Information Resilience & Fact-Checking Guide

Training Manual and a Public Resource

September 2025

ConceptHub ETHIOPIA
Powered by TIKVAH ETHIOPIA
Internews Europe
European Union

Published by:
Concepthub Ethiopia |
Powered by TIkvah Ethiopia

+251 91 313 4524
https://www.concepthub.net

Prepared By:
Kirubel Tesfaye
Fact-checking trainer and Open source investigator

Email: tesfayekirubel218@gmail.com
Phone: +251 965 57 7967

Edited By:
Bereket Gudisa
Chief Operating Officer, Tikvah Ethiopia
Founder, Concept Hub Ethiopia

Email: bereketgudisa21@gmail.com
Phone: +251 913134524

Content Directed By:
Rehobot Ayalew
Media consultant, trainer, and researcher

Email: rehobot.ay@gmail.com
Phone: +251 929 04 8772

Contributer:
Dawit Tadesse

Email: dawittad360@gmail.com
Phone: +251 920 89 6435

Layout designer:
Solomon Shewarega

Email: Solasc22@gmail.com
Phone: +251 912 83 1494

Powered by:
Tikvah Ethiopia (project lead)
Organization Contact:

Email: info@tikvahethiopia.com
Phone: +251913134524
Website: https://tikvahethiopia.com/

TABLE OF CONTENTS

PREFACE

This guidebook reflects the vision of the Voice Up! Information for Peace (VIP) Project to strengthen media literacy, fact-checking, and digital safety in Ethiopia. It serves as a practical framework for training, learning, and resilience-building, equipping individuals and institutions with the skills to verify information, counter disinformation, and promote responsible engagement in the digital age.

We extend our sincere gratitude to Internews for their technical support and to the European Union for their valuable funding and partnership. Their contribution has made it possible to produce this guidebook and ensure its role in advancing resilience against misinformation and safeguarding the integrity of information in Ethiopia.

WHAT IS THE CORE OBJECTIVE?

The objectives of this guide are to:

Equip individuals with essential fact-checking and verification skills to identify, analyze, and counter misinformation, disinformation, and propaganda.

Provide reliable tools and methods for verifying digital content, including text, images, videos, and social media posts.

Strengthen digital safety and security by promoting safe online practices, protecting personal data, and ensuring the secure use of verification tools.

Develop the capacity to monitor and interpret emerging narratives across both online and offline environments.

Promote ethical and responsible fact-checking that fosters public trust and protects individual safety.

Encourage the practical application of fact-checking skills through exercises, reporting, and community engagement, thereby supporting long-term information resilience.

ANTICIPATED OUTCOMES

Verifying local issues:

Learners are required to debunk and verify issues, particularly those that are local and will affect social, economic, and political issues in their communities.

Skill Acquisition:

Most of the participants are expected to understand the basics of core investigation and verification techniques. Most of them are expected to at least use those tools in their reporting.

Knowledge Application:

Learners will produce verified fact-checks with a standard. They are also expected to successfully identify **“red flags”** in misinformation and report them back.

Behavioral Change:

They are also expected to reduce resharing unverified claims among the society and on social media.

01

Digital Literacy

CHAPTER ONE:
Digital Literacy

The objective of the chapter is to equip learners with an understanding of digital literacy, media literacy, and the five pillars of media and information literacy, while introducing media monitoring techniques and the role of social media algorithms, so that they can critically access, evaluate, create, and share information responsibly in the digital environment.

1. Digital Literacy

Digital literacy involves the confident and critical use of a full range of digital technologies for information, communication and basic problem-solving in all aspects of life. It is underpinned by basic skills in ICT: the use of computers to retrieve, assess, store, produce, present and exchange information, and to communicate and participate in collaborative networks via the Internet.

1.1 Media Literacy and the 5 Elements

A medium is one of the means or channels of general communication, information, or entertainment in society, such as newspapers, radio, or television. Whereas Media is simply the plural form of medium, meaning all the different communication channels together.

The main pillars of media and information literacy (MIL)

Media and Information Literacy (MIL) can be understood through five practical pillars: Access, Analyze, Create, Evaluate, and Act.

Access: The ability to locate and retrieve information or media content from a variety of sources digital, print, or interpersonal effectively and efficiently.

Analyze: The capacity to critically examine information or media for credibility, bias, purpose, accuracy, and relevance.

Create: The skill to produce original content responsibly and ethically, using media tools, digital platforms, or traditional formats.

Evaluate: The ability to assess both the content and the impact of information before using or sharing it. This may include considering accuracy, ethics, and social consequences.

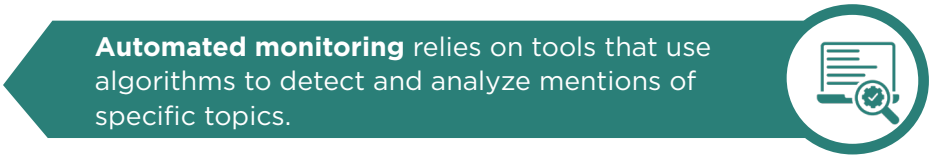
Act: The capacity to apply information and media responsibility on the ground situations. This includes sharing verified content, Liking, commenting, and making informed decisions.

1.2 Media monitoring / Social media algorithm

Media monitoring is essential for tracking, analyzing, and reporting on media material, which includes news stories, broadcasts, social media posts, and other kinds of public conversation.

Manual Monitoring Methods:

Manual Monitoring Methods:



Several free tools are available for media monitoring, including Google Alerts and other advanced search methods.

A social media algorithm is a set of rules and calculations that decides what content appears on your feed, in what order, and from whom.

1.3 How do social media algorithms work?

Social media platforms don't simply display every post in strict chronological order. Instead, they use algorithms to decide what appears in your feed. The main goal: keep you engaged by surfacing content that feels most "relevant" which also keeps you on the platform longer and boosts ad revenue.

The Algorithm Process

Collect Data-The platform tracks your actions: likes, shares, clicks, watch time, follows, and even how long you pause on a post.

Rank Content - Each post gets a score based on how likely you are to engage with it.

Reinforce Behavior -The more you interact with a type of content, the more similar posts you'll keep seeing.

Personalize Feed - The system shows you posts it predicts you'll react to not everything from your friends or sources.

Key Factors Algorithms Consider

- **Engagement** - Posts with more likes, comments, or shares are boosted.
- **Recency** - Newer posts are favored, though not always in strict chronological order (varies by platform).
- **Relationships** - Content similar to what you've interacted with in the past.
- **Format Preference** - If you watch more videos, expect to see more videos.
- **Other Signals** - Things like dwell time, negative signals (hiding/muting posts), and sometimes your location or device also play a role.

Why It Matters

The upside: You get a feed tailored to your interests, making it easier to discover content you enjoy.

The downside:

- ☑ **Limit exposure to diverse perspectives** - you mostly see content that confirms your existing beliefs.
- ☑ **Create echo chambers** - where like-minded opinions get amplified.
- ☑ **Increase polarization** - because people are less exposed to differing viewpoints, making opposing sides seem more extreme.

Summary

This chapter explored the concepts of digital literacy (skills to navigate and use digital tools safely), media literacy (understanding and analyzing media content), and the five pillars of MIL: Access, Analyze, Create, Evaluate, and Act. It also introduced media monitoring, which involves tracking and analyzing information flow across platforms, and explained how social media algorithms shape what users see online. Together, these concepts equip learners to critically evaluate, create, and share information responsibly, fostering informed digital citizenship.

02

Information disorder

CHAPTER TWO :
Information disorder

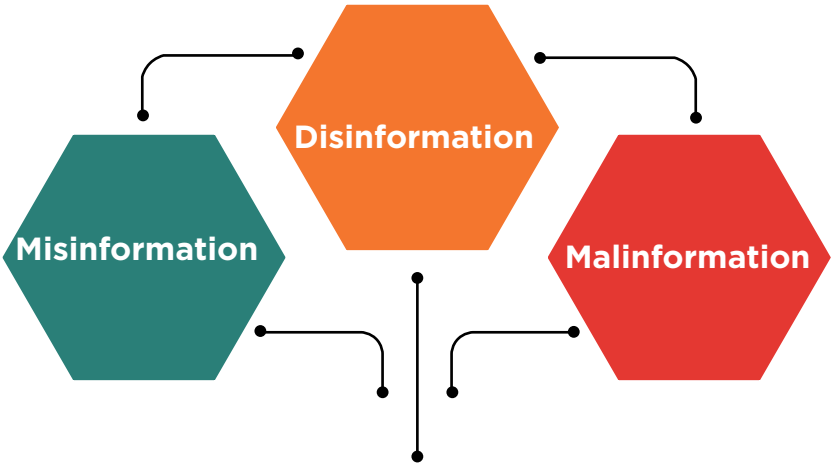
The objective of the chapter is to help learners understand the concept of information disorder, including its three main types, related phenomena such as propaganda, conspiracy theories, and rumors, and the seven forms of disinformation. It also aims to introduce learners to the role of AI in disinformation, and to equip them with practical methods for identifying and reporting false information, fostering critical thinking and responsible digital citizenship.

2. Information disorder

Information disorder is when false, misleading, or confusing information spreads and causes people to believe something that isn't true or misunderstand the facts.

2.1 Information disorder disclosed

The term **"information disorder"** refers to various types of methods used by the media and the internet to spread and magnify false and deceptive information. It includes a variety of phenomena, including disinformation, malinformation, and misinformation.



Misinformation:

Misinformation refers to the dissemination of false information by individuals who genuinely believe it to be true.

Example:

A well-meaning community health worker on a popular Telegram channel shares information he saw online that claims, "Drinking boiled tena adam (rue) water with lemon twice a day prevents COVID-19."

The person believes this traditional remedy is helpful and wants to protect their community. They did not create the graphic intentionally but shared it without checking its scientific validity with health authorities like the Ethiopian Ministry of Health or WHO.

Disinformation:

Disinformation, on the other hand, involves the deliberate spread of false information by individuals who are fully aware of its falsity.

Example:

During a period of ethnic tension, a fake letter, complete with a forged letterhead and signature, circulates on Facebook and Twitter. The letter is purportedly from a prominent political leader from Region X calling for the exclusion of an ethnic group from local administration.

The letter was actually created by a hostile group with the intent to incite violence, destabilize the region, and discredit the political leader.

Malinformation:

Malinformation comprises factually true information but is intentionally shared to cause harm or immediately threaten individuals, organizations, or countries.

Example:

Hana is a popular Ethiopian fashion and beauty influencer on TikTok and Instagram with a large following. She is known for her glamorous and conservative public image.A hacker gains access to her private phone, including intimate personal photos and videos she shared with a former partner.

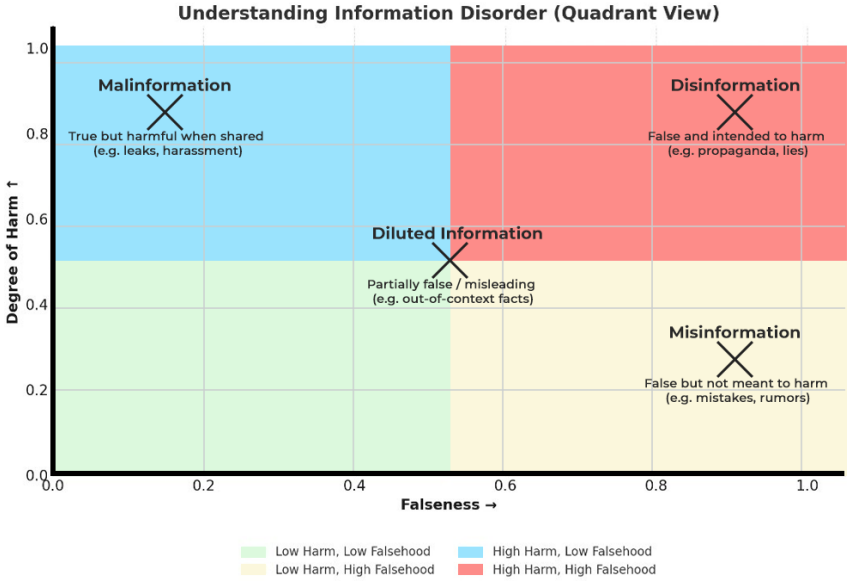
The hacker first attempts blackmail. They contact Hana via Telegram, threatening to leak all the private content to the public unless she pays a large sum of money in cryptocurrency.

Dilinformation:

A genuine information diluted with false information, misinterpreted or misrepresented with or without intent to harm.

Example:

A deceptive social media post mixes genuine photos of a peaceful protest in Shashemene with an unrelated image of a fire from years earlier in Addis Ababa to falsely claim violent suppression, using a kernel of truth to make the entire false narrative seem credible.



concepts:

Propaganda:

can be true, false, or distorted information spread to influence the audience's opinion or behavior, aiming to serve concealed goals that are not always as negative as they are generally perceived.

Conspiracy theories:

attempt to explain the ultimate causes of significant social and political events and circumstances with claims of secret plots by two or more powerful actors.

Rumor:

is unconfirmed information primarily shared in the absence or scarcity of genuine information from the relevant body involved in the incident.

2.2 Common Forms of Disinformation

In the realm of information disorder, there are seven common forms that warrants our attention:

False Connection: occurs when headlines, captions, and audio-visual elements (such as images and videos) do not align with the content they accompany.

Misleading Information: Misleading information involves the manipulation or selective use of facts to shape the perception of an issue or individual.

False Context: False context emerges when genuine content is shared alongside misleading or fabricated contextual information.

Satire/Parody: Satire and parody involve the creation of humorous content without the intention to harm.

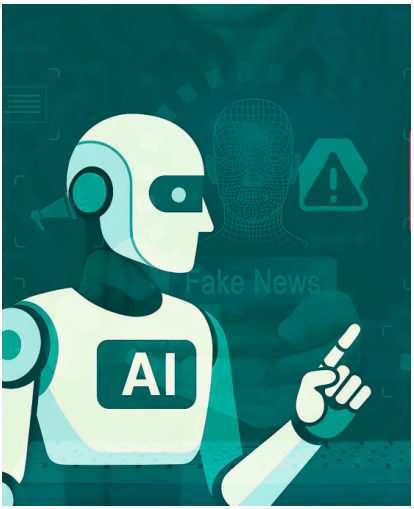
Manipulated Content: Manipulated content is the distortion or alteration of truthful information to deceive the audience.

Imposter Content: Imposter content occurs when genuine information is impersonated or falsely attributed to a source.

Fabricated Content: Fabricated content encompasses false information deliberately created to deceive and cause harm.

2.3 AI and Disinformation

Artificial intelligence (AI) is the development of computer systems that can do tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and problem-solving.



Deep-fake:
Deep fakes (AI-Powered Disinformation) are synthetic media that generate false or altered data using AI techniques, particularly machine learning algorithms.

Cheapfake:
is a type of manipulated content that is created using simple, low-cost techniques rather than advanced artificial intelligence.

2.4 Methods for Identifying False Information

It might be difficult to distinguish between what is factual and what is wrong nowadays. But getting to the truth is always worth it.

To effectively separate fact from fiction, follow these steps:

01. Read Beyond the Headlinen:
Don't rely solely on the headline; delve into the full content.

02 Ask Critical Questions:
Engage in thoroughly examining the information by asking key questions: Who is involved? What are the details? How did it happen? When did it occur?

03 Evaluate the Source: Consider the credibility of the source. Investigate the author or organization behind the information. Assess their expertise, track record, and potential biases.

04. Seek Corroboration: Verify the information by seeking confirmation from multiple reliable sources. Look for consistent reporting across reputable news outlets or statements from trusted experts.

05. Assess Media Attachments: Examine any accompanying media attachments, such as images or videos. They can be manipulated or taken out of context.

06. Consider Reactions and Responses: Examine any accompanying media attachments, such as images or videos. They can be manipulated or taken out of context.

07.Investigate the Author's Background: Research the author's background, qualifications, and previous work. Evaluate their reputation and potential biases.

08. Triangulate Information: Compare and cross-reference the information across different platforms, including reputable news sources, official statements, or academic research.

09 Consult Fact-Checking Sites:
Utilize fact-checking websites known for their rigorous verification processes.

2.5 Reporting disinformation

One key countermeasure in addressing disinformation is the proactive reporting of harmful content to digital platforms. Most social media platforms such as Facebook, YouTube, TikTok, X, Telegram and more, have developed community standards and guidelines that prohibit

Summary

The summary of the chapter is that it explores the meaning and types of information disorder, as well as related concepts like propaganda, conspiracy theories, and rumors. It outlines the seven forms of disinformation, discusses the impact of AI on spreading false information, and presents practical methods for identifying and reporting misinformation. Overall, the chapter equips learners to recognize, analyze, and respond responsibly to disinformation in the digital environment.

03

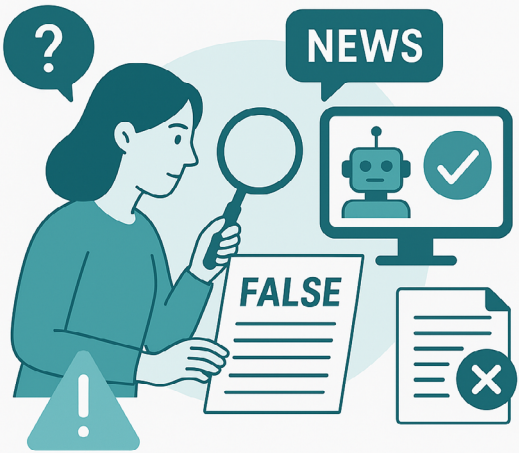
Fact-Checking and its process

CHAPTER THREE:
Fact-Checking and its process

The objective of the chapter is to equip learners with an understanding of fact-checking, its core principles, and the distinction between fact-checkable and non-fact-checkable claims. The chapter also aims to introduce learners to the step-by-step fact-checking process, identify reliable sources of information, and provide practical guidance on verifying AI-generated content and using online verification tools to ensure accurate and responsible information consumption.

3. Fact-Checking and its process

Fact-checking is a crucial process that involves using expert input and authentic external data to verify the accuracy and correctness of the information.



3.1 Core principles of fact-checking

Fact-checking is a subspecial part of journalism as it should be done in a careful way. Some of the core principles of fact-checking in which every fact-checker should follow are:

Accuracy Over Speed: Rushing to debunk can amplify errors, however verifying before sharing is essential even if it means delaying a response.

Transparency: It is always important to disclose sources, methods, and tools used, and acknowledge uncertainty If evidence is inconclusive.

Neutrality & Impartiality: Apply the same rigor to claims across the political/ideological spectrum. Our role is to verify facts, not push agendas.

Fairness & Context: Don't exaggerate or oversimplify the original statement. It is also important to provide full context in a process of explaining why a claim is misleading.

Independence: A fact-checker or fact-checking organization should reject funding/partnerships that compromise its integrity.

Evidence-Based Methodology: We need to build our evidence with official records, peer-reviewed studies, and online available documents with expert inputs. We also need to use the original video/audio in the debunked story.

Accountability: Publish clear, visible retractions if mistakes occur. It is also important to open feedback channels as it allows readers to challenge your conclusions.

3.2 Fact-checkable Vs Not Fact-checkable

Fact-checkable	Not fact-checkable
A claim that can be tested against objective evidence or verified using reliable data.	Opinions
Objective, measurable, evidence-based.	Promises
	Predictions

Table 1: Fact checkable Vs Not Fact-checkable

3.3 Fact-Checking Framework

Verify Framework

Claim Identification → Spot and select a statement and posts worth verifying.

Clarify with Source → Reach out to the person/organization who made the claim.

Gather Proof → Collect documents, data, images, videos, or records related to the claim.

Consult Specialists or Authorities → Seek insights from experts or authoritative institutions.

Analyze & Draft → Compare evidence, write a first draft, and refine the findings.

Review & Publish → Edit carefully and release the fact-check in an accessible format.

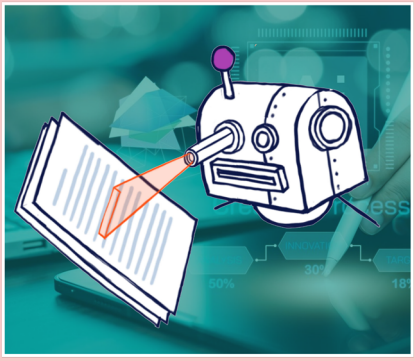
Evaluate Impact → Monitor audience response, corrections, and lessons for improvement.

3.4 Possible sources of information

- Statistical Agency (Central statistical agency)
- Data portals & Databases e.g., Ethiopia Data Portal
- International organizations
- Media and the public
- leading universities & research institutions
- Governments
- professional bodies
- Experts and Fact-checking organizations.

3.5 Verifying AI-generated content

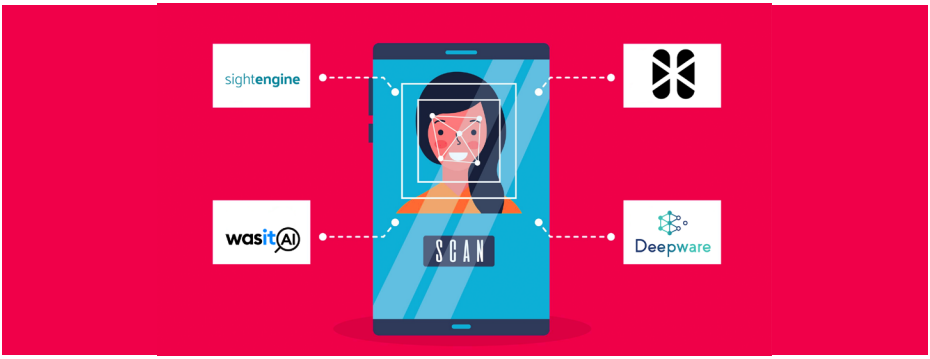
Verifying AI-generated content might be hard, but still, there are still numerous ways that can be used to assess its accuracy and reliability.



- Analyze the source
- Comparison & fact-checking
- Assess biases
- Analyze the supporting evidence
- Look for visual manipulation
- Engage with AI verification tools

AI verification tools

- ☑ Sightengine - image authentication detector
- ☑ AI or not - image authentication detector
- ☑ Wasitai.com - image authentication detector
- ☑ deepware.ai - image authentication detector



3.6 Online verification tools

Online verification tools are digital resources that help individuals check the accuracy, authenticity, and credibility of information encountered on the internet. In today's digital age,

Tool Name	Link	How to use
Google Reverse Image Search	https://images.google.com/	Upload an image or paste the URL to find similar images and track original sources.
TinEye	https://tineye.com/	Perform reverse image searches to detect image manipulation and track origin
Bing Visual Search	https://www.bing.com/visualsearch	An alternative to Google reverse image search with different index coverage.
Yandex Images	http://yandex.com/images	A reverse image search tool that allows users to find images online or by uploading images, with various filters and editing options.

Baidu's image search	https://image.baidu.com/	Baidu Image Search is a platform that allows users to search for images, similar to how Google Images works.
RevEye Reverse Image Search	https://chrome.google.com/webstore/detail/reveye-reverse-image-searc/kagpfnmlfkjnflinklcid-cmdnfhhoebb	Chrome extension that sends images to multiple reverse image tools.
Metadata2Go	http://metadata2go.com/	Drag-and-drop or upload files to view their metadata.
InVID & WeVerify	https://chrome.google.com/webstore/detail/invid-weverify/jpgagcapnkccceppgl-jfpoadahaopjdb	Chrome extension for video analysis: keyframes, metadata, reverse image search, and magnifier.

Table 1: Online verification tools

Why does fact-checking matter?

Fact-checking and debunking false information helps the public to make an informed decision, based on facts.

Summary

The summary of the chapter is that it explores the concept of fact-checking, including its core principles, the difference between fact-checkable and non-fact-checkable claims, and the steps involved in verifying information. It also highlights possible sources of reliable information, the use of AI verification tools for AI-generated content, and various online verification tools. By the end of the chapter, learners are equipped to critically verify information and apply fact-checking methods responsibly in digital and offline environments.

04

Hate Speech

CHAPTER FOUR:
Hate Speech

The objective of the chapter is to provide learners with an understanding of hate speech, its definition, and indicators, and to explain the stages of hate speech escalation. The chapter also aims to equip learners with a red flags checklist, guidance on ethical and responsible reporting, and knowledge of the legal frameworks in Ethiopia related to hate speech, enabling them to identify, report, and prevent harmful content.

4. Hate Speech

The UN Strategy and Plan of Action defines hate speech as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group based on who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor."

4.1 Indicators / Signs of Hate Speech

1. Abusive words/phrases:
Words that reinforce stereotypes, stigmatize marginalized groups, or justify discrimination or violence.

Example:

racial slurs, religious insults, sexist language, disability-related insults, xenophobic expressions, hateful slogans or chants.

2. Dehumanization:
Language that portrays individuals or groups as less than human, dangerous, or evil, reducing empathy and moral consideration.

Example:

calling people “snakes” or “hyenas.”

3.Ambiguous words/phrases:
Words that appear harmless but carry hidden negative meanings targeting specific groups.

Example:

“Junta” in certain contexts.

4. Violence-inciting speech:
Direct encouragement or condoning of violence against individuals or groups.

Example:

“kill,” “attack.”

5. Stereotypical language:
Generalizations or labels that convey negative attitudes toward particular groups.

Example:

calling a group “lazy,” “violent,” or “uncivilized.”

Evolving language/slang:
Subtle or coded expressions that hide hateful intent, making detection challenging.

Example:

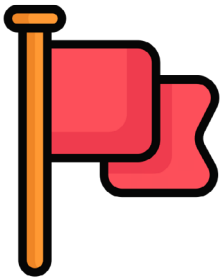
new slang terms used to target groups online.

4.2 Steps to Hate Speech Escalation

The UN Strategy and Plan of Action defines hate speech as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor."

- **Initiation (Trigger / Grievance Formation):**
  Leaders, influencers, or individuals begin to frame others (an out-group) as responsible for problems.
- **Expression (Early Speech / Online or Offline):**
  Hate-laden language starts appearing in conversations, social media posts, songs, or gatherings.
- **Amplification (Spread & Normalization):**
  The hateful messages are repeated and shared widely, often through social media, rumors, or speeches.
- **Targeting (Directed Hate):**
  The hate speech begins to name, shame, or dehumanize specific groups or individuals.
- **Polarization (Us vs. Them):**
  Society becomes divided, and neutral people are pressured to pick a side.
- **Incitement (Call to Action):**
  Hate speech escalates from words to explicit threats or encouragement of violence words such as they should be removed, and attack them can be used.
- **Violation (Real-World Harm):**
  The rhetoric results in offline consequences: harassment, attacks, displacement, or killings.

4.3 Red flags checklist



- ☑ Dehumanizing Language
- ☑ Calls for Violence or Exclusion
- ☑ Conspiratorial Accusations
- ☑ Identity-Based Slurs & Stereotypes
- ☑ Denial of Rights/Dignity
- ☑ Urgent Fear-Mongering, and Historical Revisionism

4.4 What Cases Should a Post Contain to Be Reported as Hate Speech?

◉ **Clear Harmful Content:** Attacks an individual or group based on identity (ethnicity, religion, gender, nationality)

◉ **Explicit Language or Claim:** Must contain actual words, images, or videos that show the violation (not vague or implied). Example: Calling a group “enemies” or spreading a fake video of violence.

◉ **Targeting / Call to Action:** The post identifies or dehumanizes a person, community, or group, or it incites violence, exclusion, or discrimination.

◉ **Public Availability:** The post should be publicly visible (not a private conversation) because platforms moderate public content more strictly.

◉ **Evidence of Harm or Misinformation:** The language must violate community guidelines (slurs, threats, dehumanizing terms).

Ethical Reporting

- Report only harmful content (don't misuse the tool for silencing opinions you disagree with).
- Avoid reposting harmful content yourself, instead, use screenshots with blurred identities if you need to document.
- Each platform has categories (Hate speech or symbols" on Facebook, "Misinformation" on YouTube). Choose the most accurate one.
- Some platforms (like X or Facebook) allow extra notes when reporting, mentioning why it's harmful or misleading.

4.5 Legal framework

When it comes to Ethiopia, in March 2020, the Ethiopian House of People's Representatives adopted the controversial Hate Speech and Disinformation Prevention and Suppression Proclamation No 1185/2020. The law emerged in the backdrop of a string of deadly inter-ethnic clashes across the country, which the government was quick to link to viral speech and disinformation disseminated through broadcasting, social and print media.

Ethiopia's hate speech and disinformation law defines hate speech as a speech that deliberatelypromotes hatred, discrimination, or attack against a person or a discernible group of identity,
based on ethnicity, religion, race, gender, or disability.

4.4 What Cases Should a Post Contain to Be Reported as a Hate Speech?

Ethiopia's Computer Crime Proclamation No. 958/2016, which entered into force on July 7, 2016, established a legal framework to combat cybercrimes, including illegal access, data interference, and the dissemination of defamatory or harmful content online.

The law criminalizes several acts, including illegal access to computer systems or data, illegal interception of non-public computer data, interference with computer systems, causing damage to computer data, and the intentional transmission of programs designed to cause damage.

Summary

The summary of the chapter is that it explores the concept of hate speech, highlighting its definition, common indicators, and the stages through which it escalates. It provides a red flags checklist to identify potential threats, explains what types of content should be reported, and emphasizes ethical reporting practices. Additionally, the chapter outlines Ethiopian legal frameworks governing hate speech, equipping learners to respond responsibly and take informed actions against harmful or illegal content.

05

Digital Safety Essentials

CHAPTER FIVE:
Digital Safety Essentials

The objective of the chapter is to equip learners with essential knowledge and practical skills for digital safety, including securing online accounts through strong passwords and two-factor authentication, understanding the do's and don'ts of account security, and protecting digital communication. The chapter aims to empower learners to safely navigate the digital environment, prevent cyber threats, and safeguard personal and sensitive information.

5. Digital Safety Essentials

Every fact-checking story or news we tell carries weight. Every source we protect holds truth. But in an era of digital surveillance, phishing attacks, and weaponized disinformation, our safety is the first draft of accountability journalism.

5.1 Account Security

5.1.1 Passwords

- Keep your accounts safe by using strong, unique passwords for every login.
- Avoid simple passwords like "123456" or just one word Try using longer phrases like "MyFavoriteFoodIsDoroWot"
- Avoid personal information such as birthdays, phone numbers, your family member's or pet's name...
- Mix characters: use upper case, lower case, symbols, and numbers

➔ Use longer passwords: an average of 12 - 16 characters is recommended for better protection.

➔ Avoid password reuse across multiple platforms.

5.1.2 Two-factor authentication

A Log in to Your Account Settings
Open the app → go to Settings or Security & Privacy.

B Find Security Options
Look for Login Security or Two-Factor Authentication (sometimes called 2-Step Verification).

C Choose a 2FA Method
Most platforms give you 2–3 options: Authenticator App, SMS Code, and Security Key.

D Set Up the Method

- **If using an authenticator app:** Scan the QR code provided by the platform → app generates a 6-digit code every 30 seconds.
- **If using SMS:** Enter your phone number → receive a code → confirm it.
- **If using a security key:** Register it by plugging it in or tapping (for NFC).

E Confirm & Test
The platform will ask you to enter the code from your chosen method to confirm setup.

F Save Backup Codes
Platforms usually give backup recovery codes (one-time use) in case you lose your phone or app access. Store them in a safe offline place (not on the same device you're securing).

G Enable Across All Accounts

Repeat the process for email, social media, messaging apps, and cloud services. Email is especially important since it's often the recovery point for other accounts.

Example for Facebook

- Settings → Security and Login.
- Choose Use two-factor authentication.
- Select method (App, SMS, or Security Key).
- Confirm code.
- Save recovery codes.

Example for Telegram

- Settings → Privacy and Security → Two-Step Verification.
- Set a password (in addition to SMS code).
- Add recovery email.
- Save recovery info.

The Do's and Don'ts in Account Security

Use strong, unique passwords for each account (mix letters, numbers, and symbols).	Don't reuse the same password across multiple accounts.
Enable Two-Factor Authentication (2FA) on all important accounts.	Don't rely only on a single password for protection.
Regularly update your software, apps, and devices to patch security flaws.	Don't ignore update notifications or use outdated apps.
Verify links and email addresses before clicking (phishing awareness).	Don't click on suspicious links or attachments from unknown senders.
Use privacy settings to control what information is public online.	Don't overshare personal details (location, family, workplace) publicly.

Use a password manager to store and generate secure passwords.	Don't write passwords on paper or save them in plain text.

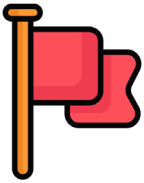
Table: 2

5.2 Communication Protection

When sharing sensitive info (like source details or story drafts), use encrypted apps like Signal or WhatsApp and use Proton Mail for encrypted email conversations. These apps scramble your messages so only you and the receiver can read them. Never discuss private matters over SMS, regular email, or social media direct messages; these are easy to intercept. For files, add a password before sending (zip files with a password sent separately).

Phishing Defense

Phishing scams try to trick you into giving away passwords or downloading viruses through fake emails, texts, or websites.



Red flags include:

- ☑ urgent demands (Verify NOW!),
- ☑ suspicious links ("bit.ly/34ghjKb"), or
- ☑ attachments you didn't expect.

Summary

The summary of the chapter is that it introduces learners to digital safety essentials, focusing on securing online accounts with strong passwords, two-factor authentication, and other protective measures. It highlights the do's and don'ts in account security and offers guidance on protecting communication channels to prevent unauthorized access, phishing, and other cyber threats.

Annex 1: Key Jargon in Fact-checking & Investigation

Social Media & Monitoring Terms

Algorithm	The system that decides what content appears in your feed.
Echo Chamber	Environment where people only see content that reinforces their beliefs.
Filter Bubble	Algorithm-driven isolation from diverse perspectives.
Virality	The rapid spread of content online.
Narrative	The storyline or framing of events being pushed online.
Bot/Troll	Automated or fake accounts spreading certain messages.
Emerging narratives (trends)	are the new storylines or framings true or false that gain traction and shape public conversations.

Fact-checking Process Terms

Rating/Label	Final verdict on a claim (True, False, Misleading, etc.).
Correction	Clarification after an error in reporting.
Transparency	Showing how you verified something (sources, methods).
Contextualization	Adding missing background to avoid misinterpretation.
Attribution	Crediting sources for their role in providing information.

Annex 2: References

First Draft News. Understanding Information Disorder.
Available at: https://firstdraftnews.org/long-form-article/understanding-information-disorder/

UNESCO. Media and Information Literacy.
Available at: https://en.unesco.org/themes/media-and-information-literacy

Digital Marketing Institute. How Do Social Media Algorithms Work?.
Available at: https://digitalmarketinginstitute.com/blog/how-do-social-media-algorithms-work

Simon Fraser University Library. How to spot fake news: Identifying propaganda, satire, and false information.
Available at: https://www.lib.sfu.ca/help/research-assistance/fake-news

Nexis. How to Fact Check Like a Pro (webinar).
Available at: https://www.lexisnexis.com/pdf/nexis/Nexis-webinar-how-to-fact-check-like-a-pro.pdf

United Nations. Understanding hate speech.
Available at: https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech/

Ministry of Justice of Ethiopia. Hate Speech and Disinformation Prevention and Suppression Proclamation No. 1185/2020.
Available at: https://justice.gov.et/en/law/hate-speech-and-disinformation-prevention-and-suppression-proclamation/

FDRE: A Proclamation to provide for the Computer Crime
Available at: https://natlex.ilo.org/dyn/natlex2/natlex2/files/download/103967/ETH103967.pdf

National domestic violence hotline
Available at: https://www.thehotline.org/resources/types-of-abuse/

ConceptHub ETHIOPIA
Powered by TIKVAH ETHIOPIA

Internews Europe

European Union