

# ISC

(BTech III Jan-April 2024)

Week#10 (9) – March 8, 2024

Dhiren Patel

# Quick discussion on Mid-sem Q3

| I/J | N | D | A | S |
|-----|---|---|---|---|
| E   | T | H | L | F |
| O   | G | M | P | V |
| B   | C | K | R | U |
| Q   | W | X | Y | Z |

INDIA SENT HALF OF THE ENGLISH TEAM TOP FIVE  
BACK BEFORE LUNCH

AAKAASH DEEP  
AA KA AS HD EE P  
AX AK AA SH DE EP  
AX AK AX AS HD EE P  
AX AK AX AS HD EX EP

Answer: YD RD YD SI DX QH OL

INDIA with Caesar cipher – key 3

INDIA with Vigenere cipher – key ROOT

- Answer – LQGLD
- Answer –
- INDIA
- ROOTR
- (?)

# Q4

1. One Time Pad – most secure?

- One input (key) out of two is completely Random
- Key is destroyed once used - not used again

2. Hill cipher (2-gram), Playfair cipher (by design it's 2-gram cipher)

Security – frequency analysis attacks, 2x2 matrix (weaknesses) v/s 5x5 key matrix

3. As per ENISA report - Two most prominent threats are: Ransomware and DDoS (Distributed Denial of Service)

Two prominent target sectors are: Public administration and Health care / targeted Individuals

# Q4

## 4. Polygram v/s Polyalphabetic

- Polygram – Hill cipher (n-gram), Playfair cipher (2-gram)
- Polyalphabetic – Vigenere cipher, One Time Pad

## 5. AI driven misinformation examples

Misleading “deepfakes” (SRT), Chatbots impersonating candidates (Election)

Mitigation strategies – E.g. DALL·E has guardrails to decline requests that ask for image generation of real people, including candidates

Original



New



# New Monalisa contains the msg

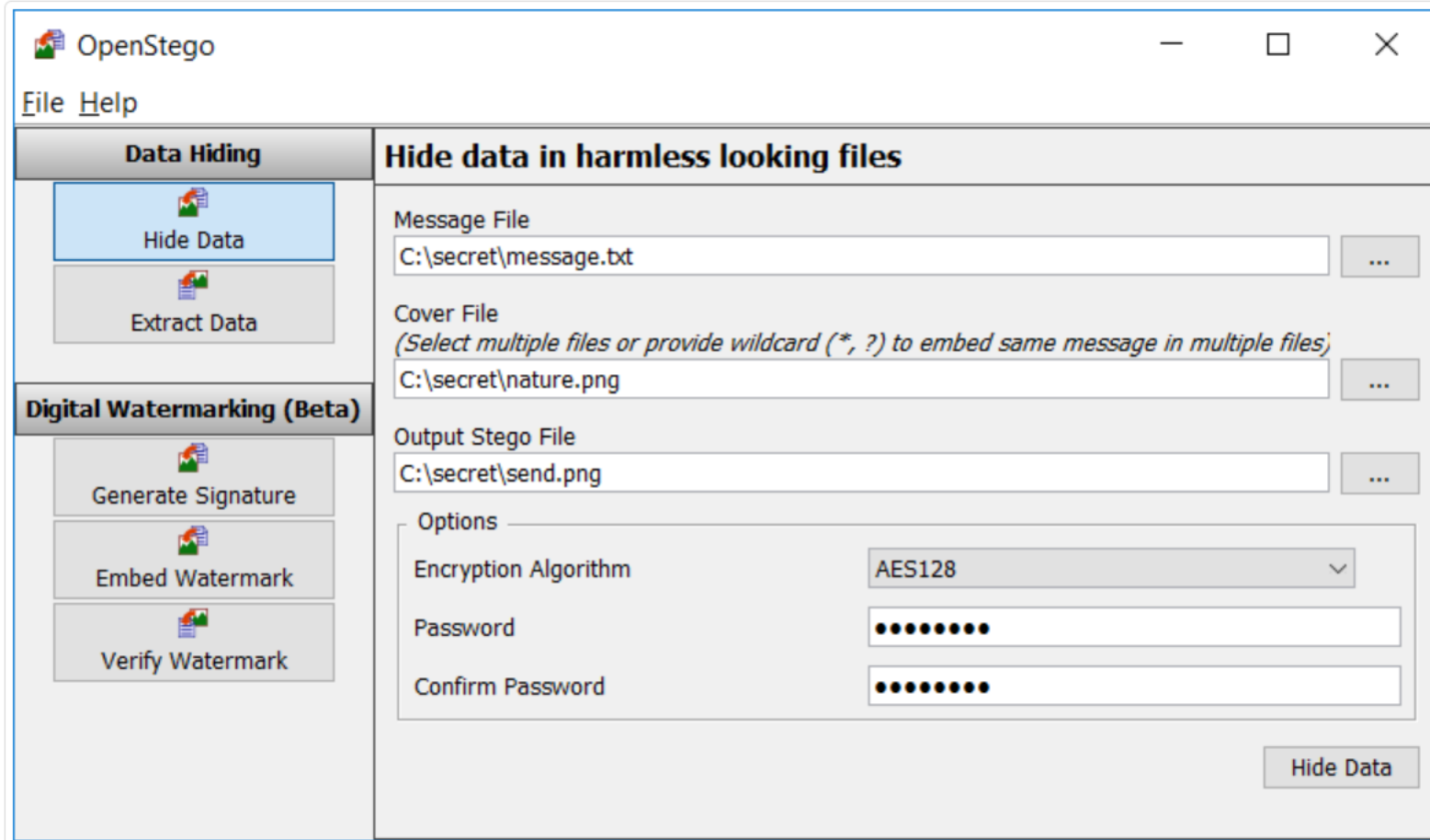
- “Hello Zelenski, Take delivery of 100 Cruise Missiles at Gdańsk port Poland on 26 Jan 2024 0200. –Biden”
- Image is a good candidate as container where small msg can fit in using least affecting bit of the pixel!
- Welcome to the world of Steganography !!
- Steganography techniques involve hiding sensitive information within an ordinary, non-secret file or message, so that it will not be detected.

or it is just that the watermark is just to preserve the source.

how does watermark help us preserve the integrity?

how will we know that it has been modified or not?

# OpenStego (<https://www.openstego.com/>)

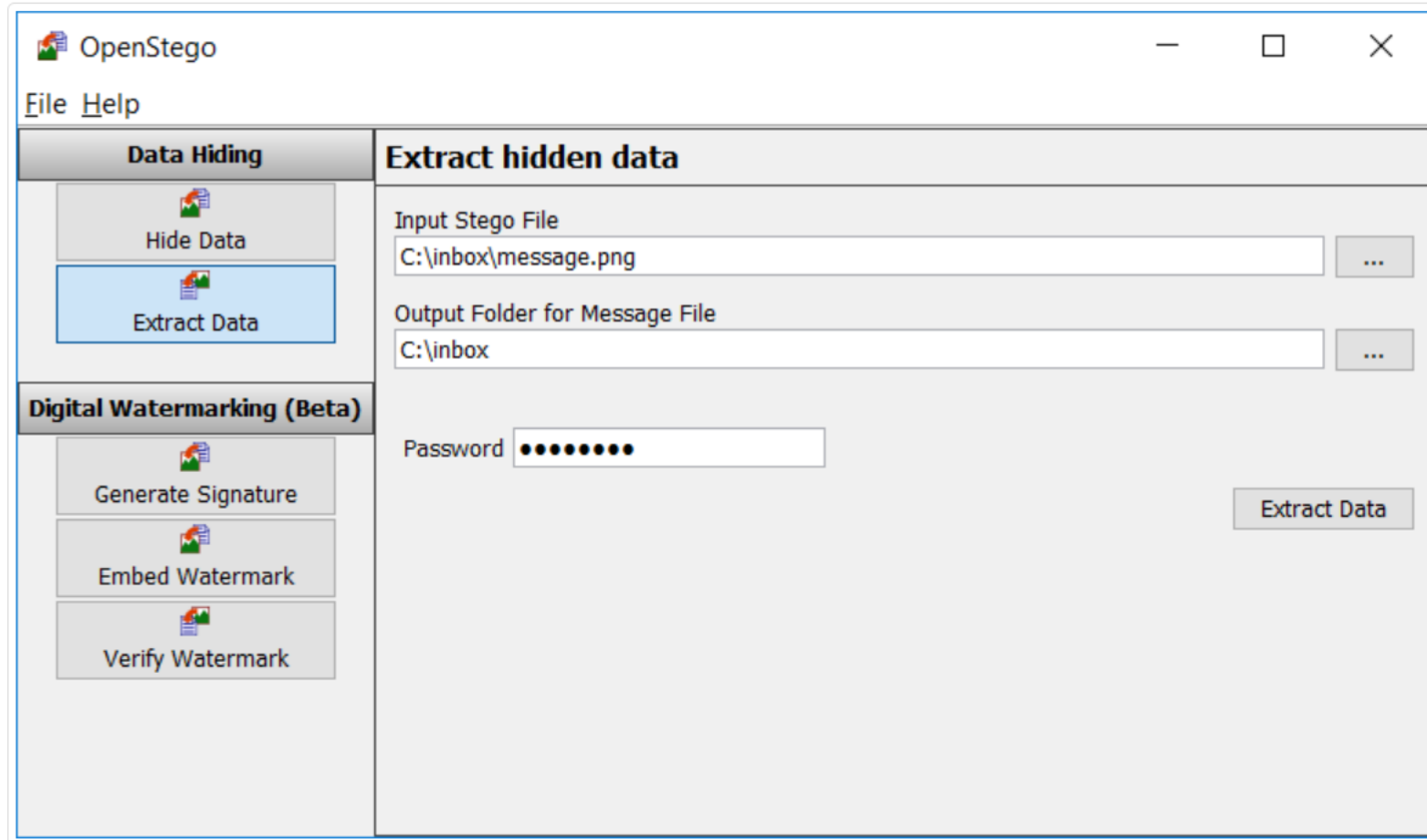


The screenshot shows the OpenStego application window. The title bar reads 'OpenStego'. Below the title bar is a menu bar with 'File' and 'Help'. The main interface is divided into two panes. The left pane has a tab labeled 'Data Hiding' and contains three buttons: 'Hide Data' (highlighted in blue), 'Extract Data', and a section titled 'Digital Watermarking (Beta)' which contains 'Generate Signature', 'Embed Watermark', and 'Verify Watermark'. The right pane is titled 'Hide data in harmless looking files' and contains the following fields and controls:

- Message File:** A text box containing 'C:\secret\message.txt' with a browse button (...).
- Cover File:** A text box containing 'C:\secret\nature.png' with a browse button (...). Below the text box is the instruction: *(Select multiple files or provide wildcard (\*, ?) to embed same message in multiple files)*.
- Output Stego File:** A text box containing 'C:\secret\send.png' with a browse button (...).
- Options:** A section containing:
  - Encryption Algorithm:** A dropdown menu set to 'AES128'.
  - Password:** A text box filled with 10 dots.
  - Confirm Password:** A text box filled with 10 dots.
- Hide Data:** A button at the bottom right of the right pane.

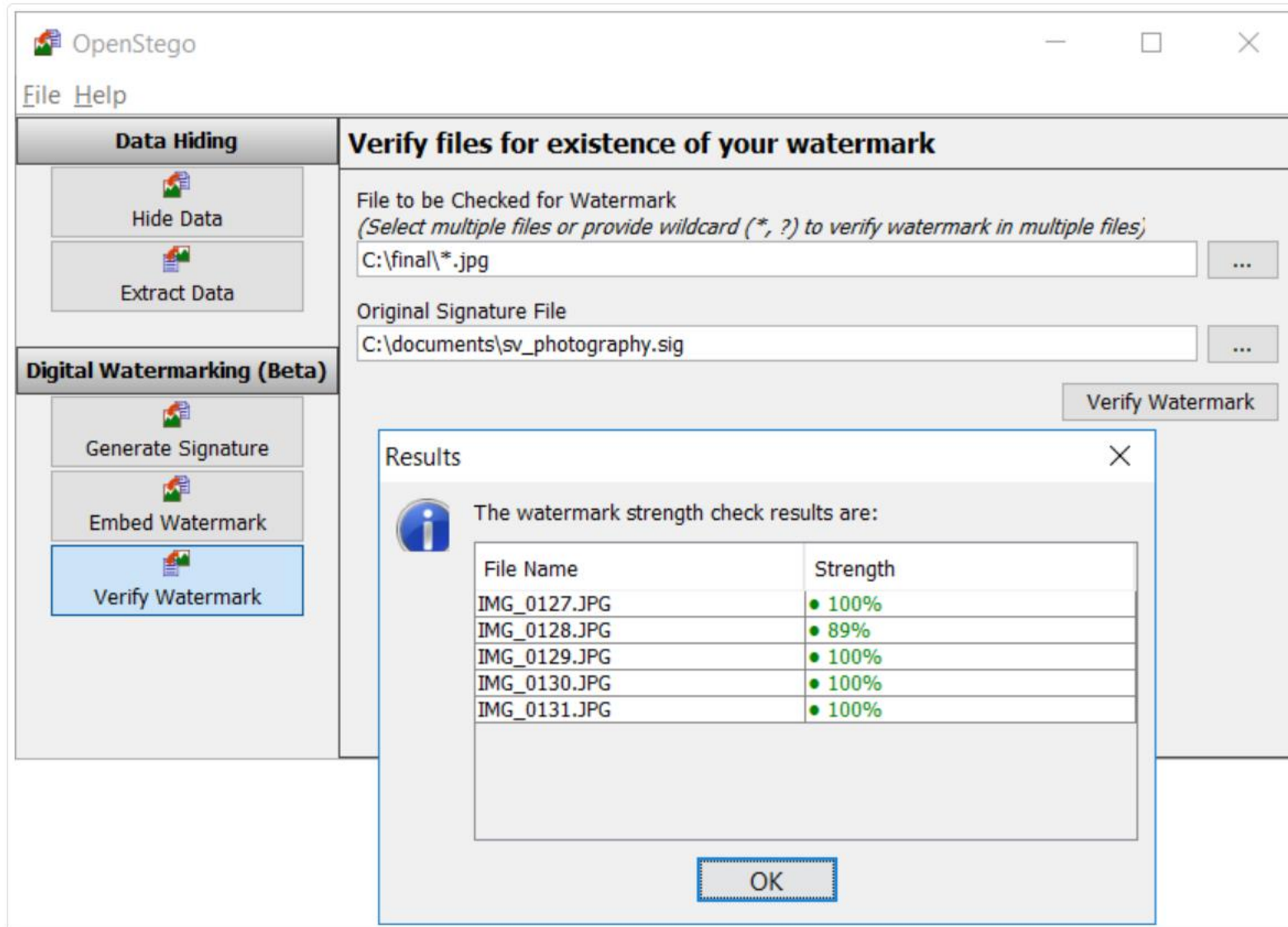


Extract data



what does that percentage indicates?  
what do you mean by the strength of the result.

Verify watermark



# Steganography - more

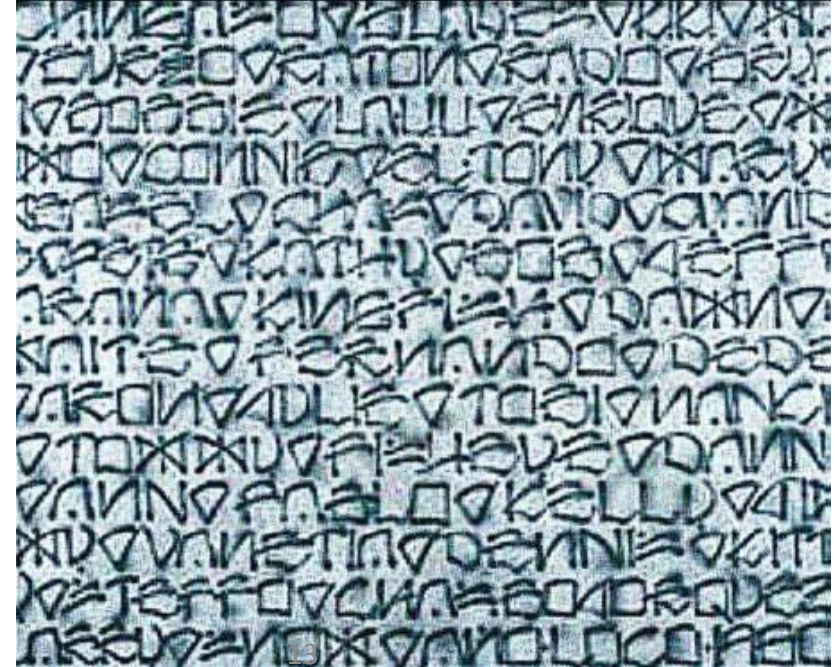
- Core Goal: The primary aim of steganography is to hide the existence of a message within a cover object.
- The hidden message itself doesn't hold any inherent value to the cover object.
- Applications - Steganography is often used for covert communication, where the goal is to transmit a secret message undetected by unintended recipients.
- It can also be used for applications like copyright protection, where the hidden message might identify the owner but not be readily noticeable.

# Hidden message within a message using Null Cipher

- Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on byproducts, ejecting suets and vegetable oils.
- pershingsailsfromnyjune1
- pershing sails from ny june 1
- Pershing sails from NY June 1

# More classical techniques

- Pin puncture,
- over written characters,
- writing using invisible ink,
- Modified alphabet
- Language translation



# Modern Steganography

- In a digital message, the secret information is inserted or "hidden" into the "container data" <any type of digital data file>
- It doesn't appear to be anything other than what it is eg. A picture or music file.
- An encrypted file on the other hand cries out '*I contain sensitive information!!!*'
- Computer files (images, sounds recordings, even disks) contain unused or insignificant areas of data.
- Steganography takes advantage of these areas, replacing them with information.

# More techniques

- Holography
- IR controls
- Pagers
- Colored glasses to filter wavelengths
- Inks – magnetic, thermo-chromic, photo-cromic
- Jargon speak
- Blank areas on Memory
- HTML code

least significant bit of every pixel => so that the information is stored inside it.

# Modern Techniques

- Least Significant Bit
- Low-Bit Encoding (image)
- Spread Spectrum
- Echo Data Hiding (audio)
- Perceptual Masking (audio—differential sounds)
- Discrete Cosine Transform (video)
- Textual steganography (spacing, coding)
- unused space in the packet headers

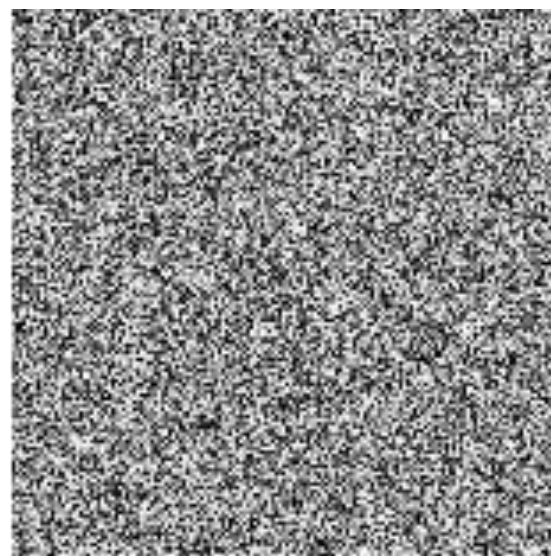


# Digital Watermarking

- Digital watermarking embeds a signal (watermark) within a digital object (image, audio, video) to convey specific information. The watermark can be used for various purposes, such as copyright protection, content authentication, or tracking distribution.
- Watermarking is widely used in multimedia content to deter unauthorized copying, identify ownership, or track distribution channels. It's also used in applications like document authentication and tamper detection.



Original Image



Watermark



Target image with  
watermark



# Watermarking

- watermarks might be designed to be detectable under specific circumstances (e.g., copyright verification software)



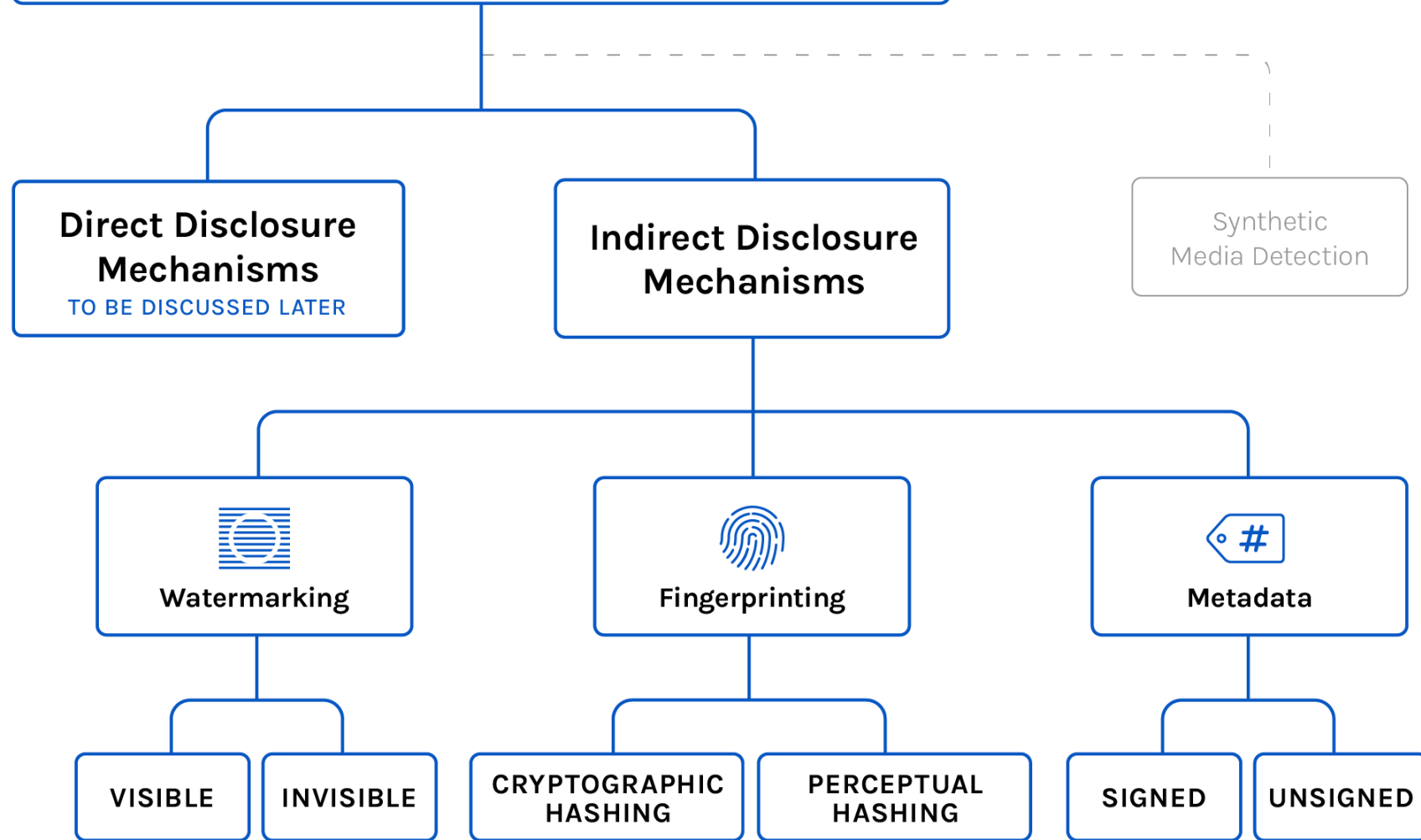




# Coalition for Content Provenance and Authenticity (C2PA)

- The C2PA standard is designed to tackle the issue of trust by tracking the origin and history of online assets via Content Credentials.
- An open technical standard providing publishers, creators, and consumers the ability to trace the origin of different types of media.
- This technology allows content creators to add provenance data to original digital assets in a standardized manner via cryptographically signed metadata attached to assets called C2PA manifests.
- For instance, Leica or Nikon digital cameras can add signed C2PA manifests to pictures at the source.

# Synthetic Media Transparency Methods



[syntheticmedia.partnershiponai.org](https://syntheticmedia.partnershiponai.org)

# Generative AI

- The Authors Guild and 17 authors filed a class-action suit against OpenAI in the Southern District of New York for copyright infringement of their works of fiction on behalf of a class of fiction writers whose works have been used to train GPT. (Sept 2023)
- They don't object to the development of generative AI, but OpenAI had no right to develop their AI technologies with unpermitted use of the authors' copyrighted works. OpenAI could have 'trained' their large language models on works in the public domain or paid a reasonable licensing fee to use copyrighted works."



# Biased AI – serious problem

- The UK health secretary, has announced a review into systemic racism and gender bias in medical devices in response to concerns it could contribute to poorer outcomes for women and people of colour. (Nov. 2021)
- Oximeters estimate the amount of oxygen in a person's blood, and are a crucial tool in determining which Covid patients may need hospital care
- Concerns have been raised, however, that the devices work less well for patients with darker skin. NHS England and the Medicines and Healthcare products Regulatory Agency (MHRA) say pulse oximeters can overestimate the amount of oxygen in the blood



# Other devices

- Respirator masks - Medical-grade respirators are crucial to help keep healthcare workers safe from Covid because they offer protection to the wearer against both large and small particles that others exhale.
- In order to offer the greatest protection, however, filtering face piece (FFP) masks must fit properly and research has shown they do not fit as well on people from some ethnic backgrounds!!
- Spirometers measure lung capacity, but experts have raised concerns that there are racial biases in the interpretation of data gathered from such gadgets!!

# Why?

- If we only train our AI using mostly data from white patients it cannot help our population as a whole.
- We need to make sure the data we collect is representative of our nation and the world (because every one is interested to get examined by imported medical machines!)
- <Such concerns were raised in relation to AI systems for diagnosing skin cancers.
- Researchers revealed that few freely available image databases that could be used to develop such AI are labelled with ethnicity or skin type.
- Of those that did have such information recorded, only a handful were of people recorded as having dark brown or black skin>

# Web 3

- Web3 is an idea for a new iteration of the World Wide Web which incorporates concepts such as decentralization, blockchain technologies, and token-based economics
- decentralized – rather than controlled by governments and corporations
- It gives power back to the users in the form of ownership

# Web 3

- Web 1.0 is the "read-only Web,"
- Web 2.0 is the "participative social Web," and
- Web 3.0 is the "read, write, execute Web."
- Web 3 is a decentralized and fair internet where users control their own data, identity and destiny
- A good example of a web3 trustless transaction would be sending Bitcoin directly to another person – not via an online exchange or wallet stored on a centralized server