सत्यमेव जयते
Ministry of Education
Government of India

विद्यया ऽमृतमश्नुते
एन सी ई आर टी
NCERT

CyberPeace

CYBERPEACE CORPS · DIGITAL WELLBEING SERIES

Information Integrity & Cognitive Security

Disinformation, Misinformation & Malinformation

Understanding the information disorder — how false and harmful content spreads, and how to recognize and resist it.

WHAT YOU WILL LEARN

Session Roadmap

- Understanding the Digital world and information Disorder
- Fake News and Deepfake
- Social Media Manipulation
- Fact-Checking
- Provisions in the Indian Laws

Digital/ Cyberworld today -1

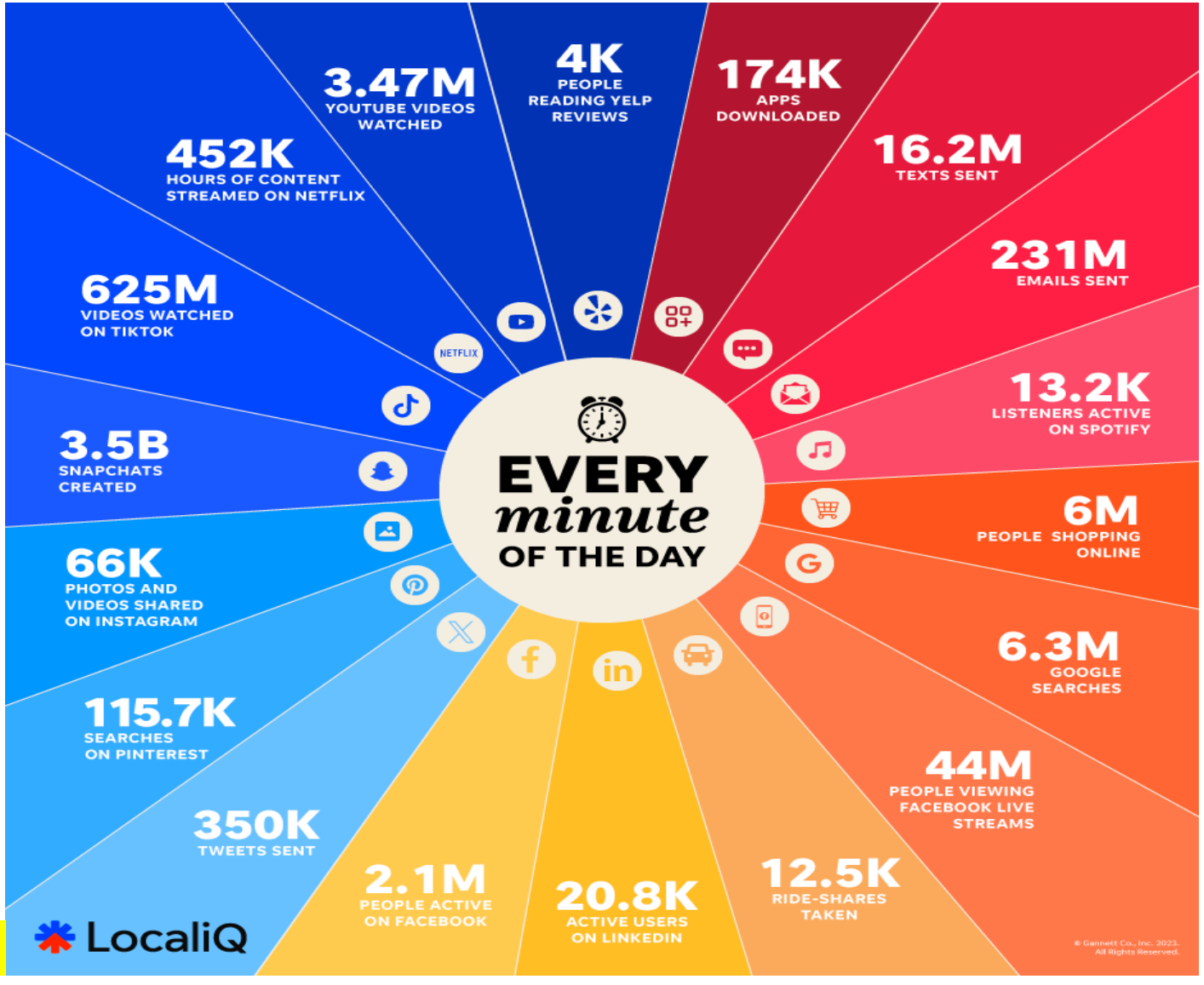
- Cyber space Open, Ubiquitous, virtual, borderless and (mostly) anonymous.
- > 5.5 billion of 8.1 billion people connected on Internet and Social media.
- Connect/ interact with anyone and everyone across the world, instantaneously.
- The Internet a platform of to express ourselves
- Social media and the Digital Media Content
  - Around 5.5 billion Social media Users!
  - Around 200 million active websites
  - Around 3+ million Apps on Android
  - **In India we have around 600 million on WhatsApp, 400 million on FB…..**
- **Any available domain name can be purchased from anywhere.**
- **There are around 20 crore active Websites**
- Virality/ Proliferation and hence the ability to create Online harms is much faster & higher than the physical world.
- AI has changed the whole paradigm by creating abilities to generate, predict, analyse and carry out automated tasks in almost every sector.
- Large number of Apps providing creation, editing features

Digital/ Cyberworld today -2

- The world generates 402.74 million terabytes of data daily, which equals 147 zettabytes per year.

- Vast amount of Information on Internet.
  - **Some of it is also misleading, incorrect, harmful, fake/false, obscene, have potential to breach privacy or is unlawful in some way.**

One Minute on Internet in 2025



Source: LocaliQ

India and the Digital World

- We have about 100 crore Internet connections.
- Increasing Internet penetration, Data Speeds and hours of time spent in digital.
- One of the most used Digital Public Infrastructure (DPI).
- Multi cultural, multi religious, multi language – Huge Variation all across.
- Freedom of speech and expression a fundamental right.

Information perception Vs Reality
(The story of blind men and an elephant)

Understanding the Information Disorder

The vocabulary and framework behind false and harmful content

Defining the Information Disorder

- "Information disorder" is a term researchers use to describe the full landscape of false and harmful information problems online — first widely framed by researchers Claire Wardle and Hossein Derakhshan.
- It moves beyond the single label "fake news" to distinguish between different types of problematic content based on two questions: is it false, and is it intended to cause harm?
- Understanding these distinctions helps you respond appropriately — a genuine mistake needs a correction; a coordinated deception needs a different kind of response.
- The three core categories are misinformation, disinformation, and malinformation.

Not all "false" content is created the same way, and not all "true" content is shared for good reasons. Intent and accuracy are two separate questions.

What Are Deepfakes?

- A deepfake is synthetic media — video, audio, or images — generated or altered using artificial intelligence to convincingly show someone doing or saying something they never did.
- The name blends "deep learning" (the AI technique) with "fake."
- Deepfakes range from harmless entertainment (face-swap filters) to serious threats: fabricated political speeches, fraudulent voice calls impersonating executives, and non-consensual explicit content.
- The technology has become dramatically cheaper and easier to use, moving from research labs to free consumer apps in a few years.

As of recent years, generating a convincing deepfake voice clip can require just seconds of sample audio.

Information Dis-order and its impact

- News and Stories around
  - Operation Sindoor.
  - US/ Israel and Iran War.
  - Recent Students/ CJP Protests.
  - Digitals Scams including “Digital Arrests”.
  - Rumours of Paper leaks
  - Election time Mis-information and Deepfakes
- Impact on
  - Individuals including Students.
  - Society or a particular section of a society. (Public Order)
  - Sovereignty and Integrity of India/ Security of State

Spotting Deepfakes: Red Flags

- Unnatural blinking patterns or lack of blinking
- Unnatural skin texture, lighting, or shadows
- Blurring or warping around the edges of the face or hairline
- Flat or robotic tone in cloned voices
- Audio that doesn’t quite sync with lip movements
- No corroborating source for a dramatic claim

SECTION 4 OF 5

Fact-Checking Techniques & Verification Tools

Practical methods to check claims before you believe or share them

Overcoming the Information Disorder

- The onboarding of fact checkers
  - Utilised by Online news platforms.
  - Utilised by Social media platforms
  - Governments' own fact checking agencies.
- Legislative initiatives by the Government
  - Information Technology (IT) Act
  - The IT Rules, 2021 as amended in 2026.

Independent Fact-Checkers

- Organizations that verify claims and publish transparent, sourced methodology — often signatories of the International Fact-Checking Network.

Media Literacy Organizations

- Groups like CyberPeace Corps provide training, guides, and reporting tools for the public.

The IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (IT Rules, 2021)

Journey of IT Rules, 2021 and its amendments

- Evolution Timeline
  - 2011 Rules → 2021 Overhaul → 2022–23 Tightening → 2026 AI/SGI Regulation
  - The 2026 amendment has come into effect from 20th Feb., 2026.
- Scope
  - Practically all Digital platforms (Intermediaries and Publishers)
- Philosophy of the 2021 Rules
  - Accountability | Traceability | Transparency | User Safety | Sovereign Oversight
- Due Diligence Architecture
  - Policies | Notice | Takedowns | Cooperation | Retention
- Prohibited Content framework
  - Illegal, defamatory, obscene, invasive, misleading, harmful to minors
- Grievance Redressal
  - Grievance officer, acknowledgement, disposal timelines
- Traceability Requirement
  - Identify first originator for serious offences.
- Grievance Appellate Committee (GAC)
  - Govt appointed appellate body (3 committees) for moderation disputes.

Broad areas of 2026 Amendments

Major drivers :

- Address AI-driven misinformation
- Prevent impersonation and deepfake abuse
- Protect victims from non-consensual intimate imagery (NCII)
- Safeguard public trust and digital integrity
- Accelerate enforcement timelines in view of the potential damage it causes because of the Virality.

New Definition

'Synthetically Generated Information' formally defined Rule 2(wa)

Faster Timelines

Govt orders: 36 hrs to 3hrs.
Grievances: 15 days to 7 days.
Action on Personal harm : 24 hrs to 2 hrs

Due Diligence (Rule 3(3))

Mandatory AI content labeling, metadata embedding, tracking

Addl. Due Diligence by SSMI (Rule 4(1A))

SSMIs must require user declarations, verify accuracy, display AI labels on SGI

BNS Replaces IPC

Rule 7 updated: Bharatiya Nyaya Sanhita 2023 replaces Indian Penal Code

Identity Disclosure

Platforms can disclose violating user identity to victims and authorities

Conclusion and Way forward

- Internet is a rich source of information. But all information need not be correct and authentic.
- Lot of Mis-information, Dis-information and mal-information on the Internet.
- Always visit authentic Sites 9e.g. .gov.in or .nic.in for Govt websites).
- Always use apps downloaded from the official play stores.
- Verify the authenticity of the information including news from authentic sources or the fact checking sites.
- Do not forward unless you are sure about the authenticity of information.
- Seeing is not believing!. So be aware.

Questions and Answers

Thank You

Building resilience against the information disorder — one verified claim at a time.

LEARN MORE

- CyberPeace Corps — cyberpeace.org — cyber hygiene and digital wellbeing resources
- Google Fact Check Explorer — search existing fact-checks on a claim
- InVID/WeVerify — browser tools for image and video verification
- Practice the SIFT method and lateral reading on content you encounter this week