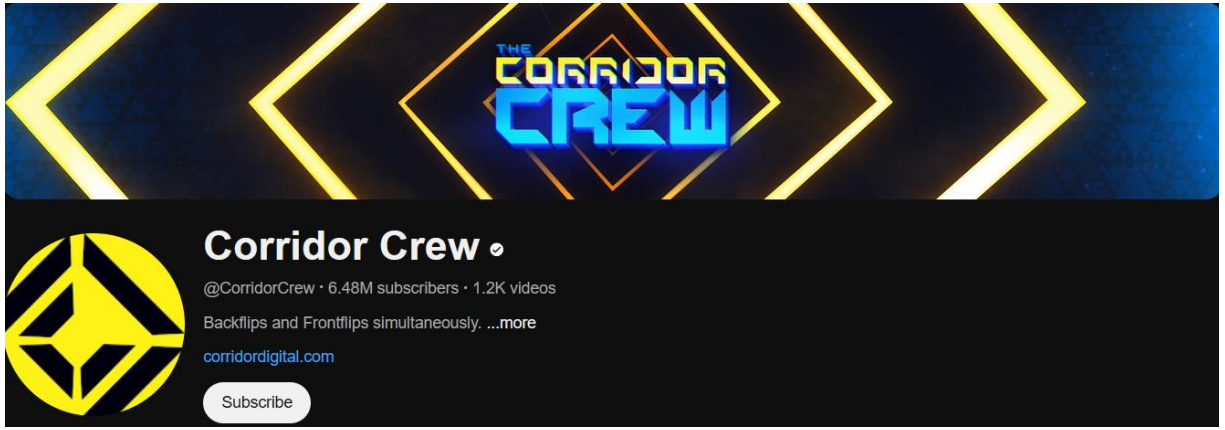


VFX Artists DEBUNK Laser UFO Videos

Some Good DEBUNKERs As Well







Faked: A Known UFO Hoax

Faked: A Known UFO Hoax

Breaking the Problem Down: Three Basic Options

1. **We can check for faked images/videos and hoaxes**

Breaking the Problem Down: Three Basic Options

1. **We can check for faked images/videos and hoaxes**

2. **Ensure authenticity of the images/videos**
   - *With an increase in hoax civil penalties*

Breaking the Problem Down: Three Basic Options

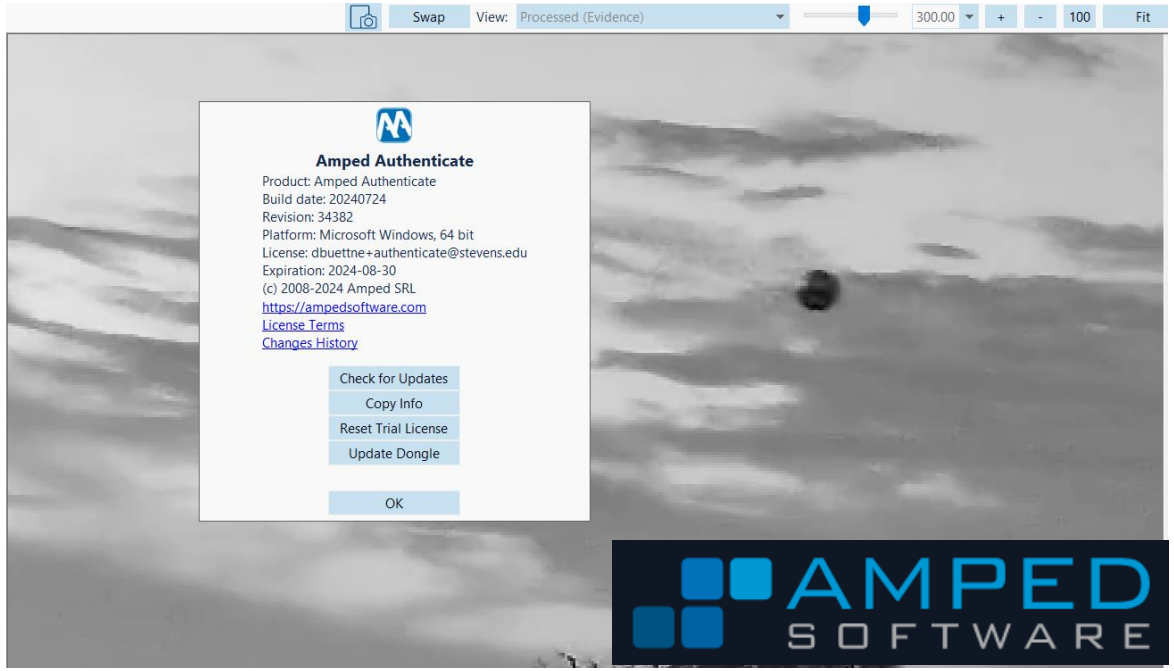
1. **We can check for faked images/videos and hoaxes**

2. **Ensure authenticity of the images/videos**
   - *With an increase in hoax civil penalties*

3. **… or we can do nothing**

Fake or Real? The UFO/UAP Hoax Arms Race…!

- ***Forensic Image/Video Analysis Tools Used by Law Enforcement Agencies***
  - Amped Authenticate
  - Cognitech TriSuite64
- ***Faked Data in Biomedical Papers***
  - Office of Research Integrity (ORI)
- ***Other Online Tools & Resources***
  - FotoForensics (analysis only)
  - Ghiro (open source licensed!)
  - Forensic Protection (open-source)
  - Forevid Video Analyzer



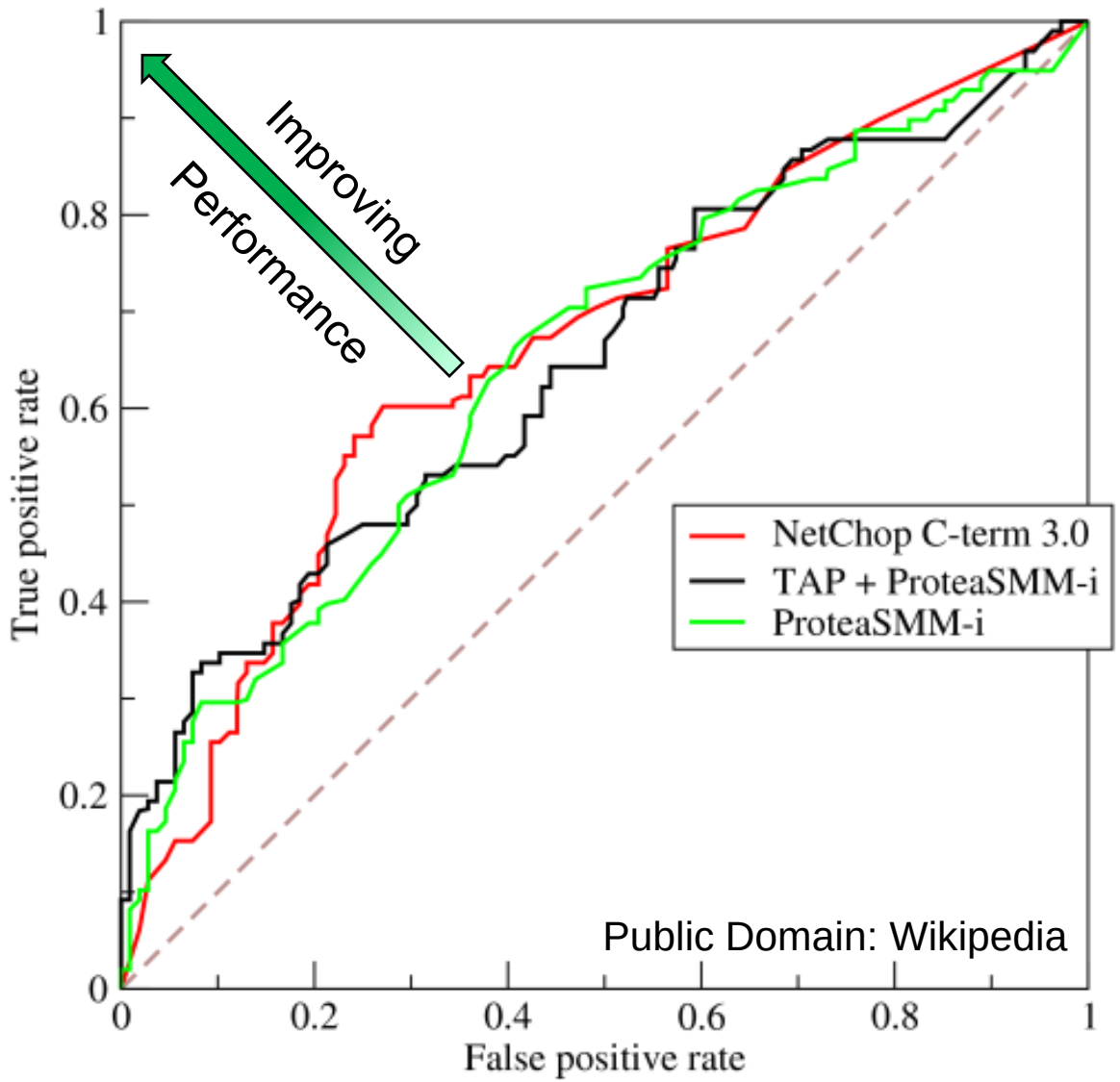
Fake or Real? The UFO/UAP Hoax Arms Race…!

- ***Fake Detection Algorithms Try to Detect Image Tampering***
  - Photoshop artifacts
  - Altered Iimage compression artifacts
  - CGI/VFX technology detection
  - Deep fake generation detection
- ***Semantic Forensics (SemaFor) DARPA Program***
  - Analytic catalog containing open-source resources developed under SemaFor for use by researchers and industry
  - *Provides a variety of detectors for images and videos (face-based and general), news, audio and text*
  - An open community research effort called AI Forensics Open Research Challenge Evaluation (AI FORCE), which aims to develop innovative and robust machine learning, or deep learning, models that can accurately detect synthetic AI-generated images
    - Detect DeepFakes: How to counteract misinformation created by AI (MIT Project)
- ***AI-generated or Real? Northwestern University tests how well you can detect faked images***

Fake or Real? Receiver Operating Characteristics (ROC)

ROC Curves

- Used to measure detection performance
- Plots **False positive rate** along the X-axis and the **True positive rate**
- **True positive rate** is the probability of a positive test result, conditioned on the individual (item) being truly being positive.
- **True negative rate** is the probability of a negative test result, conditioned on the individual truly being negative.
- **False positive rate** (or "false alarm rate") usually refers to the expectancy of the **false positive ratio.**
- **False positive ratio** (also known as **fall-out** or **false alarm ratio**) is the probability of falsely rejecting the null hypothesis for a particular test.



https://en.wikipedia.org/wiki/File:Roccurves.png

Fake or Real? Receiver Operating Characteristics (ROC)

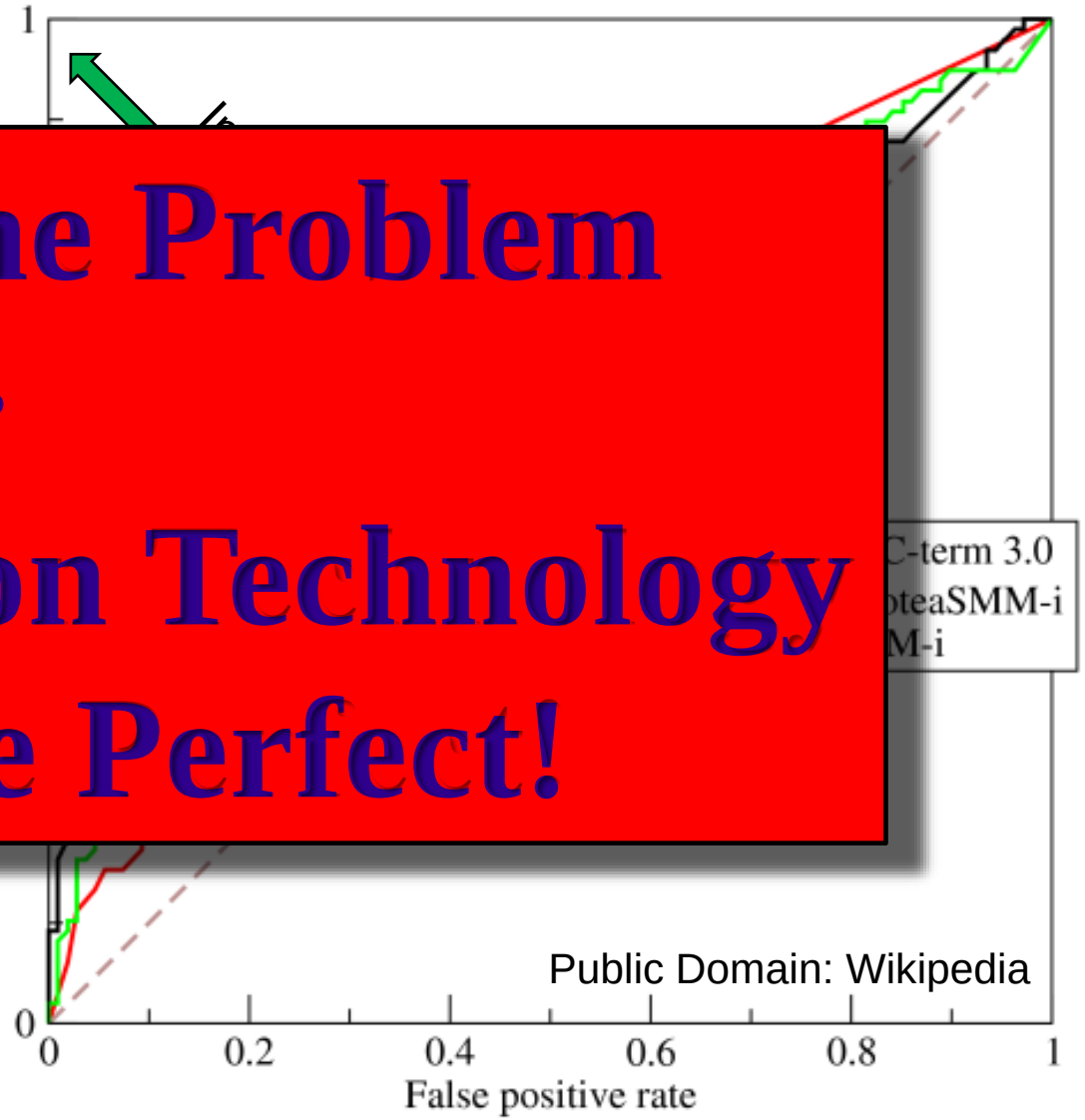
ROC Curves

- Used to measure detection performance
- Plots **R**
  the **Tru**
- **True p**
  positiv
  individ
- **True n**
  negati
  individ
- **False**
  usually
  positive ratio.
- **False positive ratio** (also known as **fall-out** or **false alarm ratio**) is the probability of falsely rejecting the null hypothesis for a particular test.

Illustrating The Problem

. . .

No Fake Detection Technology Will Ever Be Perfect!



Public Domain: Wikipedia

https://en.wikipedia.org/wiki/File:Roccurves.png

AND BEFORE YOU CAN ACCUSE ME
… OF BEING BIASED …
AND THAT I BELIEVE ALL UAP REPORTS ARE FAKED

Not Everything Can Be Easily Explained

SCORE [77]

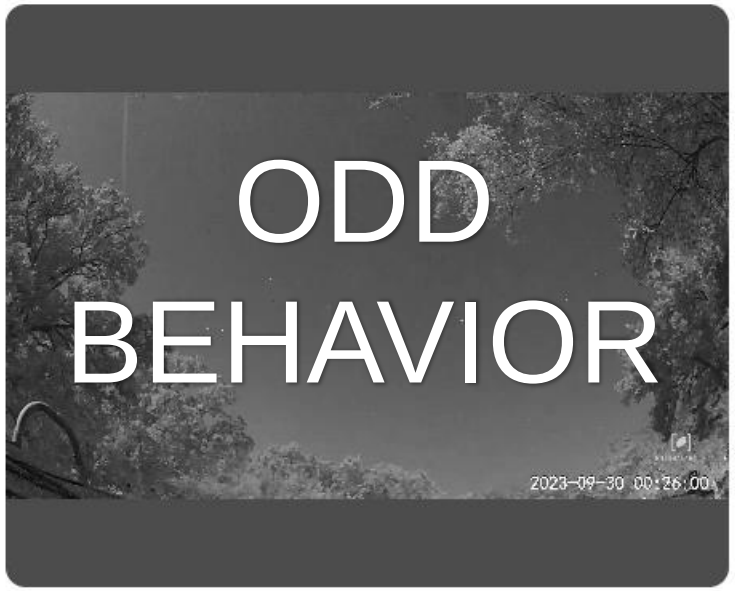
Enigma UFO Sighting #282353

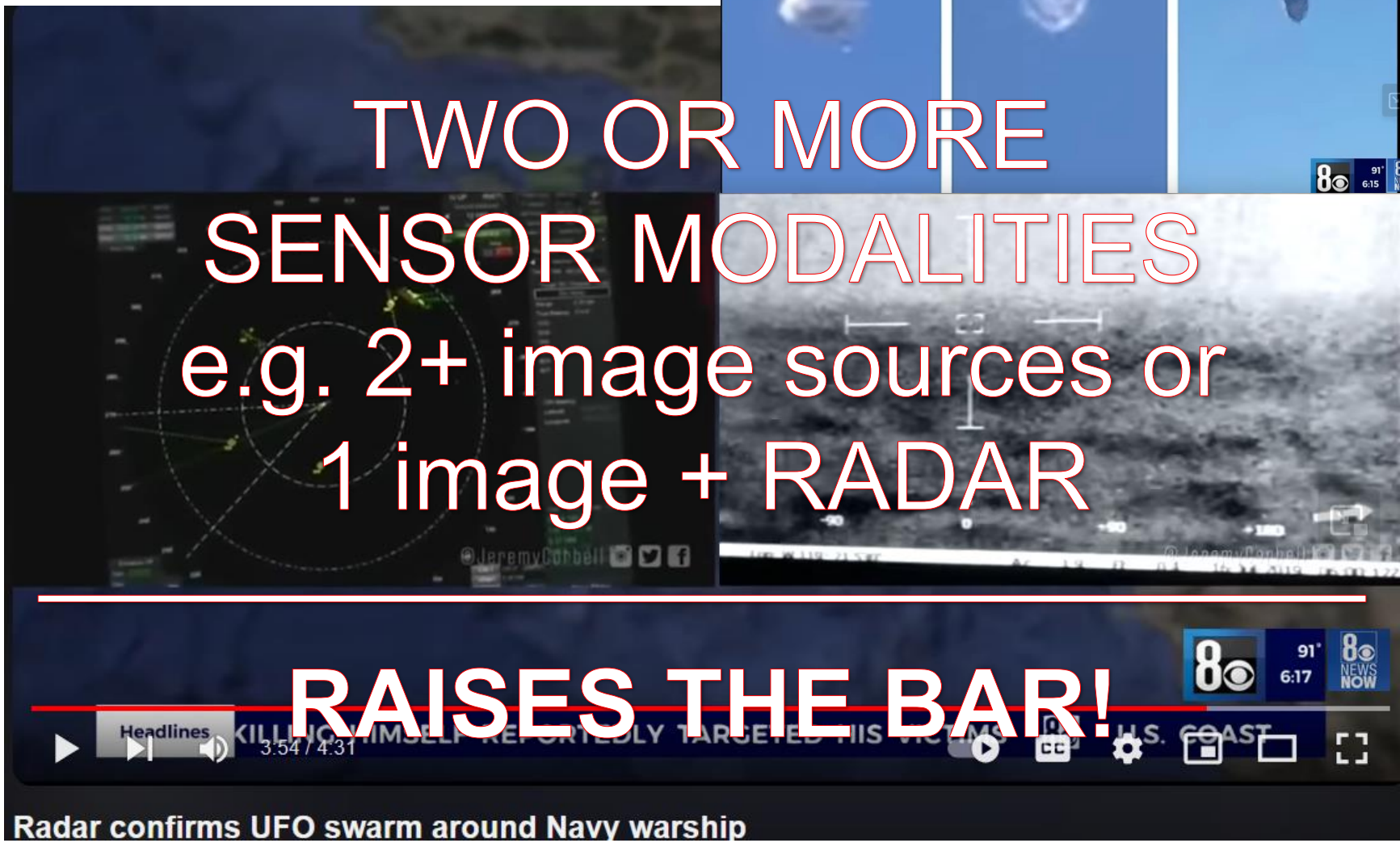
2023 SEP 30 • 12:26:00 AM CDT

SIGNAL MOUNTAIN, TENNESSEE, UNITED STATES

WITNESSES 2

10 Minute video, Comes in bottom right (Northwest), at around 5:30/36 something appears to either shoot away from the UAP or is something entering/leaving the atmosphere based on speed and light. Camera is shit quality WYZE pushed to a bridge that re... Read more ↓





Not Everything Can Be Easily Explained

SCORE [77]

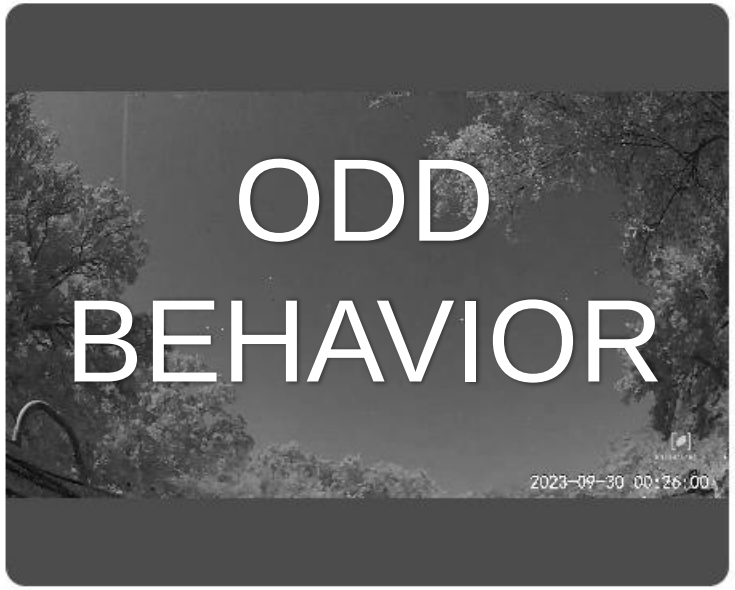
Enigma UFO Sighting #282353

2023 SEP 30 • 12:26:00 AM CDT

SIGNAL MOUNTAIN, TENNESSEE, UNITED STATES

WITNESSES 2

10 Minute video, Comes in bottom right (Northwest), at around 5:30/36 something appears to either shoot away from the UAP or is something entering/leaving the atmosphere based on speed and light. Camera is shit quality WYZE pushed to a bridge that re... Read more ↓

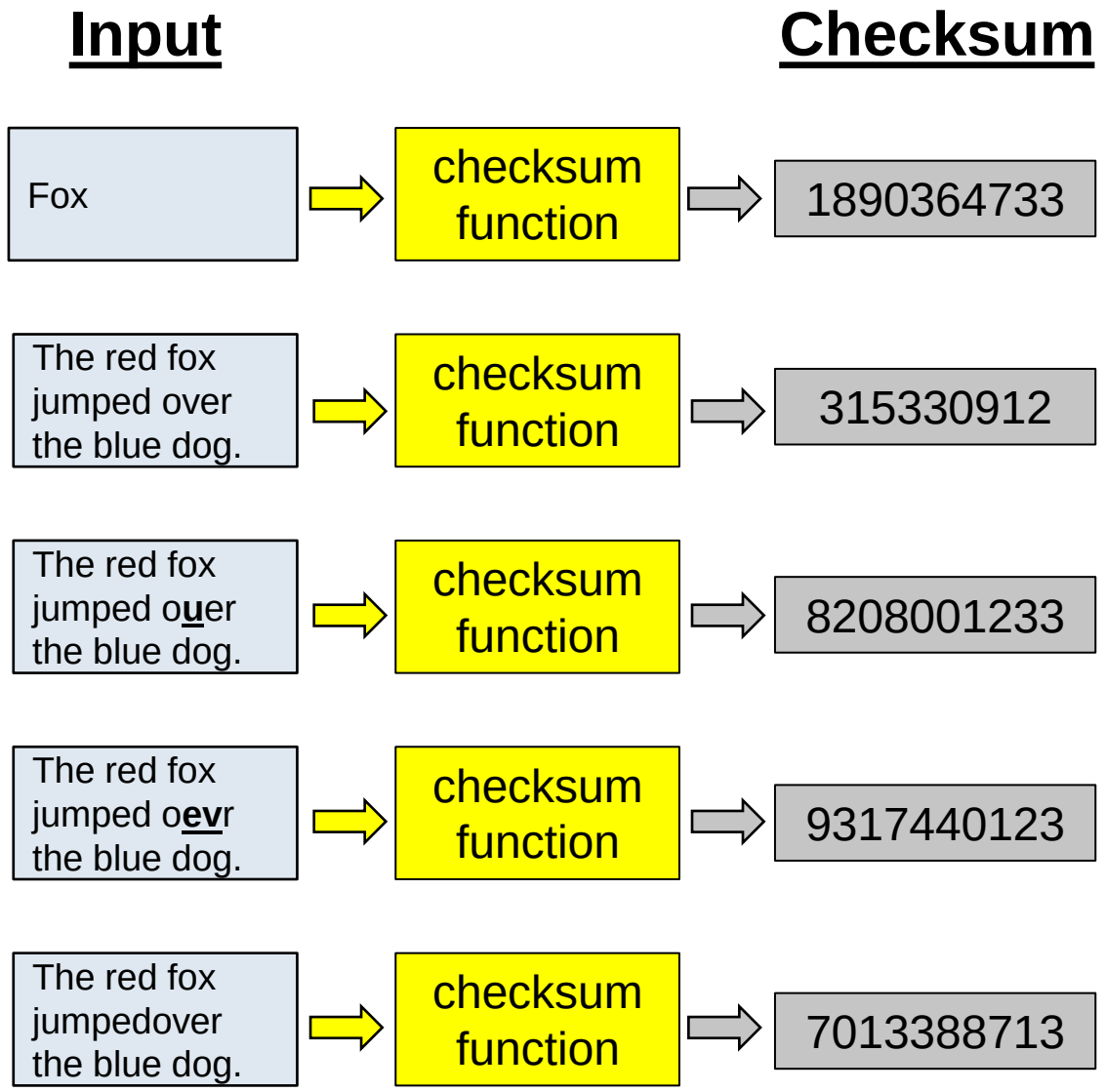


Data Integrity Technologies

- *Cryptographic Image/Video & Metadata Content Checksums*
- *Digitally Signing Images/Videos*
  - Concerns with misuse by a repressive government
    - *Could be provided as a device specific signing approach, i.e. not related to a specific owner but to the creation of the data on the device itself.*
- *Source Based Image/Video Encryption*
- *Blockchain*
- *Certification/Authentication Infrastructure (an example)*
- *Zero Trust Security Architecture (NIST Explanatory Document)*
  - IEEE Paper Zero Trust Architecture (ZTA): A Comprehensive Survey

What is a Checksum?

A **checksum** is a small-sized block of data derived from another block of digital data for the purpose of detecting errors that may have been introduced during its transmission or storage. By themselves, **checksums are often used to verify data integrity but are not relied upon to verify data authenticity**.



https://en.wikipedia.org/wiki/Checksum

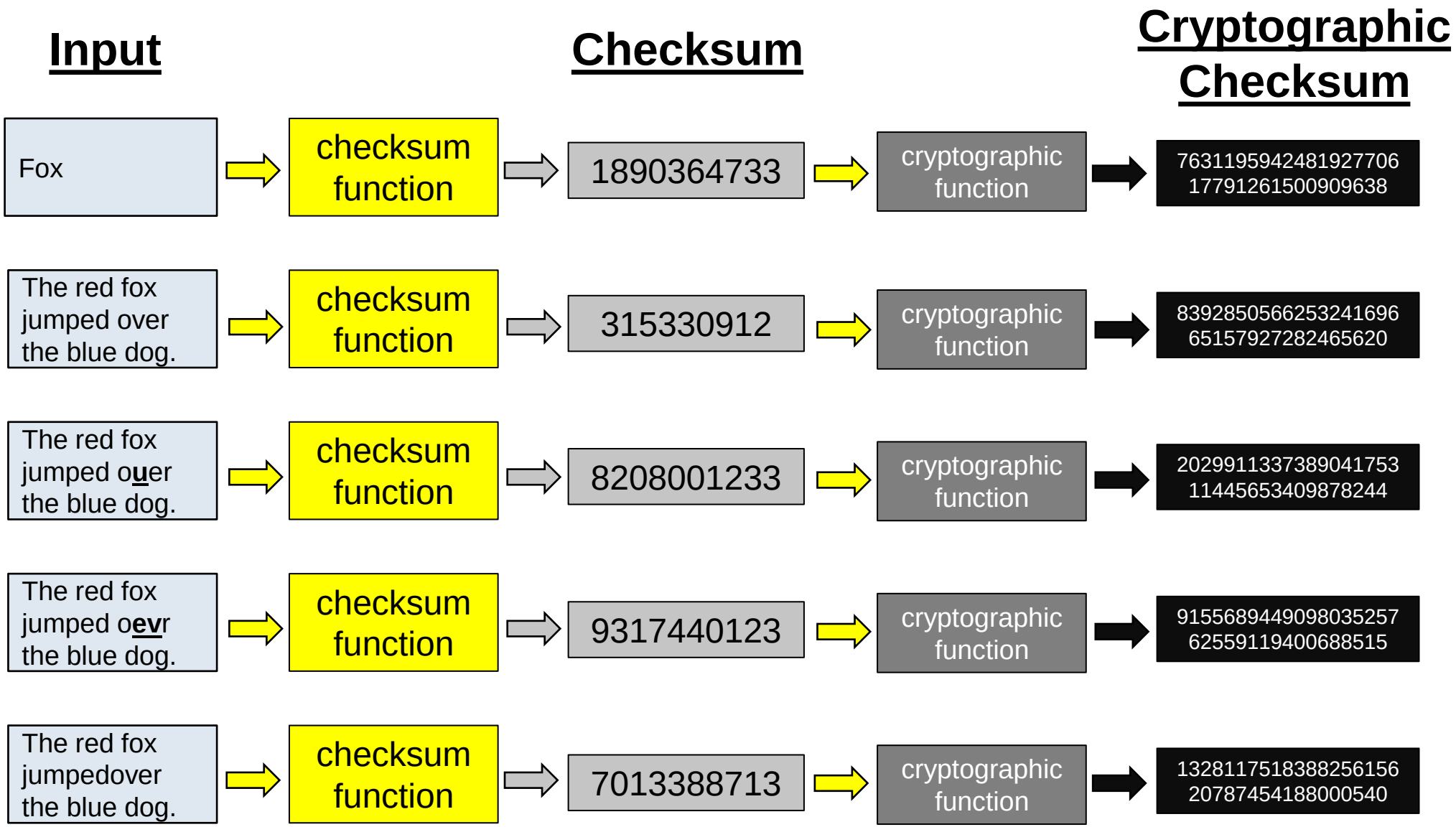
What is a Cryptographic Checksum?

A **checksum** is a mathematical value (called a checksum) that is derived from a file, an image or other data and is used to "test" at a later date to verify that the data has not been maliciously changed.

A **cryptographic checksum** is created by performing a complicated series of mathematical operations (known as a cryptographic algorithm) that translates the checksum into a fixed string of digits called a hash value, which is then used as a checksum. Without knowing which cryptographic algorithm and specific assigned private key was used to create the hash value, it is highly unlikely that an unauthorized person would be able to change data without inadvertently changing the corresponding checksum.

https://www.cybertraining365.com/cybertraining/Topics/Cryptographic_checksum

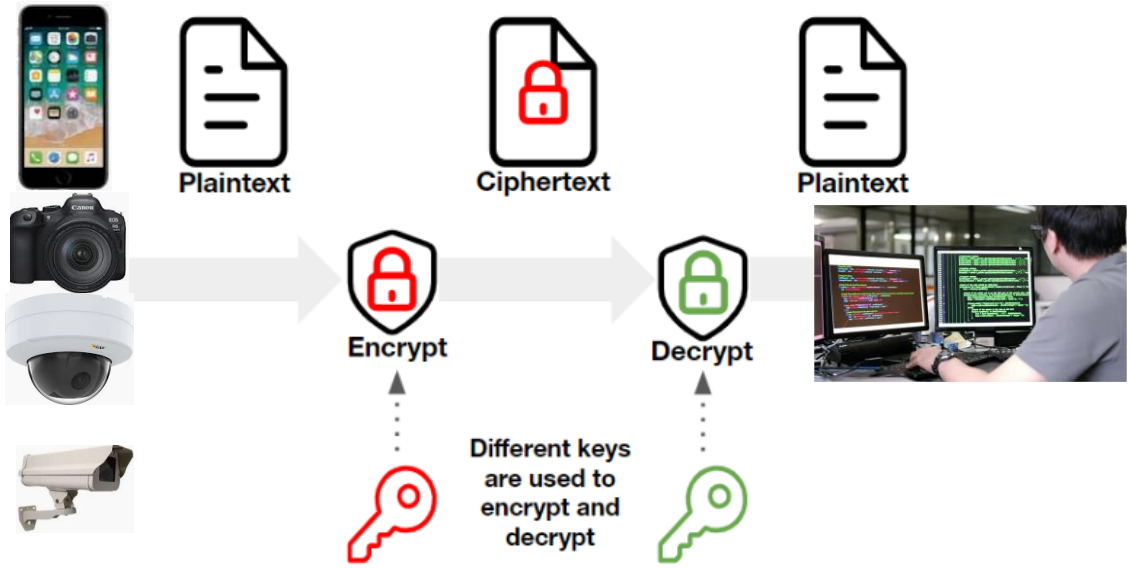
What is a Cryptographic Checksum?



Where is a Cryptographic Checksum Calculated?

- **On the device generating the data!**
  - *Devices need an embedded cryptographic chip*
    - E.g. Public Key Infrastructure (PKI) based encryption can be provided on new devices and then pass new digital certificates to devices that people want to use for UAP and paranormal studies.
    - The certificates are used to encrypt the checksum of the data and updated according to a risk profile.
    - Verification that the cryptographic checksum is correct can be used by UAP researchers to validate that the data has ***likely*** not been tampered with.
    - ***likely***, is used here to indicate that nothing is 100% rock solid foolproof, but it does significantly raise the bar, protecting us scientists from fraud.

- **Just encrypting a checksum of the image, meta- and scientific-data is likely sufficient!**
  - *Saves on the amount of data being encrypted and thus likely low overhead!*



https://www.hashicorp.com/blog/what-is-public-key-infrastructure-pki

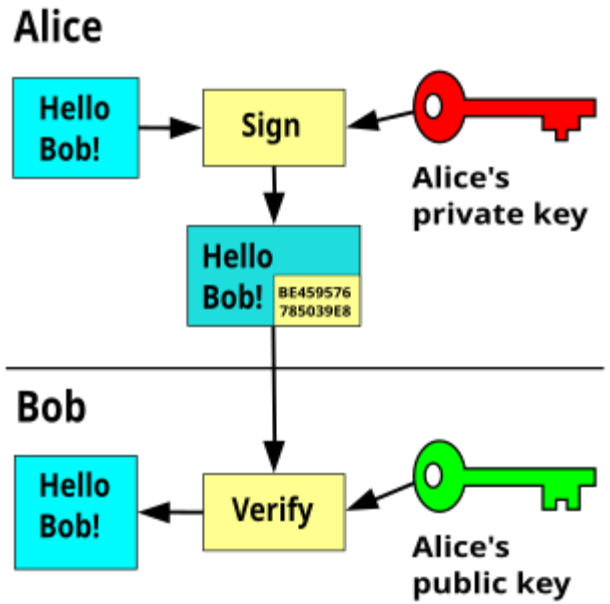
Digitally Signing Images/Videos

A **digital signature** is a mathematical scheme for verifying the authenticity of digital messages or documents. A valid digital signature on a message gives a recipient confidence that the message came from a sender known to the recipient.

Digital signatures are a standard element of most cryptographic protocol suites, and are commonly used for software distribution, financial transactions, contract management software, and in other cases where it is important to detect forgery or tampering.

Significant privacy concerns from repressive governments requiring personally digitally signing images and videos.

However, device digital signatures not traceable to the actual owner is a possibility!

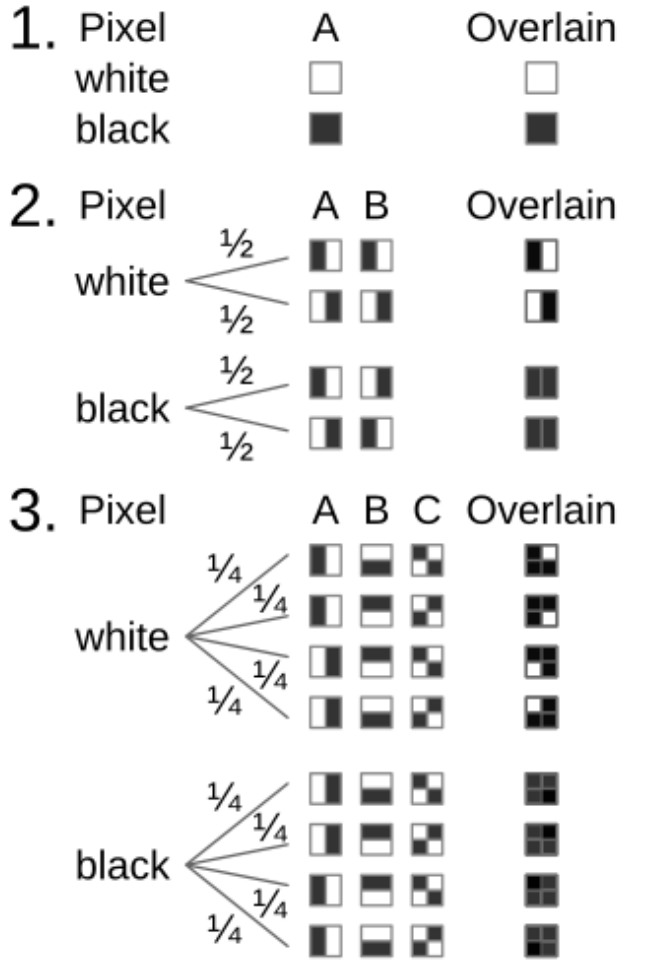


https://en.wikipedia.org/wiki/Digital_signature

Source-based Video Encryption

Source-based Image/Video Encryption (Visual cryptography) is a cryptographic technique which allows visual information (pictures, text, etc.) to be encrypted in such a way that the decrypted information appears as a visual image.
[…] a visual secret sharing scheme, where a binary image was broken up into n shares so that only someone with all n shares could decrypt the image, while any n − 1 shares revealed no information about the original image.

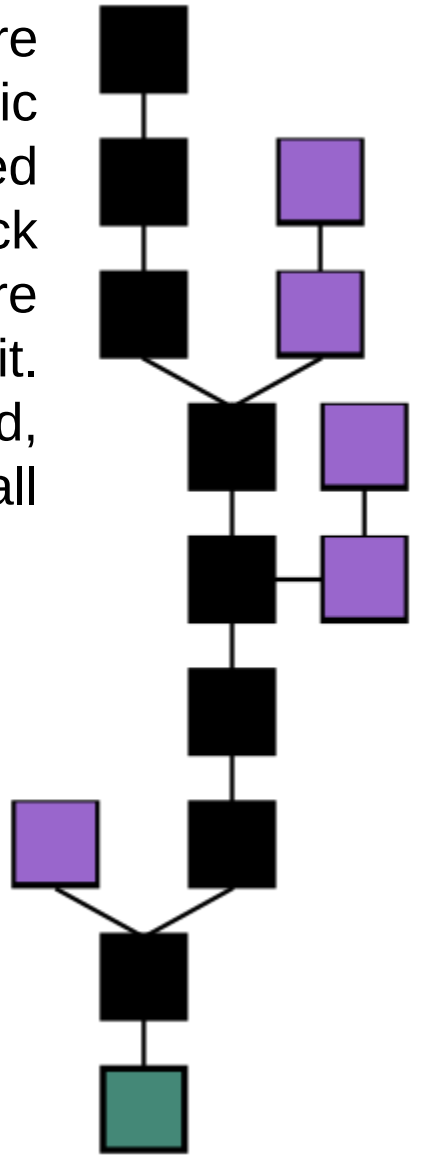
Development of masks to let overlaying n transparencies A, B,... printed with black rectangles reveal a secret image — n = 4 requires 16 (24) sets of codes each with 8 (24-1) subpixels, which can be laid out as 3×3 with the extra bit always black



https://en.wikipedia.org/wiki/Visual_cryptography

Blockchain

A **blockchain** is a distributed ledger with growing lists of records (blocks) that are securely linked together via cryptographic hashes. Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data (generally represented as a Merkle tree, where data nodes are represented by leaves). Since each block contains information about the previous block, they effectively form a chain (compare linked list data structure), with each additional block linking to the ones before it. Consequently, blockchain transactions are irreversible in that, once they are recorded, the data in any given block cannot be altered retroactively without altering all subsequent blocks. […]

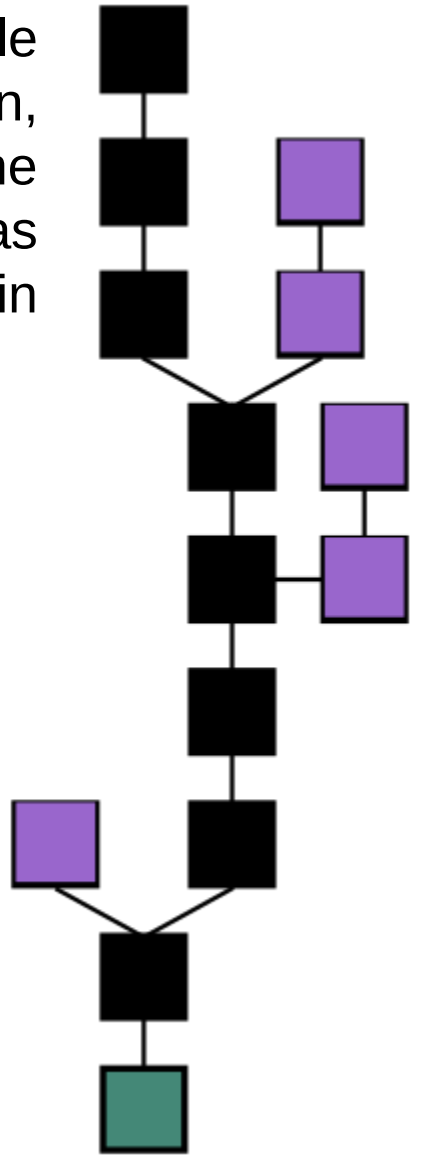


Blockchain formation: The main chain (black) consists of the longest series of blocks from the genesis block (green) to the current block. Orphan blocks (purple) exist outside of the main chain.

https://en.wikipedia.org/wiki/Blockchain

Blockchain: Blocks

Blocks hold batches of valid transactions that are hashed and encoded into a Merkle tree. Each block includes the cryptographic hash of the prior block in the blockchain, linking the two. The linked blocks form a chain. This iterative process confirms the integrity of the previous block, all the way back to the initial block, which is known as the genesis block (Block 0). To assure the integrity of a block and the data contained in it, the block is usually digitally signed. […]



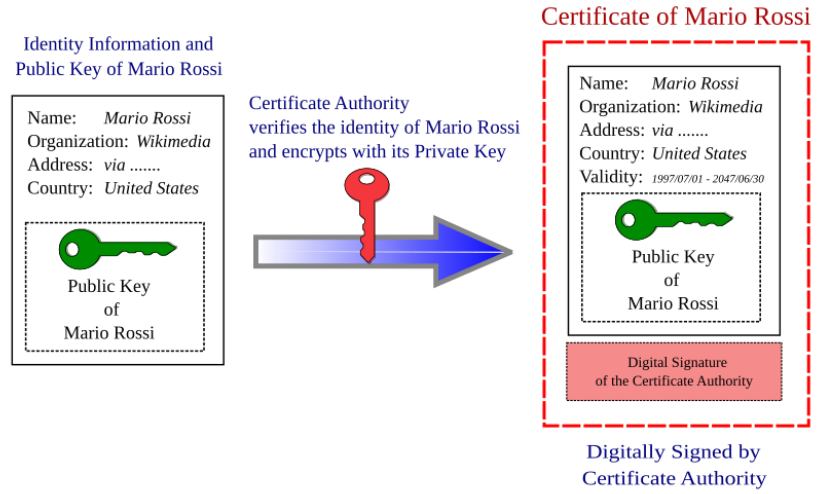
Blockchain formation: The main chain (black) consists of the longest series of blocks from the genesis block (green) to the current block. Orphan blocks (purple) exist outside of the main chain.

https://en.wikipedia.org/wiki/Blockchain

Certificate Authority

In cryptography, a **certificate authority** or **certification authority (CA)** is an entity that stores, signs, and issues digital certificates. A digital certificate certifies the ownership of a public key by the named subject of the certificate. This allows others (relying parties) to rely upon signatures or on assertions made about the private key that corresponds to the certified public key. A CA acts as a trusted third party—trusted both by the subject (owner) of the certificate and by the party relying upon the certificate. The format of these certificates is specified by the X.509 or EMV standard.

One particularly common use for certificate authorities is to sign certificates used in HTTPS, the secure browsing protocol for the World Wide Web. Another common use is in issuing identity cards by national governments for use in electronically signing documents.



https://en.wikipedia.org/wiki/Certificate_authority

Authentication and Authorization Infrastructure

Authentication and authorization infrastructure (AAI) refers to a service and a procedure that enables members of different institutions to access protected information that is distributed on different web servers.

Traditional approaches to authorization and access control in computer systems are not sufficient to address the requirements of federated and distributed systems, where infrastructural support may be required. Authentication and authorization infrastructure solutions address such limitations. With an AAI, access control is not managed by a central register, but by the respective organization of the user who wishes to access a specific resource.

https://en.wikipedia.org/wiki/Authentication_and_authorization_infrastructure

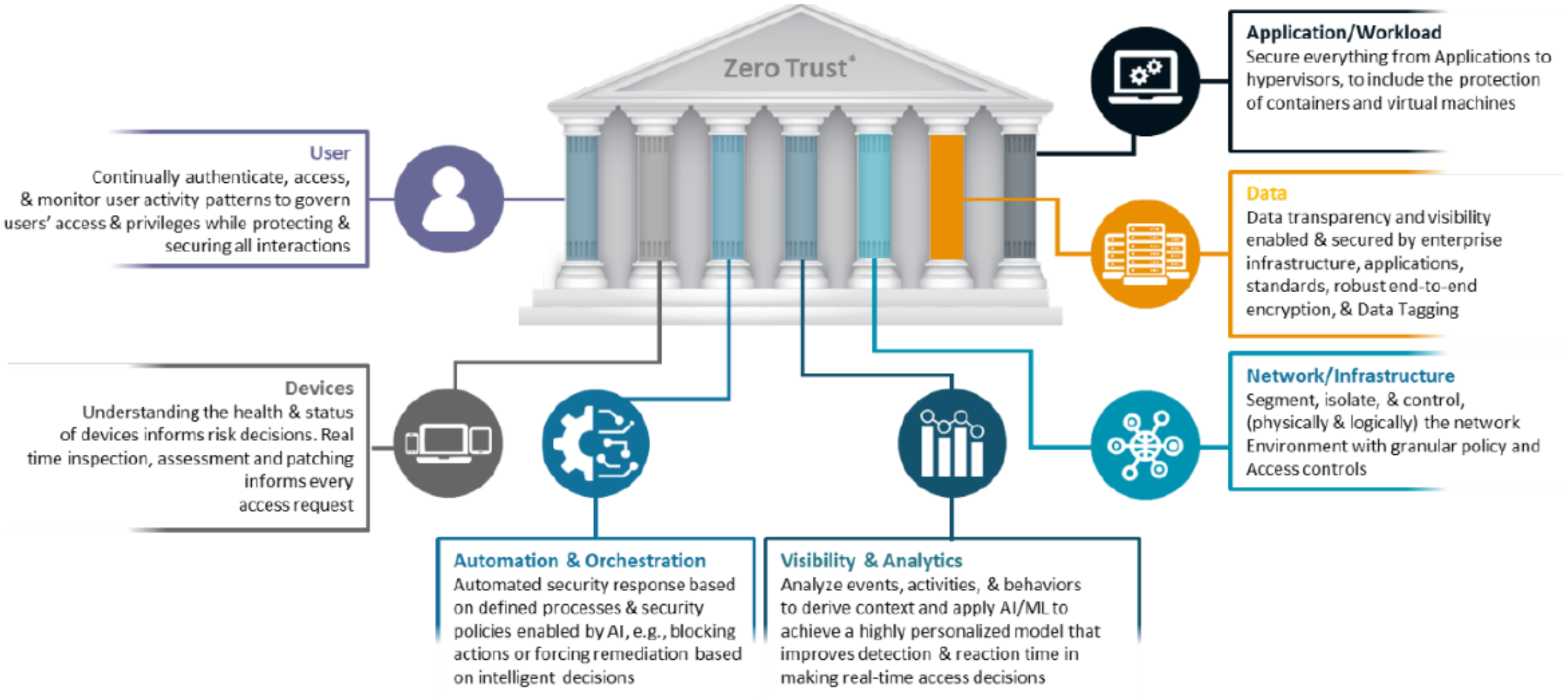
Zero Trust Security Architecture

The **zero trust security model**, also known as **zero trust architecture (ZTA)**, and **perimeterless security** describes an approach to the strategy, design and implementation of IT systems. The main concept behind the zero-trust security model is "**never trust, always verify**", which means that users and devices should not be trusted by default, even if they are connected to a permissioned network such as a corporate LAN and even if they were previously verified.

ZTA is implemented by establishing strong identity verification, validating device compliance prior to granting access, and ensuring least privilege access to only explicitly-authorized resources. Most modern corporate networks consist of many interconnected zones, cloud services and infrastructure, connections to remote and mobile environments, and connections to non-conventional IT, such as IoT devices.

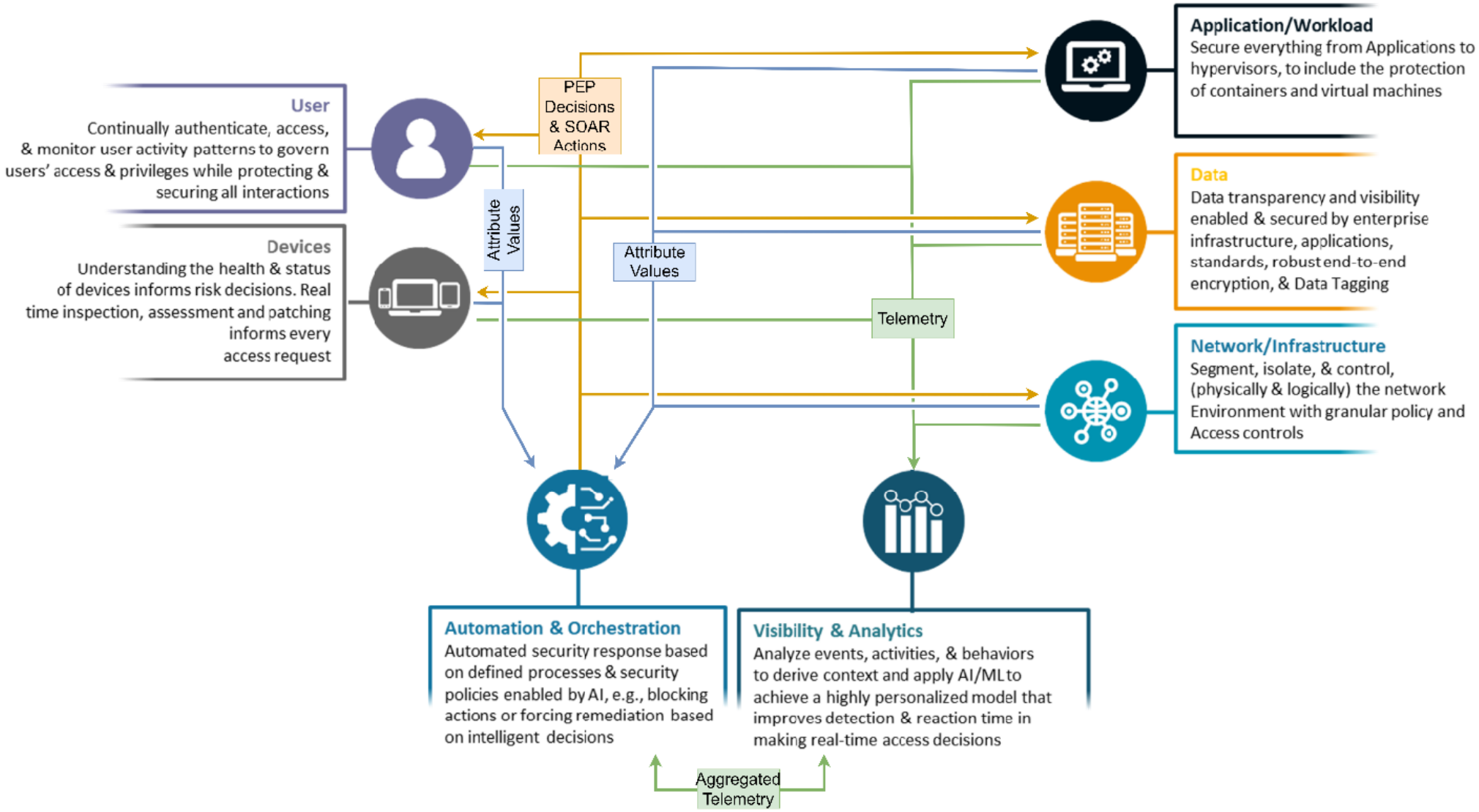
https://en.wikipedia.org/wiki/Zero_trust_security_model

ZTA: Trust Pillars



https://dodcio.defense.gov/Portals/0/Documents/Library/ZeroTrustOverlays.pdf

ZTA: Integration and Interoperability



https://dodcio.defense.gov/Portals/0/Documents/Library/ZeroTrustOverlays.pdf

What is a Data Architecture? According to IBM

A ***data architecture*** describes how data is managed–from collection through to transformation, distribution, and consumption. It sets the blueprint for data and the way it flows through data storage systems. It is foundational to data processing operations and artificial intelligence (AI) applications.

[…]

https://www.ibm.com/topics/data-architecture

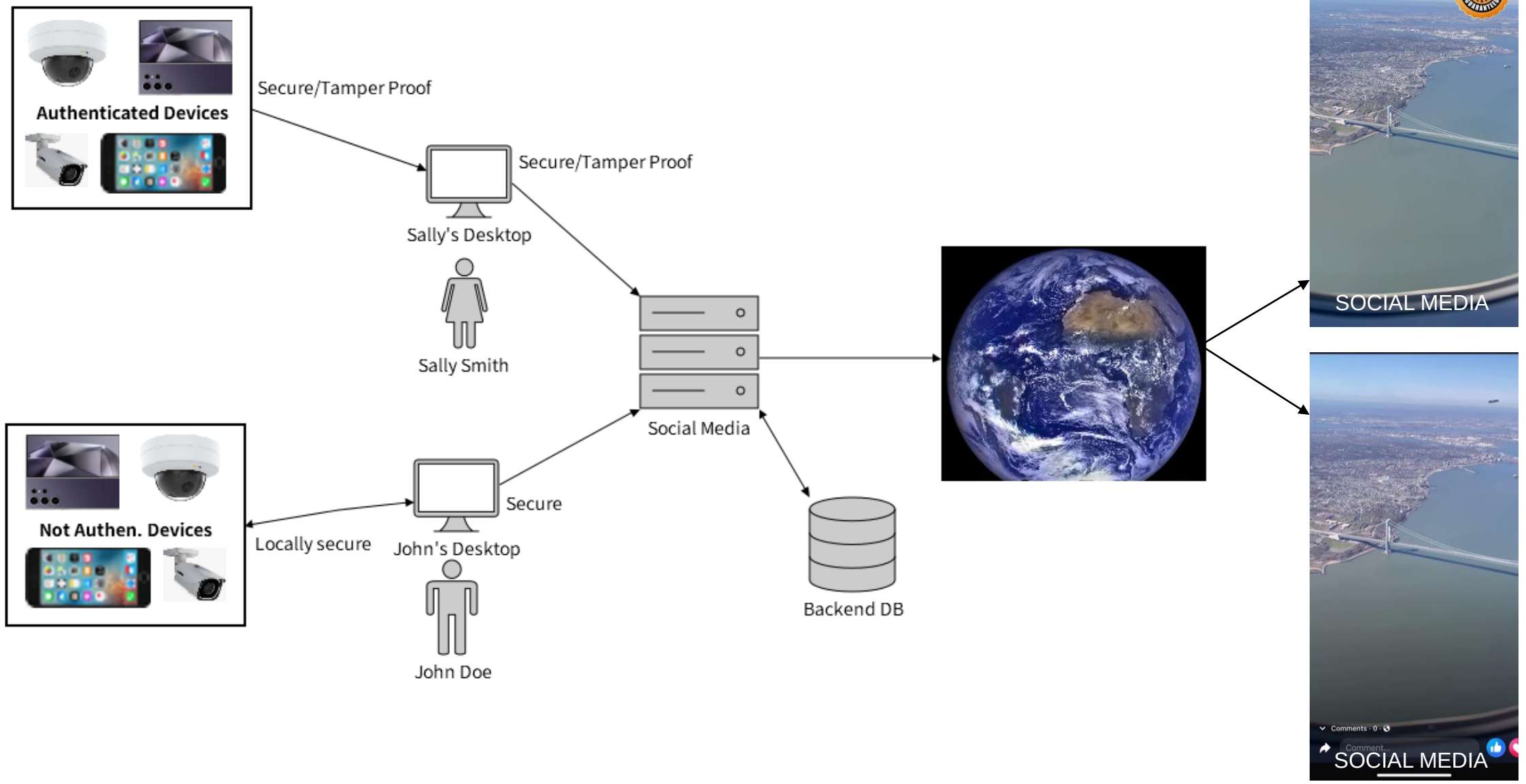
What is a Data Architecture? According to IBM

A data architecture's documentation includes:

Conceptual data models: They are also referred to as domain models and offer a big-picture view of what the system will contain, how it will be organized, and which business rules are involved. Conceptual models are usually created as part of the process of gathering initial project requirements. Typically, they include entity classes (defining the types of things that are important for the business to represent in the data model), their characteristics and constraints, the relationships between them and **relevant security and data integrity requirements**.

https://www.ibm.com/topics/data-architecture

Conceptual Data Model: For UAP Studies



As a Scientist,
based on what I just showed you…

Which data do you think I want to analyze?

Recommendations: How Can Citizen Scientists Help?

- **Please, don't start by posting your UAP videos on social media**
  - Instead … first do the following …
- **Report the sighting to a reputable scientific organization**
  - **A short list of examples include:**
    - *Scientific Coalition for UAP Studies (SCU)* → *https://www.explorescu.org/report-a-uap*
    - *Enigma Labs* → *https://enigmalabs.io/*
    - *Mutual UFO Network (MUFON)* → *https://mufon.com/cms-ifo-info/*
    - Some TV Hosts also do a good job of properly investigation reports → https://benhansen.com/
  - **Reporting:**
    - *Provide as much detail as possible in the sighting report and all original data*
    - *Many of these organizations will keep who reported the sighting confidential if you wish to remain anonymous, but they may need to have investigators contact you for an interview to get additional details*

Recommendations: How Can Citizen Scientists Help?

- **… after reporting:**
  - Feel free to post on social media if you wish, however, please identify the organization that took your report and your report identification # to allow us to follow-up with you if the organization does not have their own supporting researchers!

- **Get Involved!**
  - Join the Society for UAP Studies or SCU and work with a rapidly growing community of volunteer scientists and other professionals to help weed out the mundane phenomena from the truly anomalous to help us get to the bottom of UAPs!
  - For example, the SCU has numerous citizen scientists' projects they are working on now!

Legal Disclaimer

All trademarks, service marks, and trade names are the property of their respective owners.

All videos, images and diagrams are the property of their respective owners. Videos were only included after receiving explicit permission from their ownsers to use in this presentation.

Inclusion in this presentation of any company's technology is not an endorsement by Stevens Institute or the University of Utah.

All opinions are mine, and mine alone.

BACK UP

What is a Data Architecture? According to IBM

A data architecture's documentation includes:

Logical data models: They are less abstract and provide greater detail about the concepts and relationships in the domain under consideration. One of several formal data modeling notation systems is followed. These indicate data attributes, such as data types and their corresponding lengths, and show the relationships among entities. Logical data models don't specify any technical system requirements.

https://www.ibm.com/topics/data-architecture

What is a Data Architecture? According to IBM

A data architecture's documentation includes:

Physical data models: The physical data model is the most detailed and specific of the three. It defines the actual implementation of the database, including table structures, indexes, storage and performance considerations. It focuses on the technical aspects of how the data will be stored and accessed and is used for database schema creation and optimization.

https://www.ibm.com/topics/data-architecture

Fake?: What is CGI/VFX?

Computer-generated imagery (**CGI**) is a specific-technology or application of computer graphics for creating or improving images in art, printed media, simulators, videos and video games. These images are either static (i.e. still images) or dynamic (i.e. moving images). CGI both refers to 2D computer graphics and (more frequently) 3D computer graphics with the purpose of designing characters, virtual worlds, or scenes and special effects (in films, television programs, commercials, etc.). The application of CGI for creating/improving animations is called *computer animation*, or *CGI animation*.

https://en.wikipedia.org/wiki/Computer-generated_imagery

Fake?: What is CGI/VFX?

Visual effects (sometimes abbreviated **VFX**) is the process by which imagery is created or manipulated outside the context of a live-action shot in filmmaking and video production. The integration of live-action footage and other live-action footage or CGI elements to create realistic imagery is called VFX.

VFX involves the integration of live-action footage (which may include in-camera special effects) and generated-imagery (digital or optics, animals or creatures) which look realistic, but would be dangerous, expensive, impractical, time-consuming or impossible to capture on film. Visual effects using computer-generated imagery (CGI) have more recently become accessible to the independent filmmaker with the introduction of affordable and relatively easy-to-use animation and compositing software.

https://en.wikipedia.org/wiki/Visual_effects