सत्यमेव जयते
Ministry of Education
Government of India

विद्यया ऽ मृतमश्नुते
एन सी ई आर टी
NCERT

CyberPeace

CYBERPEACE CORPS · COGNITIVE SECURITY SERIES

Theme 1 · Information Integrity & Cognitive Security

Information Integrity Fundamentals

Understanding accuracy, authenticity, and trust in the information we create, share, and consume every day.

WHAT YOU WILL LEARN

Session Roadmap

सत्यमेव जयते
Ministry of Education
Government of India

विद्यया ऽ मृतमश्नुते
एन सी ई आर टी
NCERT

CyberPeace

1. Information Integrity Fundamentals

The principles of accuracy, authenticity, and trust that anchor reliable information.

2. Verifying Sources & Content Authenticity

Practical tools and techniques to check whether content is genuine and trustworthy.

3. Integrity in Digital Communications

Applying these principles to email, messaging, and social media in daily life.

Why Information Integrity Matters

- Every day, billions of people make decisions — about health, money, elections, and relationships — based on information they encounter online.
- When that information is inaccurate, fabricated, or manipulated, real harm follows: wrong medical choices, financial fraud, damaged reputations, and eroded public trust.
- Information integrity is the foundation that lets individuals, institutions, and societies function on shared, reliable facts.
- This session builds the core knowledge you need to create, verify, and share information responsibly.

Information Integrity means information that can be trusted: accurate, authentic, and handled responsibly across its lifecycle.

Information Integrity Fundamentals

- Authenticity, Accuracy and Trust- principles of Information Integrity and Cognitive Security
- Misinformation, Disinformation and Fake News
- Need for Verifying Sources and Content

ALTERED MEDIA – COMMUNICATION & INFO

- Single Device for News, Views and Chat

ALTERED MEDIA – COMMUNICATION & INFO

The mobile is mightier than the Sword

COMMUNICATION ?
EDUCATION TOOL?
NETWORKING ?
PUBLIC INFORMATION ?
BROADCAST AGENT ?
PAYMENT TOOL?

- Single Device for News, Views and Chat

YOU ARE THE TARGET AND WEAPON

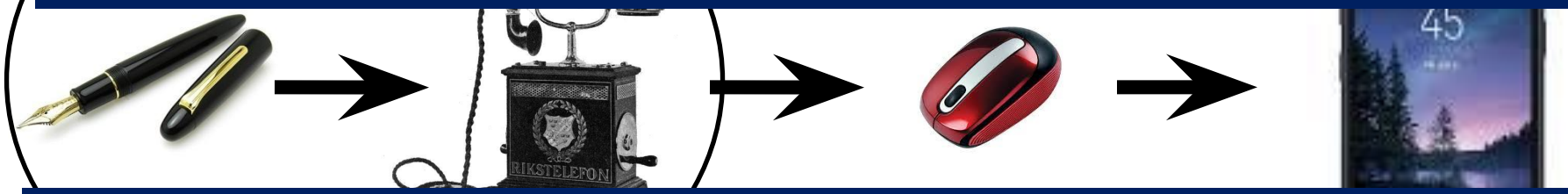
ONE TO ONE

ONE TO MANY

YOU ARE THE TARGET AND WEAPON

ONE TO ONE

UNVERIFIED BULK TRANSMISSION



ABSENCE OF EDITORIAL OVERSIGHT

ONE TO MANY

IMPACT OF AI- MORE CREDIBLE IW TOOLS, OVERRIDING REALITY

Information Integrity Fundamentals

- Authenticity, Accuracy and Trust- principles of Information Integrity and Cognitive Security
- Misinformation, Disinformation and Fake News
- Need for Verifying Sources and Content

'There's No Market Place Of Facts' : CJI DY Chandrachud Expresses Concerns About Online Disinformation & Hate Speech

LIVELAW NEWS NETWORK

2 Dec 2023 10:35 AM

“Simply by virtue of the scale of dissemination, fake news drowns out true information, replacing the character of discourse from truth-seeking to the loudest voice,”

If we cannot agree on the veracity of basic facts, debate stops, partisanship hardens, and social solidarity breaks down

A cursory glance at newspapers brings to the fore violence resulting out of rumours and disinformation campaigns.

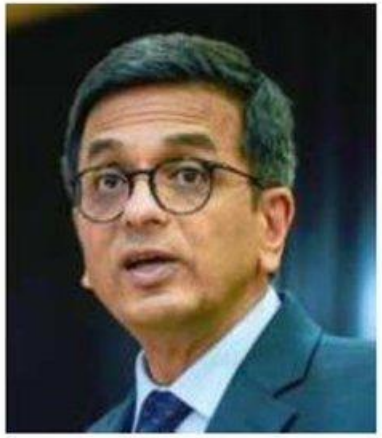
Disinformation can impair democratic discourse: CJI

Krishnadas Rajagopal
NEW DELHI

Chief Justice of India D.Y. Chandrachud on Friday said fake news had drowned out true information simply by virtue of the scale of dissemination in the digital age, replacing the character of discourse from truth-seeking to the loudest voice.

"Disinformation has the power of impairing democratic discourse forever, pushing a marketplace of free ideas to the point of collapse under the immense weight of fake stories," Chief Justice Chandrachud said in his 14th Justice V.M. Tarkunde Memorial Lecture on the topic "Upholding civil liberties in the digital age: Privacy, surveillance and free speech".

He said freedom of speech was traditionally considered an essential part of civil rights activism for the fear that the government would gag certain kinds of speech. However, he said, "With the advent of troll armies and organised disinformation campaigns across social media platforms, the fear is that there is an overwhelming barrage of speech that distorts the truth".


Justice D.Y. Chandrachud

He referred to how social media has ruptured the "traditional state-activist-corporation relationship." Civil rights activists no longer place the corporation within the traditional box of an entity whose power is to be restricted. In fact, to the contrary, they rely on social media corporations such as Facebook, Twitter, and YouTube to expand their freedom of speech and expression, often in opposition to the government.

However, he said there was a flip side to adopting privately owned platforms as the medium for dissent, activism, and expression of free speech. Unlike state actors who are held accountable by the Constitution and the electorate, social media platforms are relatively unregulated. This is another novel challenge, he said.

"Simply by virtue of the scale of dissemination, fake news drowns out true information, replacing the character of discourse from truth-seeking to the loudest voice,"

If we cannot agree on the veracity of basic facts, debate stops, partisanship hardens, and social solidarity breaks down

A cursory glance at newspapers brings to the fore violence resulting out of rumours and disinformation campaigns.

EAM - TECHNOLOGY PENETRATES BOUNDARIES AND ALL BARRIERS

Information Integrity Fundamentals

- Authenticate by alternate means before taking any action- even if told urgency or threat
- Never reveal locations, school or colony names, or contact details online
- Zero Trust ethos- you do not need to expand social network beyond known persons
- Absorb knowledge- do not broadcast news

They're Not Hacking Your Tech, They're Hacking You: The Psychology of Social Engineering Attacks

Attackers leverage cognitive biases, emotional triggers, and social norms to deceive their targets :-

1. **Authority**: People are more likely to comply with requests from figures of authority. Impersonate managers, executives, or IT personnel

2. **Urgency**: Creating a sense of urgency can lead individuals to act quickly without fully considering the consequences.

3. **Reciprocity**: offer something seemingly beneficial, such as a free software download, in exchange for personal information.

4. **Consistency**: small initial agreements, which can lead to larger, more compromising actions

5. **Social Proof**: fake testimonials or references to credible sources to gain trust

What Is Information Integrity?

- Information integrity is the quality of information being accurate, complete, consistent, and unaltered from creation through use.
- It applies to every format: news articles, research data, official records, photos, videos, and everyday messages.
- It is not just about whether something is "true" — it also covers whether it stays intact, unmanipulated, and properly attributed as it travels.
- Strong information integrity practices protect individuals, organizations, and democratic institutions from manipulation and error.

Think of information integrity as a chain: if any link — creation, storage, or sharing — is weak, the whole chain of trust breaks.

The Three Pillars of Information Integrity

Accuracy

Information correctly reflects facts, figures, and reality — free from errors, exaggeration, or omission.

Authenticity

Content genuinely comes from the source it claims to, and has not been faked, altered, or misattributed.

Trust

The confidence people place in a source because it has consistently been accurate and authentic over time.

PILLAR 1

Why Accuracy Matters

What Accuracy Looks Like

- Facts are checked against reliable, primary sources
- Numbers, dates, and quotes are reported correctly
- Context is included, not stripped away
- Corrections are made openly when errors occur

What Happens Without It

- Wrong health or safety decisions
- Financial losses from bad data
- Public confusion during emergencies
- Long-term erosion of trust in institutions

PILLAR 2

Why Authenticity Matters

Signs of Authentic Content

- Clear, verifiable author or organization
- Consistent with the source’s known history and style
- Original files, unedited metadata, matching timestamps
- Traceable back to a primary origin

Red Flags to Watch For

- Anonymous or unverifiable authorship
- Content that "looks off" compared to the source’s usual work
- Edited images/videos with inconsistent lighting or artifacts
- No way to trace the content back to its origin

PILLAR 3

Why Trust Matters

सत्यमेव जयते
Ministry of Education
Government of India

विद्यया ऽ मृतमश्नुते
एन सी ई आर टी
NCERT

CyberPeace

Slow

to build — trust is earned through repeated accuracy and authenticity over time

Fast

to lose — a single major error or deception can undo years of credibility

Shared

in effect — when trust in information ecosystems falls, everyone bears the cost

Takeaway: Trust is the outcome of consistently practicing accuracy and authenticity — not a separate goal you can shortcut to.

WHAT YOU WILL LEARN

Session Roadmap

सत्यमेव जयते
Ministry of Education
Government of India

विद्यया ऽ मृतमश्नुते
एन सी ई आर टी
NCERT

CyberPeace

1. Information Integrity Fundamentals

The principles of accuracy, authenticity, and trust that anchor reliable information.

2. Verifying Sources & Content Authenticity

Practical tools and techniques to check whether content is genuine & trustworthy.

3. Integrity in Digital Communications

Applying these principles to email, messaging, and social media in daily life.

SECTION 2 OF 3

Verifying Sources & Content Authenticity

Practical techniques to check what you read, see, and share

Why Verify Sources?

- Anyone can publish anything online — there is no universal filter for accuracy before content reaches you.
- AI tools now make it easy to generate convincing fake text, images, audio, and video at scale.
- Verification is a skill, not a one-time act — it should become a habit before you believe, react to, or share content.
- Taking a few extra minutes to verify protects you and everyone you might share the content with.

Rule of thumb: If a piece of content triggers a strong emotional reaction, pause and verify before sharing it further.

A PRACTICAL FRAMEWORK

The CRAAP Test for Evaluating Sources

Currency

When was it published or updated? Is it recent enough for the topic?

Relevance

Does it actually relate to your question, and at the right depth and level?

Authority

Who is the author or publisher? What are their credentials or track record?

Accuracy

Is it supported by evidence, and can it be verified elsewhere?

Purpose

Why was it created — to inform, sell, entertain, or persuade?

VERIFICATION TOOL

Reverse Image Search

How It Works

- Upload or paste an image link into a search tool
- The tool scans the web for matching or similar images
- Results show where else — and when — the image has appeared
- Popular tools: Google Images, TinEye, Bing Visual Search

What It Reveals

- Whether an image is old and being reused out of context
- The image's original source and caption
- If the image has been digitally altered from an earlier version
- Whether it truly shows what a post claims it shows

Metadata & EXIF Data

- Most digital photos and files carry hidden metadata: camera or device model, date/time, GPS location, and editing software used.
- Checking metadata (EXIF) can confirm — or contradict — when and where a photo claims to have been taken.
- Free tools like Jeffrey’s Exif Viewer or built-in file "Properties/Details" panels let anyone inspect this information.
- Caution: metadata can be stripped or faked, so treat it as one clue among several, not final proof.

Many social platforms automatically strip metadata on upload, so absence of metadata is not itself suspicious.

EMERGING STANDARD

Digital Watermarking & Content Provenance

- Digital watermarks embed an invisible or visible marker in content to signal its origin or that it was AI-generated.
- Content provenance standards, such as C2PA (Coalition for Content Provenance and Authenticity), attach a secure, tamper-evident history to a file: who created it, what tools were used, and what edits were made.
- Major camera, software, and AI companies are adopting these standards to help audiences tell real content from synthetic content.

Provenance tools do not judge whether content is "good" or "bad" — they simply make its history visible so you can decide.

Red Flags of Unreliable Content

- No named author, publisher, or verifiable source
- Urges you to "share before it’s deleted"
- Extreme emotional language or clickbait headlines
- Claims that seem too dramatic or convenient to be true
- Images/video with odd lighting, blurring, or mismatched shadows
- No other credible outlet is reporting the same story

GROUP ACTIVITY

Spot the Fake

THE SCENARIO

A social media post shows a dramatic photo claiming to be from a "breaking" local emergency, with a caption urging viewers to share immediately and no named source.

KEY FACTS

- A reverse image search shows the same photo published years earlier for a different event in another country
- No official news outlet or authority has reported the claimed event
- The account posting it was created recently and has no verified history

TAKEAWAY: Apply CRAAP, run a reverse image search, and check for independent confirmation before you believe or share breaking claims.

SECTION 3 OF 3

Integrity in Digital Communications

Applying these principles to email, messaging, and social media

EVERYDAY APPLICATION

Email & Messaging Integrity

Sender Spoofing

Attackers fake a "From" address to look like a trusted contact or organization — always check the full email address, not just the display name.

Phishing Signals

Urgent requests, unexpected attachments, and links that don’t match the claimed destination are classic warning signs.

Verify Before You Click

Hover over links to preview the real URL, and verify unusual requests through a separate, known channel.

Secure Channels

Use encrypted messaging apps and enable account verification features where available.

EVERYDAY APPLICATION

Social Media & Integrity

- Context collapse: a post written for one audience can be screenshotted and reshared to a completely different audience, stripped of its original context.
- Edited screenshots, cropped images, and quote-out-of-context posts are common ways integrity breaks down on social platforms.
- Algorithms often reward engagement, not accuracy — sensational or misleading content can spread faster than careful corrections.
- Before resharing, ask: do I know the original context, source, and date of this content?

Sharing is a form of publishing. When you hit "share," you become part of the information's chain of custody.

Digital Signatures & Encryption Basics

- A digital signature is a cryptographic proof that a message or document came from a specific sender and has not been altered since signing.
- Encryption scrambles content so only the intended recipient with the correct key can read it, protecting integrity and confidentiality in transit.
- Look for signed documents (e.g., PDF signature panels) and "end-to-end encrypted" labels in trusted messaging apps as everyday examples.
- These tools don’t replace critical thinking — they confirm origin and prevent tampering, not truthfulness of content itself.

A padlock icon in your browser confirms an encrypted connection to a website — it does not confirm that the website’s content is trustworthy.

Building a Culture of Integrity Online

- Pause before you share — verify first, especially emotional content
- Question content that seems designed to provoke outrage
- Credit original sources and creators accurately
- Use privacy and security settings to protect your own data
- Correct your own mistakes openly and promptly
- Encourage friends and classmates to verify before resharing

Your Role: Being an Information-Integrity Champion

Be a Careful Creator

Whatever you post, write, or record — make sure it is accurate and properly sourced before you publish it.

Be a Careful Verifier

Apply CRAAP, reverse image search, and metadata checks as everyday habits, not just for "big" claims.

Be a Careful Sharer

Treat every share as a small act of publishing — you are responsible for what you help spread.

Be a Peer Educator

Share what you learn today with friends, family, and classmates to multiply its impact.

WRAP-UP

Key Takeaways

1. Information integrity rests on three pillars: accuracy, authenticity, and trust.
2. Data integrity protects information at every stage of its lifecycle — creation, storage, sharing, and disposal.
3. Verification tools like CRAAP, reverse image search, metadata checks, and provenance standards help confirm authenticity.
4. Integrity applies to everyday digital communication — email, messaging, and social media — not just formal publishing.
5. Everyone has a role to play as a careful creator, verifier, sharer, and educator.

Thank You

Building a more accurate, authentic, and trustworthy information ecosystem — together.

LEARN MORE

- CyberPeace Corps — cyberpeace.org — cyber hygiene and digital wellbeing resources
- C2PA Content Credentials — contentcredentials.org — verify content provenance
- Google Fact Check Explorer — for checking existing fact-checks on claims
- Practice the CRAAP test and reverse image search on content you encounter this week

Thank you