

डीपफेक और सिंथेटिक मीडिया की पहचान करना

परिचय

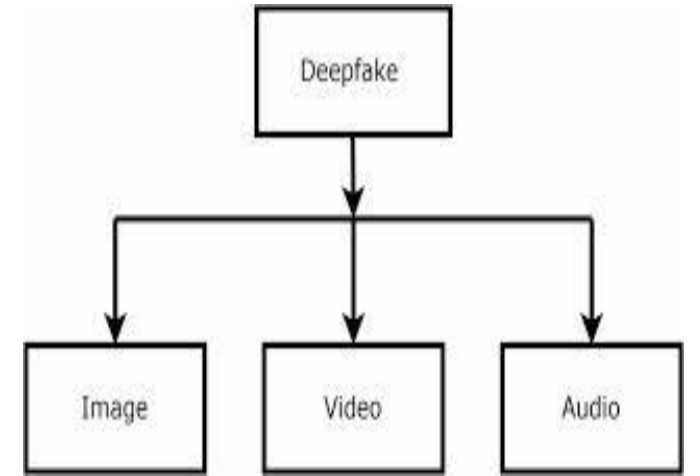
परिचय

• डीपफेक तकनीक आर्टिफिशियल इंटेलिजेंस का एक शक्तिशाली अनुप्रयोग है जो अत्यधिक यथार्थवादी नकली वीडियो, चित्र और ऑडियो बना सकता है।

डीपफेक सिंथेटिक मीडिया हैं जो उन्नत एआई तकनीकों, विशेष रूप से डीप लर्निंग का उपयोग करके बनाए गए हैं।

डीपफेक एआई-जनरेटेड या हेरफेर किए गए मीडिया हैं: ऑडियो/वीडियो/छवियां।

डीपफेक एक ऐसा मीडिया है जिसे एआई आधारित टूल या ऑडियो-वीडियो एडिटिंग सॉफ्टवेयर का उपयोग करके बनाया या बदला गया है।



•सिंथेटिक मीडिया टेक्स्ट, इमेज, ऑडियो और वीडियो सहित विभिन्न डिजिटल प्रारूपों में एक डिजिटल सामग्री है जिसे स्वचालित रूप से और कृत्रिम रूप से उत्पादित या हेरफेर किया गया है। यह डीपफेक सामग्री का उत्पादन करने के लिए जनरेटिव एआई के उपयोग को संदर्भित करता है। वे मुख्य रूप से संगीत संश्लेषण, पाठ निर्माण, मानव छवि संश्लेषण, भाषण संश्लेषण आदि के लिए उपयोग किए जाते हैं। वे किसी व्यक्ति के चेहरे या आवाज को प्रामाणिक दिखाने के लिए उसमें हेरफेर करते हैं या उसे बदल देते हैं।



सामान्य उपयोग

- **फिल्म और मनोरंजन** - अभिनेताओं के युवा संस्करणों को फिर से बनाना, अलग-अलग फिल्मों को डब करना
यथार्थवादी लिप-सिंक वाली भाषाएँ। उत्पादन लागत कम करता है और संपादन में समय बचाता है।
वाँयस क्लोनिंग - किसी व्यक्ति की आवाज़ को सटीक रूप से दोहराएँ। ऑडियोबुक और कथन, कई भाषाओं में सामग्री डब करना, सेलिब्रिटी आवाज़ों की नकल करना।
 - **वर्चुअल असिस्टेंट** - ह्यूमन-लाइक स्पीच इंटरैक्शन, कस्टमर सर्विस ऑटोमेशन, इंटरएक्टिव लर्निंग टूल्स, सिरी और एलेक्सा जैसे असिस्टेंट डीपफेक-आधारित वाँयस और फेशियल सिमुलेशन का उपयोग करके अधिक उन्नत हो सकते हैं।
सोशल मीडिया सामग्री - मनोरंजन वीडियो और मीम्स, सामग्री निर्माण, रचनात्मक अभिव्यक्ति, बढ़ी हुई व्यस्तता।
- हालाँकि, दुरुपयोग दुनिया भर में एक बढ़ती चिंता का विषय बन गया है।



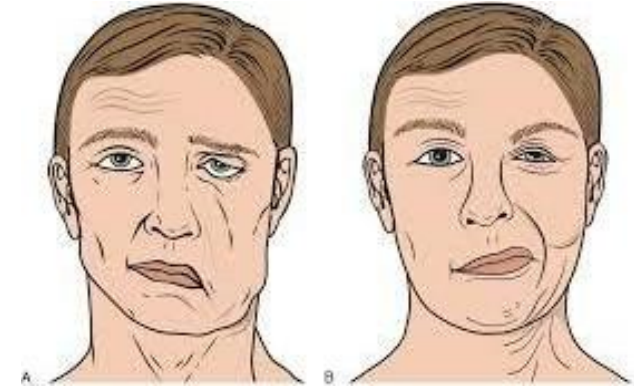
सुरा
ग

दृश्य सुराग

- **अप्राकृतिक चेहरे के भाव:** प्राकृतिक की खराब प्रतिकृति

□ मानवीय भावनाएँ

मुस्कान मजबूर या असमान लग सकती है
भौहें और गाल एक साथ नहीं चल सकते हैं
भावनाएं स्थिति से मेल नहीं खा सकती हैं



- **लिप-सिंक बेमेल :** चेहरे की गतिविधियों के साथ आँडियों को संरेखित करने में त्रुटियाँ

□ शब्द और मुंह की गति थोड़ी सिंक से बाहर हैं
उच्चारण होंठ के आकार से मेल नहीं खाता है

LIP ROUNDING	
ROUNDED	UNROUNDED
/ʊ/ cook	/ɪ/ sit
/uː/ root	/iː/ seat
/ɔː/ thought	/æ/ cat
/ɒ/ lot	/ʌ/ cup

• **अनियमित पलक झपकना** : एआई प्राकृतिक आंखों के व्यवहार की नकल करने के लिए संघर्ष कर रहा है

- बहुत अधिक पलक झपकना या बिल्कुल भी पलक नहीं झपकना विषम अंतराल पर पलक झपकना

• **धुंधले किनारे**: मूल वीडियो के साथ नकली चेहरे का अपूर्ण सम्मिश्रण

- बालों, कानों या जबड़े के आसपास के धुंधले क्षेत्र जब व्यक्ति चलता है तो टिमटिमाते किनारे

ऑडियो सुराग

□ रोबोटिक या फ्लैट ट्यून:

- मोनोटोन लगता है (पिच में कोई वृद्धि और गिरावट नहीं)
- प्राकृतिक मानवीय भावनाओं का अभाव है और यांत्रिक या कृत्रिम लगता है

□ अप्राकृतिक ठहराव :

- विराम गलत जगहों पर दिखाई देते हैं
- भाषण की लय असमान लगती है
शब्द अलग-थलग महसूस हो सकते हैं

भावनात्मक बेमेल :

- तटस्थ या उदास स्वर में बोले गए सुखद शब्द
- उचित भावना के बिना दिए गए गंभीर संदेश

□ उच्चारण त्रुटियाँ:

- गलत नाम या कठिन शब्द
शब्दांशों पर गलत तनाव
अप्राकृतिक उच्चारण बदलाव

प्रासंगिक जांच

□ चरित्र व्यवहार से बाहर:

- अप्राकृतिक व्यवहार जो उनके ज्ञात व्यक्तित्व या मूल्यों का खंडन करता है (रीलों में)

(https://youtube.com/shorts/_jRJgQ_1W_s?si=2ydrSJx9YpHi-ODk)

- स्वर, भाषा या दृष्टिकोण में अचानक परिवर्तन

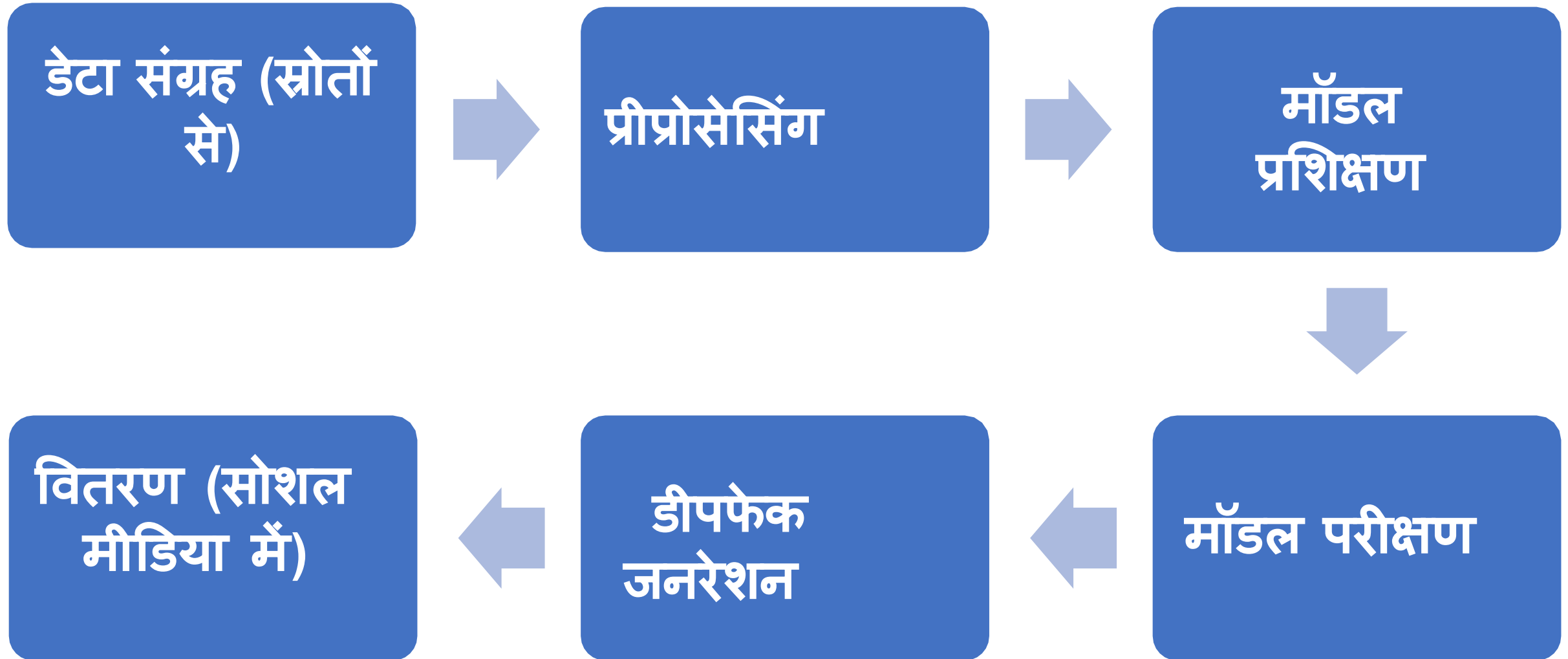
□ कोई विश्वसनीय समाचार कवरेज नहीं :

- विश्वसनीय स्रोतों द्वारा रिपोर्ट नहीं किया गया
- केवल रैंडम सोशल मीडिया अकाउंट्स पर देखा गया

□ एकाधिक परस्पर विरोधी संस्करण:

- अलग-अलग कैप्शन के साथ एक ही वीडियो
- अलग-अलग अर्थ दिखाने वाली संपादित क्लिप

चाल के पीछे की तकनीक



डीपफेक डिटेक्शन में तकनीकी संकेतक

- तकनीकी संकेतक **वैज्ञानिक और विश्वसनीय तरीके** हैं जिनका उपयोग डीपफेक का पता लगाने के लिए किया जाता है, वीडियो, छवियों और ऑडियो की डिजिटल संरचना का विश्लेषण करके। जबकि दृश्य और श्रव्य सुराग सामान्य उपयोगकर्ताओं के लिए उपयोगी हैं, तकनीकी तरीके प्रदान करते हैं **गहरा और अधिक सटीक सत्यापन**। इन विधियों का उपयोग अक्सर **डिजिटल फॉरेंसिक और साइबर सुरक्षा** के विशेषज्ञों द्वारा किया जाता है।

- **मेटा डेटा विश्लेषण**
फ्रेम-दर-फ्रेम निरीक्षण
एआई आधारित डिटेक्शन मॉडल

- **मेटा डेटा विश्लेषण:** किसी फ़ाइल की उत्पत्ति और प्रामाणिकता की पुष्टि करना - दिनांक और निर्माण का समय, प्रयुक्त उपकरण (कैमरा, सॉफ्टवेयर), संपादन इतिहास
- **फ्रेम-दर-फ्रेम निरीक्षण:** एक समय में एक फ्रेम के वीडियो का विश्लेषण - फ्रेम के बीच अचानक गड़बड़ी, विकृत चेहरे की विशेषताएं, असंगत प्रकाश या छाया, झिलमिलाहट या अप्राकृतिक संक्रमण
- **एआई आधारित डिटेक्शन मॉडल:** स्वचालित रूप से डीपफेक का पता लगाने के लिए एआई (एमएल, डीएल) का उपयोग करके निर्मित उन्नत उपकरण - छवियों और वीडियो में पैटर्न का विश्लेषण करें, चेहरे की गतिविधियाँ और बनावट में विसंगतियों का पता लगाएं, वास्तविक बनाम नकली डेटा की तुलना करें।

कायदा

अगर सामग्री है चौंकाने वाला, वायरल, असत्यापित
इसे संदिग्ध मानें ...



SIFT विधि

SIFT विधि

SIFT विधि: जानकारी के प्रकार की पहचान करने के लिए

SIFT का अर्थ है

S – स्टॉप

I - स्रोत की जांच करें

F - बेहतर कवरेज खोजें

T - ट्रेस मूल संदर्भ में वापस दावा करता है

चरण 1: रुकें

- पहला कदम **प्रतिक्रिया करने से पहले रुकना है।**
जब आप ऑनलाइन चौंकाने वाली, भावनात्मक या आश्चर्यजनक जानकारी देखते हैं, तो आपकी तत्काल प्रतिक्रिया इसे साझा करने की हो सकती है। ठीक इसी तरह गलत सूचना फैलती है।

मुख्य कार्रवाई :

- तुरंत शेयर न करें
सोचने के लिए कुछ समय निकालें
पूछो: "क्या यह सच होने के लिए बहुत अच्छा (या बुरा) है?"

रुकने से झूठी जानकारी के प्रसार को रोकने में मदद मिलती है।

चरण 2 : स्रोत की जांच करें

- जानकारी पर भरोसा करने से पहले, जांच लें कि इसे किसने बनाया है।

मुख्य क्रियाएं (पूछने के लिए प्रश्न):

- क्या स्रोत विश्वसनीय है?
क्या यह एक प्रसिद्ध संगठन या एक अज्ञात वेबसाइट है?
लेखक की पृष्ठभूमि क्या है?

आधिकारिक समाचार वेबसाइटों की जानकारी आम तौर पर यादृच्छिक सोशल मीडिया पोस्ट की तुलना में अधिक विश्वसनीय होती है।

चरण 3: बेहतर कवरेज खोजें

- कई विश्वसनीय स्रोतों से एक ही जानकारी देखें।

मुख्य कार्रवाई :

- Google पर विषय खोजें
विभिन्न समाचार स्रोतों की तुलना करें
एक पोस्ट पर निर्भर रहने से बचें
यह कदम सटीकता की पुष्टि करने में मदद करता है

यदि जानकारी सत्य है, तो इसे कई विश्वसनीय प्लेटफार्मों द्वारा रिपोर्ट किया जाएगा।

चरण 4 : मूल संदर्भ के दावों का पता लगाएं

- कई पोस्ट जानकारी को संदर्भ से बाहर ले जाते हैं।

मुख्य कार्रवाई :

- छवि, वीडियो या कथन का मूल स्रोत ढूँढें
जांचें कि क्या इसे संपादित किया गया है या दुरुपयोग किया गया है
दिनांक और स्थान सत्यापित करें

एक पुराने वीडियो को ऐसे साझा किया जा सकता है जैसे कि यह एक वर्तमान घटना है

यह फर्जी खबरों के प्रसार
को कम करता है

आलोचनात्मक सोच
को प्रोत्साहित
करता है

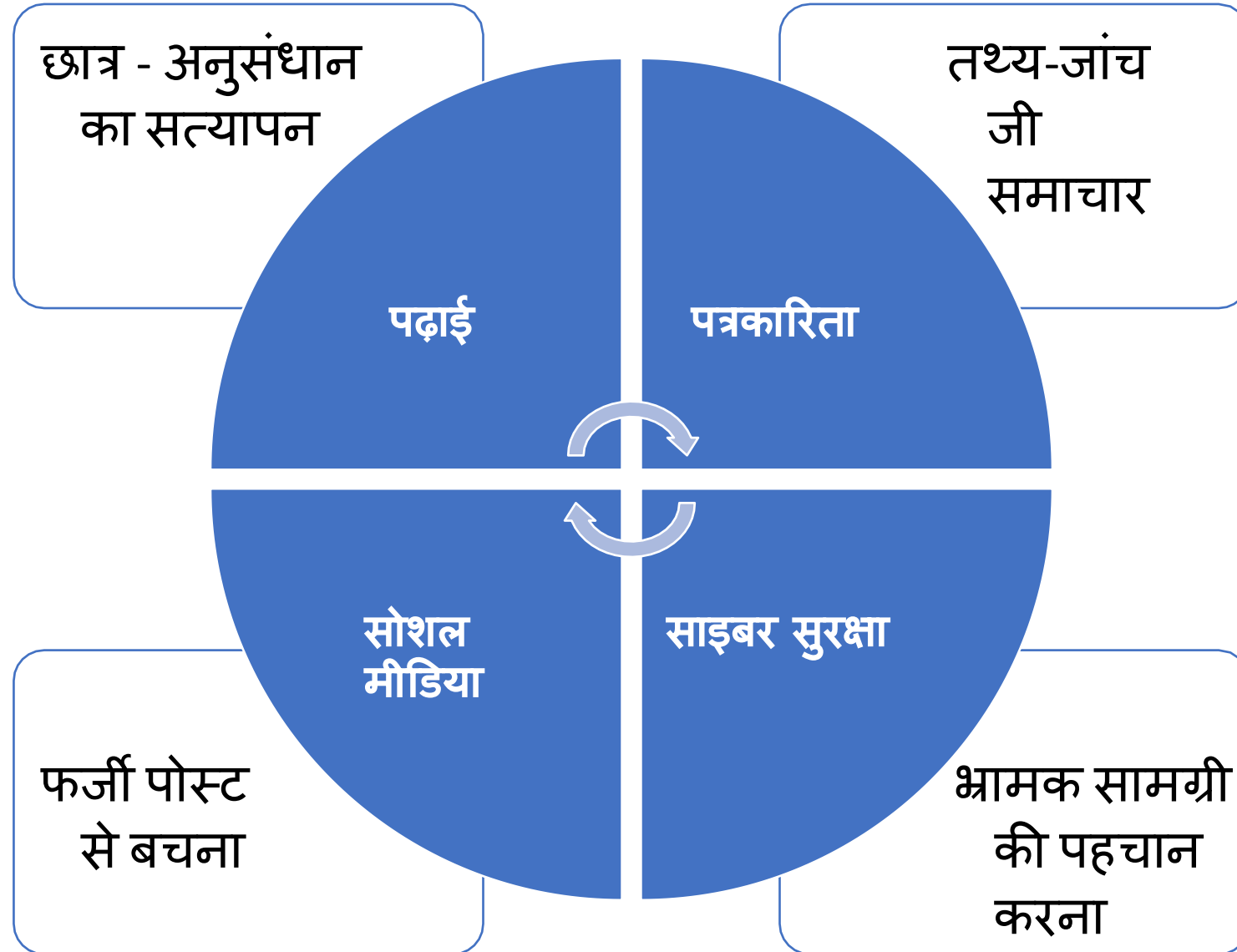
SIFT विधि का महत्व

लोगों को गलत सूचना
से बचाता है

जिम्मेदार डिजिटल
व्यवहार को बढ़ावा
देता है

डीपफेक और एआई-जनित सामग्री के युग में, यह विधि पहले से कहीं अधिक प्रासंगिक है।

SIFT विधि के अनुप्रयोग



कानूनी और नैतिक सीमाएँ

कानूनी और नैतिक सीमाएँ

- डीपफेक के मनोरंजन और शिक्षा में उपयोगी अनुप्रयोग हैं, यह गंभीर कानूनी और नैतिक चिंताएँ भी पैदा करता है।
- डीपफेक के गलत सूचना, धोखाधड़ी और उत्पीड़न के बढ़ते प्रसार के लिए दुरुपयोग के साथ, यह उन सीमाओं को समझता है जो उनके महत्वपूर्ण उपयोग को नियंत्रित करती हैं।

कानून और विनियम

- डीपफेक को नियंत्रित करने के लिए विभिन्न देश कदम उठा रहे हैं-
- कुछ क्षेत्रों में **गैर-सहमति वाली डीपफेक सामग्री** के खिलाफ कानून हैं
साइबर अपराध और प्रतिरूपण के लिए सख्त दंड
डेटा संरक्षण कानूनों के तहत विनियम
- हालाँकि, तेजी से तकनीकी प्रगति के कारण प्रवर्तन एक चुनौती बनी हुई है।

डीपफेक के नैतिक मुद्दे

- यहां तक कि जब कानूनी हो, डीपफेक अभी भी अनैतिक हो सकते हैं।

□ प्रमुख नैतिक चिंताएँ:

- सहमति का अभाव
दर्शकों को गुमराह करना
जनता की राय में हेरफेर करना
- डिजिटल मीडिया में विश्वास की कमी

नैतिकता इस बात पर ध्यान केंद्रित करती है कि कानूनी क्या सही और जिम्मेदार है, जो कानूनी है।

समाज पर प्रभाव

पॉज़िटिव

- फिल्मों और दृश्य प्रभाव
शिक्षा और प्रशिक्षण सिमुलेशन
ऐतिहासिक मनोरंजन
अभिगम्यता (आवाज सहायता)

नेगटिव

- फर्जी खबरों का प्रसार और
गलत सूचना
राजनीतिक हेरफेर
साइबरबुलिंग और उत्पीड़न
मीडिया और संस्थानों में विश्वास की
कमी
वे जनता की राय को प्रभावित कर
सकते हैं और यहां तक कि चुनाव को
भी प्रभावित कर सकते हैं

चुनौती जिम्मेदारी के साथ नवाचार को संतुलित करने की है ...

निवारक उपाय

डीपफेक के दुरुपयोग को नियंत्रित करने के लिए:

- मजबूत कानून और नियम
जन जागरूकता और शिक्षा
पता लगाने वाले उपकरणों का उपयोग
डेवलपर्स के लिए नैतिक दिशानिर्देश
एआई तकनीक का जिम्मेदार उपयोग

साझा करने से पहले व्यक्तियों को जानकारी सत्यापित भी करनी चाहिए।

कक्षा के लाल झंडे

CLASSROOM RED FLAGS

Signs that content might be misleading or fake

<p>1. NO RELIABLE SOURCE The information has no trustworthy source.</p> 	<p>2. SHOCKING OR UNBELIEVABLE It is too shocking or unbelievable to be true.</p> 	<p>3. VIRAL BUT UNVERIFIED It is widely shared but not verified by experts or official sources.</p> 
<p>4. EMOTIONAL MANIPULATION It tries to make you angry, scared or excited to get a reaction.</p> 	<p>5. NO OTHER NEWS COVERAGE No other trusted news outlets are reporting the same story.</p> 	<p>6. MULTIPLE CONFLICTING VERSIONS Different versions of the same story do not match.</p> 

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

निष्कर्ष

निष्कर्ष

- डीपफेक तकनीक एक **दोधारी तलवार** है। जबकि यह रोमांचक संभावनाएं प्रदान करता है, दुरुपयोग होने पर यह गंभीर जोखिम भी पैदा करता है। आज के डिजिटल युग में, **सूचित होना** पर्याप्त नहीं है; हमें होना चाहिए **जिम्मेदार**। इसकी **कानूनी और नैतिक सीमाओं** को समझना यह सुनिश्चित करने के लिए आवश्यक है कि इसका उपयोग जिम्मेदारी से किया जाए। सरकारों, संगठनों और व्यक्तियों को एक **सुरक्षित डिजिटल वातावरण** बनाने के लिए मिलकर काम करना चाहिए। सावधानीपूर्वक अवलोकन और उपकरणों से पता लगाया जा सकता है।

धन्यवाद!!!