

IDENTIFYING DEEPPFAKES & SYNTHETIC MEDIA

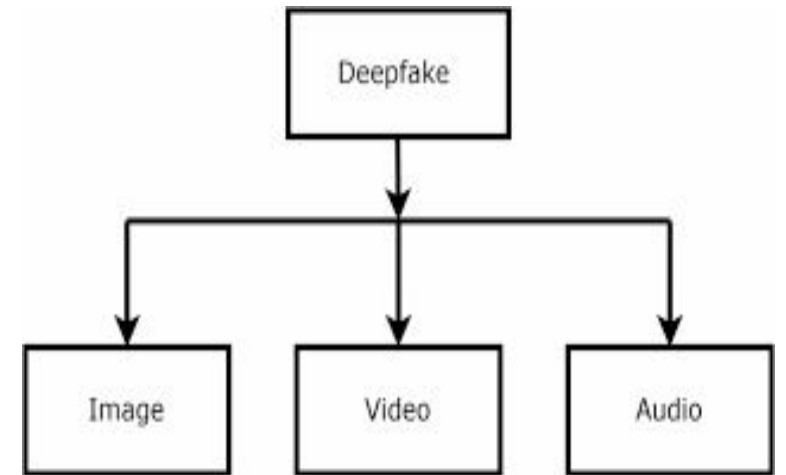
Dr. G. Sofia

*Associate Professor,
Department of Computer Science,
Lady Doak College,
Madurai.*

INTRODUCTION

INTRODUCTION

- **Deepfake** technology is a **powerful application** of **Artificial Intelligence** that can create **highly realistic fake videos, images, and audio**.
- **Deepfakes** are **synthetic media** created using advanced AI techniques, especially **Deep Learning**.
- Deepfakes are **AI-generated** or **manipulated** media: **Audio/ Video/ Images**.
- **Deepfake** is a media that has been created or altered using AI based tools or audio – video editing software.



- Synthetic media is a **digital content** in various digital formats including text, image, audio and video which has been **automatically** and **artificially** produced or manipulated.
- It refers to the use of Generative AI to produce the Deepfake content.
- They are primarily used for music synthesis, text generation , human image synthesis, speech synthesis etc.
- They manipulate or replace a person's face or voice to make it appear authentic.



COMMON USES

- **Film and entertainment** - Recreating younger versions of actors, Dubbing movies in different languages with realistic lip-sync. Reduces production cost and Saves time in editing.
- **Voice cloning** - Replicate a person's voice accurately. Audiobooks and narration, Dubbing content in multiple language, mimic celebrity voices.
- **Virtual assistants** - Human-like speech interaction, Customer service automation, Interactive learning tools, Assistants like Siri and Alexa can become more advanced using deepfake-based voice and facial simulation.
- **Social media content** - Entertainment videos and memes, Content creation, Creative expression, Increased engagement.

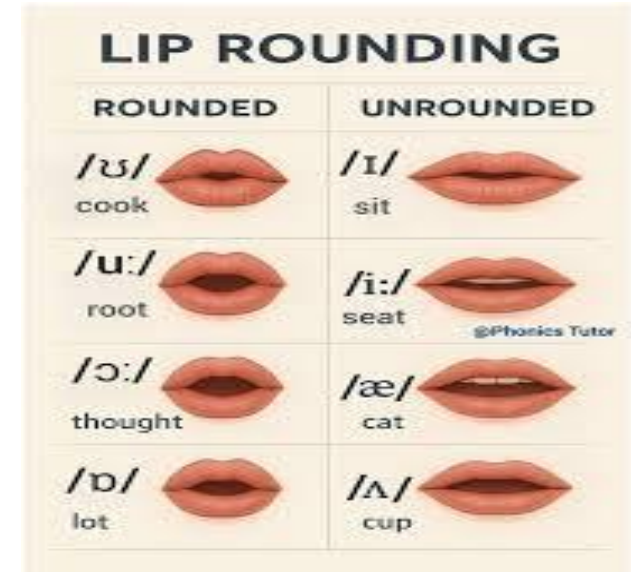
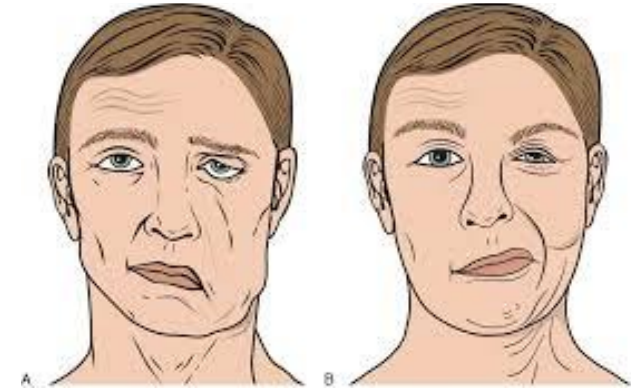
However, misuse has become a growing concern worldwide.



CLUES

VISUAL CLUES

- **Unnatural Facial Expressions** : Poor replication of natural human emotions
 - Smiles may look forced or uneven
 - Eyebrows and cheeks may not move together
 - Emotions may not match the situation
- **Lip-sync Mismatches** : Errors in aligning audio with facial movements
 - Words and mouth movement are slightly out of sync
 - Pronunciation doesn't match lip shape



- **Irregular Blinking** : AI struggling to mimic natural eye behavior
 - Too much blinking or no blinking at all
 - Blinking at odd intervals

- **Blurred Edges** : Imperfect blending of the fake face with the original video
 - Blurry areas around hair, ears, or jawline
 - Flickering edges when the person moves

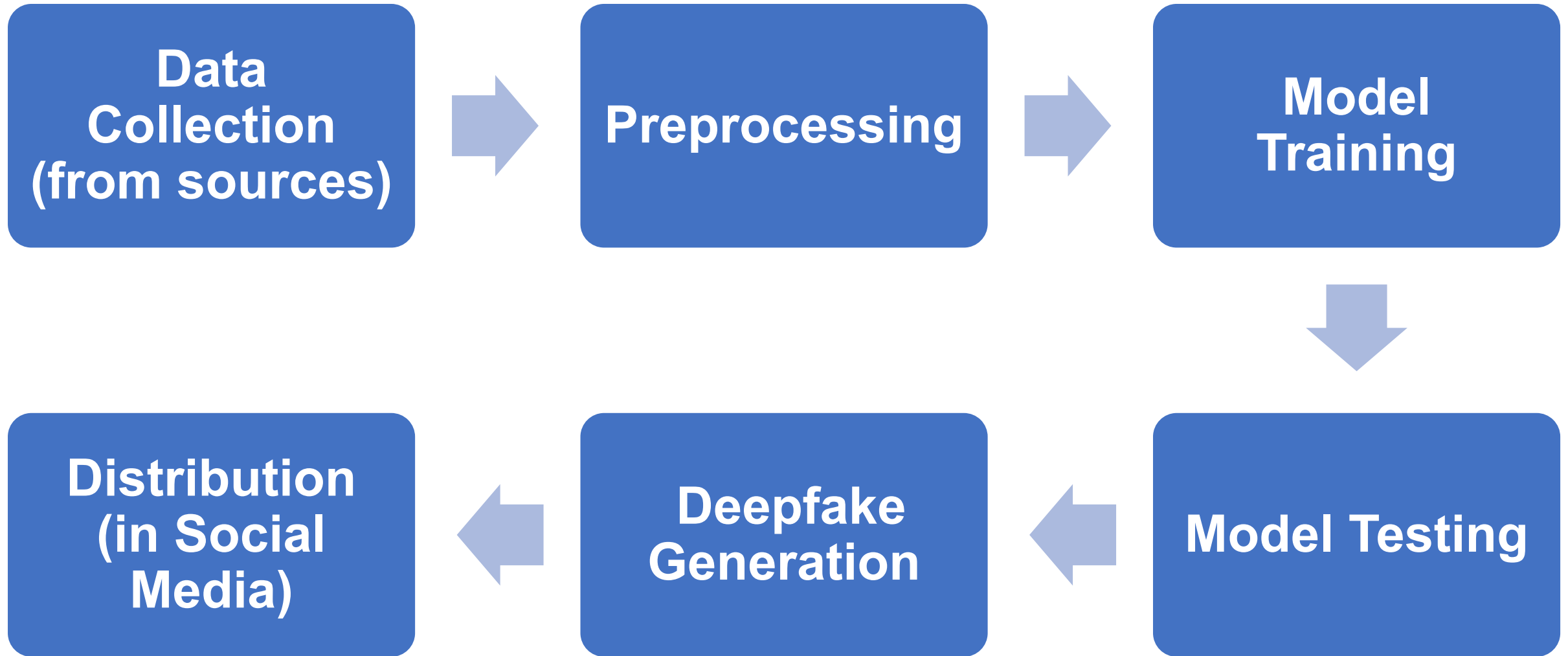
AUDIO CLUES

- **Robotic or Flat Tune :**
 - Sounds monotone (no rise and fall in pitch)
 - Lacks natural human emotion & Feels mechanical or artificial
- **Unnatural Pauses :**
 - Pauses appear in the wrong places
 - Speech rhythm feels uneven
 - Words may feel disconnected
- **Emotional Mismatch :**
 - Happy words spoken in a neutral or sad tone
 - Serious messages delivered without proper emotion
- **Pronunciation errors:**
 - Mispronounced names or difficult words
 - Incorrect stress on syllables
 - Unnatural accent shifts

CONTEXTUAL CHECKS

- **Out of character behaviour :**
 - Unnatural Behavior that contradicts their known personality or values (in Reels)
(https://youtube.com/shorts/_jRJgQ_1W_s?si=2ydrSJx9YpHi-ODk)
 - Sudden change in tone, language, or attitude
- **No reliable news coverage :**
 - Not reported by trusted sources
 - Seen only on random social media accounts
- **Multiple conflicting versions :**
 - Same video with different captions
 - Edited clips showing different meanings

TECHNOLOGY BEHIND THE TRICK



TECHNICAL INDICATORS IN DEEPPFAKE DETECTION

- Technical indicators are **scientific and reliable methods used to detect deepfakes** by analyzing the digital structure of videos, images, and audio.
- While visual and audio clues are useful for general users, technical methods offer **deeper** and more **accurate verification**.
- These methods are often used by experts in **Digital Forensics** and **Cyber Security**.
 - Meta Data Analysis
 - Frame-by-Frame Inspection
 - AI based Detection Models

- **Meta Data Analysis** : Verifying the origin and authenticity of a file - Date and time of creation, Device used (camera, software), Editing history
- **Frame-by-Frame Inspection** : Analyzing a video one frame at a time - Sudden glitches between frames, Distorted facial features, Inconsistent lighting or shadows, Flickering or unnatural transitions
- **AI based Detection Models** :Advanced tools built using AI(ML,DL) to detect deepfakes automatically - Analyze patterns in images and videos, Detect anomalies in facial movements and textures, Compare real vs fake data.

RULE OF THUMB

If the content is

SHOCKING, VIRAL, UNVERIFIED

Treat it as suspicious ...



SIFT METHOD

SIFT METHOD

SIFT Method : To identify the type of information

SIFT stands for

S – Stop

I – Investigate the Source

F – Find Better Coverage

T – Trace Claims back to the Original Context

Step 1 : Stop

- The first step is to **pause before reacting**.
- When you see shocking, emotional, or surprising information online, your immediate reaction might be to share it. This is exactly how misinformation spreads.

Key actions:

- Do not share immediately
- Take a moment to think
- Ask: “Is this too good (or bad) to be true?”

Stopping helps to prevent the spread of false information.

Step 2 : Investigate the Source

- Before trusting information, check **who created it**.

Key Actions (Questions to ask) :

- Is the source reliable?
- Is it a well-known organization or an unknown website?
- What is the author's background?

Information from official news websites is generally more reliable than random social media posts.

Step 3 : Find Better Coverage

- Look for the same information from **multiple trusted sources**.

Key Actions:

- Search the topic on Google
- Compare different news sources
- Avoid relying on a single post
- This step helps to confirm accuracy

If the information is true, it will be reported by several reliable platforms.

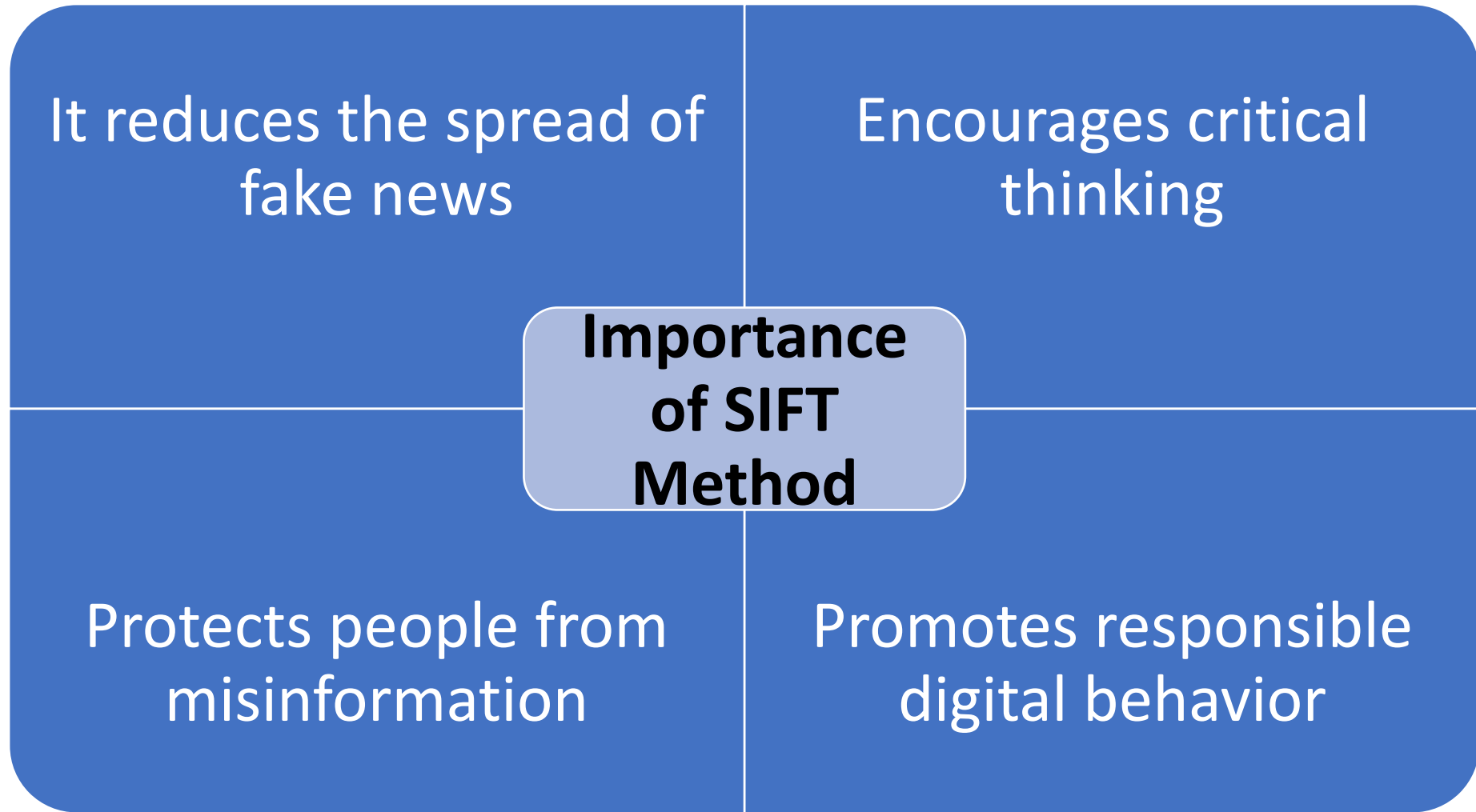
Step 4 : Trace Claims to Original Context

- Many posts take information **out of context**.

Key Actions:

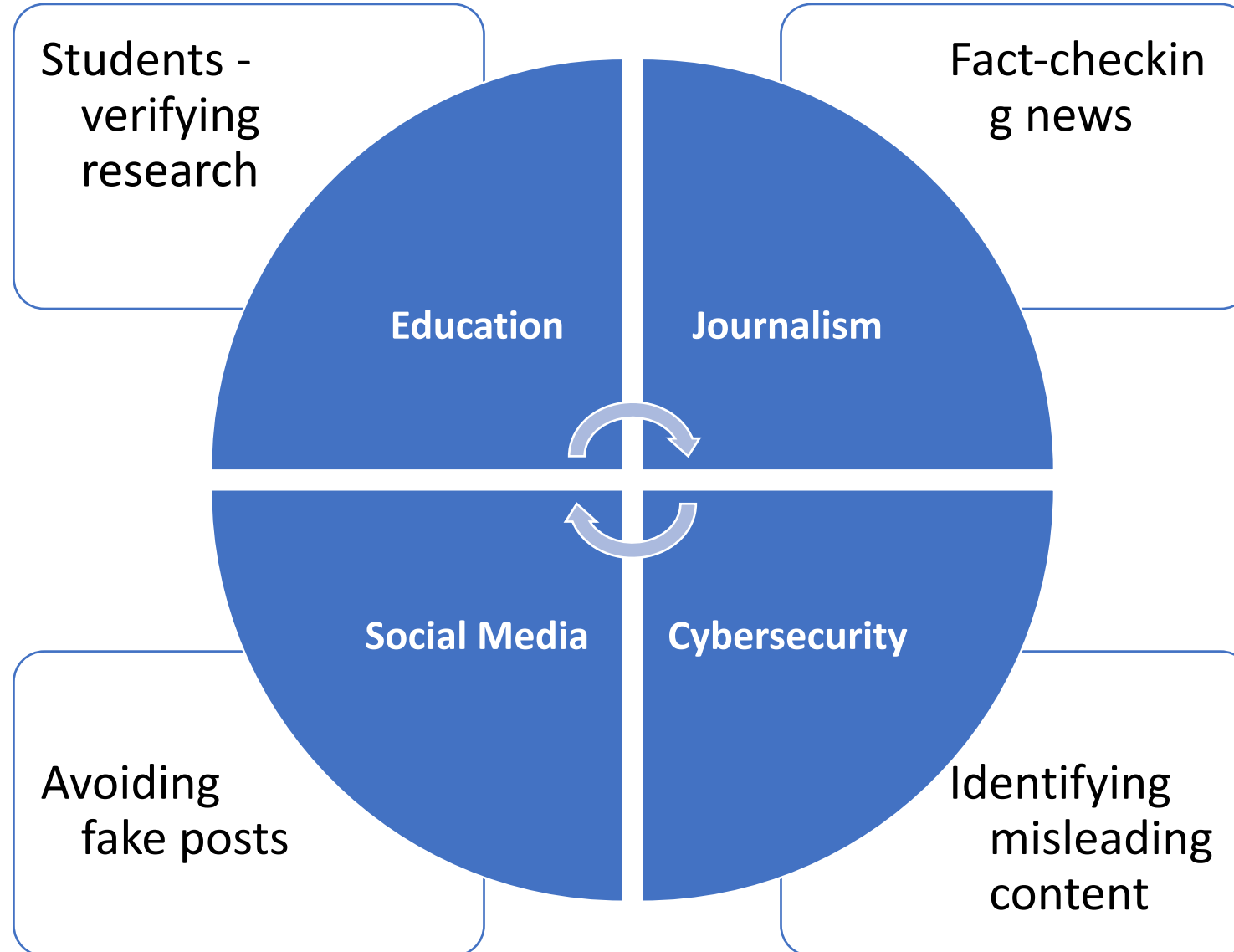
- Find the original source of the image, video, or statement
- Check if it has been edited or misused
- Verify dates and locations

An old video may be shared as if it is a current event



In an age of deepfakes and AI-generated content, this method is more relevant than ever.

APPLICATIONS OF SIFT METHOD



LEGAL AND ETHICAL BOUNDARIES

LEGAL & ETHICAL BOUNDARIES

- While Deepfake has useful applications in entertainment and education, it also raises **serious legal and ethical concerns**.
- With the increasing misuse of deepfakes for spreading misinformation, fraud, and harassment, it is important to understand the boundaries that govern their use.

LAWS AND REGULATIONS

- Different countries are taking steps to control deepfakes:
 - Some regions have laws against **non-consensual deepfake content**
 - Strict penalties for **cybercrime and impersonation**
 - Regulations under **data protection laws**
- However, enforcement remains a challenge due to rapid technological advancement.

ETHICAL ISSUES OF DEEPPFAKES

- Even when legal, deepfakes may still be unethical.
- **Key ethical concerns:**
 - Lack of consent
 - Misleading audiences
 - Manipulating public opinion
 - Loss of trust in digital media
- Ethics focuses on what is right and responsible, beyond what is legal.

IMPACT ON SOCIETY

POSITIVE

- Movies and visual effects
- Education and training simulations
- Historical recreations
- Accessibility (voice assistance)

NEGATIVE

- Spread of fake news and misinformation
- Political manipulation
- Cyberbullying and harassment
- Loss of trust in media and institutions
- They can influence public opinion and even affect election

The challenge is to balance innovation with responsibility ...

PREVENTIVE MEASURES

To control misuse of deepfakes:

- Stronger laws and regulations
- Public awareness and education
- Use of detection tools
- Ethical guidelines for developers
- Responsible use of AI technology


Individuals must also verify information before sharing.

CLASSROOM RED FLAGS

CLASSROOM RED FLAGS

Signs that content might be misleading or fake

1. NO RELIABLE SOURCE
The information has no trustworthy source.



2. SHOCKING OR UNBELIEVABLE
It is too shocking or unbelievable to be true.

OMG!
Is this REAL?



3. VIRAL BUT UNVERIFIED
It is widely shared but not verified by experts or official sources.

10K+ SHARES



4. EMOTIONAL MANIPULATION
It tries to make you angry, scared or excited to get a reaction.



5. NO OTHER NEWS COVERAGE
No other trusted news outlets are reporting the same story.

NEWS
NO RESULTS



6. MULTIPLE CONFLICTING VERSIONS
Different versions of the same story do not match.

A ≠ B



THINK BEFORE YOU SHARE!
Check. Verify. Then Believe.



CONCLUSION

CONCLUSION

- Deepfake technology is a **double-edged sword**. While it offers exciting possibilities, it also poses serious risks if misused
- In today's digital age, **being informed** is not enough; we must also be **responsible**.
- Understanding its **legal and ethical boundaries** is essential to ensure it is used responsibly.
- Governments, organizations, and individuals must work together to create a **safer digital environment**.
- Can be detected with careful observation and Tools.

THANK YOU !!!