Data and Information Literacy
(Thinking with Data in the AI)
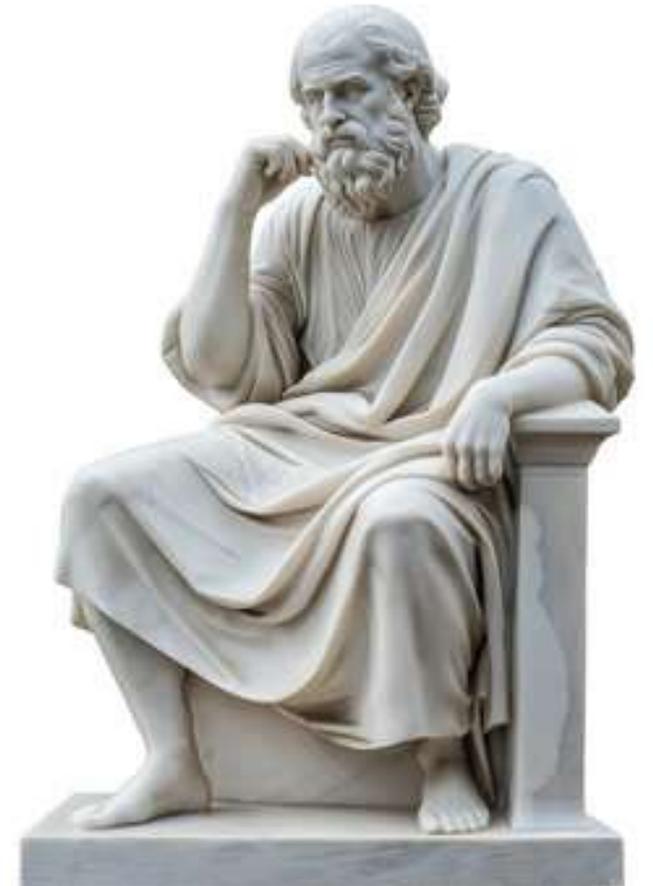
By

Chathurangika Kahandawaarachchi

Program Coordinator, BScHons IT(ISE)

Department of Computer Systems Engineering
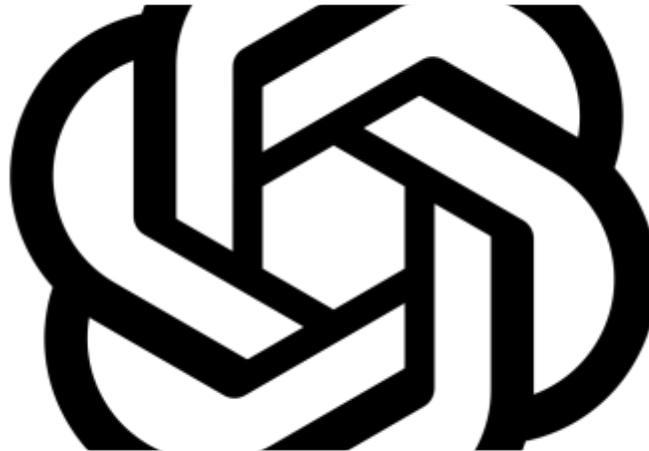
Faculty of Computing

Learning Outcomes

- Thinking with Information in the age of AI.

Think

Do you trust Google results?

Do you trust AI answers?

Do you trust social media news?

**What makes you trust it? What makes you doubt it?**

Data vs Information vs Knowledge

Data	Information	Knowledge
Raw facts or observations	Processed data with meaning	Insights gained from interpreting information
Input for analysis	Supports understanding	Enables decision-making
Low value	Moderate value	High value

Social Media Post Performance Example

Data	Information	Knowledge
• 10,000 likes • 1,200 comments • 3,000 shares • Posted at 9:30 PM	• Engagement increased 25% vs last week • Most engagement came from 18–24 age group	• Short videos posted at night perform best for this audience • Post consistently at 8–10 PM

The Pipeline

Data → Cleaning/Processing → Information

Interpretation → Knowledge → Decision making /Action
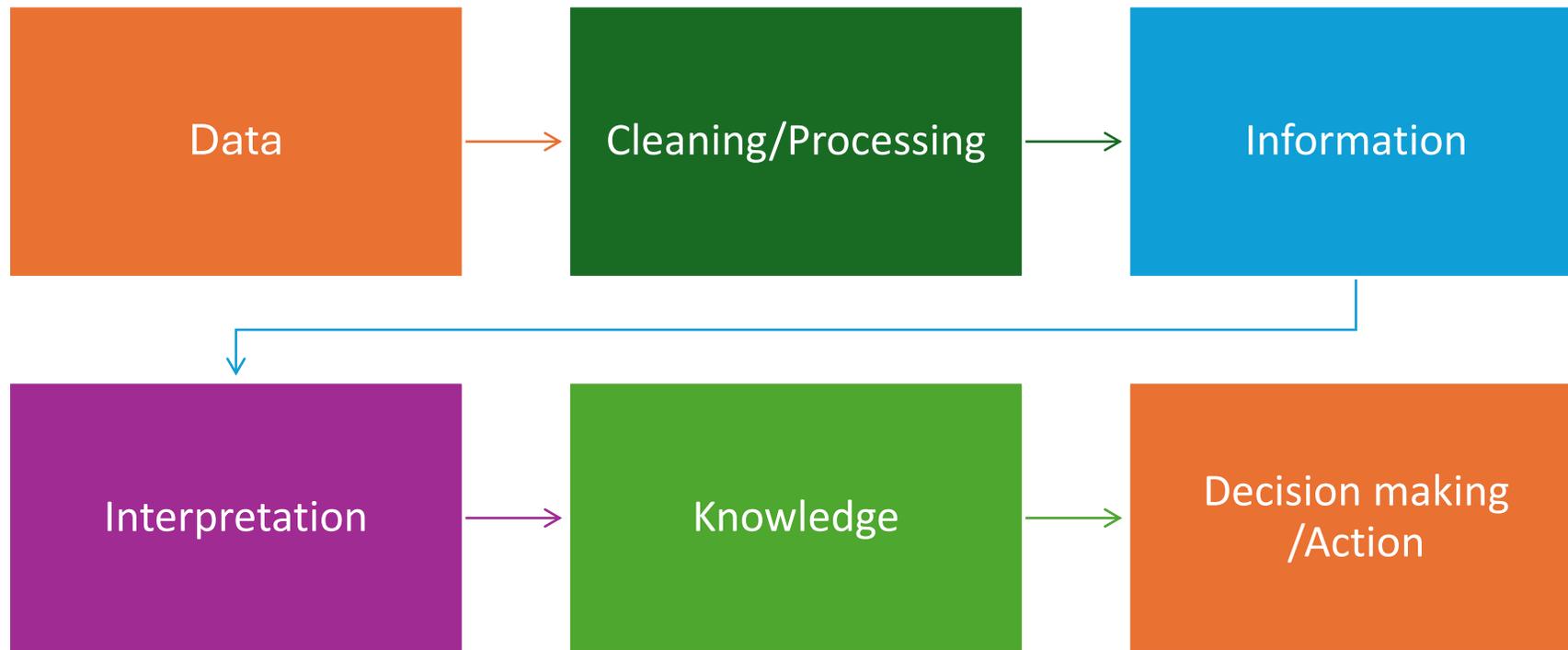
Garbage in, garbage out. Low-quality data leads to bad decisions.

Mini-Activity - Data / Information / Knowledge

- **Step 1:** Choose your example (Attendance, fitness app, gaming stats).
- **Step 2:** Write 1 example each.
  - Data
  - Information
  - Knowledge
- **Step 3:** Share one with the class.

Types of Data

- Know what you're dealing with before you analyze it
  - **By structure**
    - Structured (tables, spreadsheets)
    - Semi-structured (JSON, logs)
    - Unstructured (images, video, text)
  - **By nature**
    - Quantitative (numbers)
    - Qualitative (opinions, descriptions)
  - **By time**
    - Real-time (live feeds)
    - Historical (past records)
  - **By source**
    - Primary (collected directly)
    - Secondary (collected by others)

Examples Students See Every Day

Structured / Quantitative

- Bank transactions

- Grades

- E-commerce orders value

- Weather temperatures

Unstructured / Qualitative

- TikTok videos & comments

- Instagram photos

- Voice notes

- Customer reviews

Data Quality Basics (Why analysis fails)

- **Accuracy:** is it correct?

- **Completeness:** missing values?

- **Consistency:** same format/meaning across sources?

- **Timeliness:** up to date?

- **Validity:** does it follow rules (e.g., valid dates, ranges)?

Credibility & Misinformation

- Not everything online is information.
    - **Misinformation**
        - False or misleading information shared without intent to harm.
    - **Disinformation**
        - False information shared deliberately to mislead.
    - **Malinformation**
        - True information used to cause harm.

Credibility Quick Check

- **Currency:** When was it published/updated?

- **Authority:** Who is behind it? What are their credentials?

- **Evidence:** Sources, data, citations, and transparency

- **Bias/Purpose:** Inform? sell? persuade? provoke?

- **Cross-check:** Can you confirm via two or more independent sources?

Common Misinformation Tactics

- Emotional triggers (fear/anger) to push sharing
- **Cherry-picking** - showing only part of the data
- **False context** - real image/video, wrong story
- **Fake experts** or anonymous claims
- **AI-generated content** (deepfakes, synthetic images, hallucinations)

Data Biases

- Data isn't neutral. Bias can hide in collection, labeling, and algorithms.

- What is Bias?
  - A systematic distortion that makes results unfair, inaccurate, or unrepresentative.
  - Bias can enter before analysis even starts (how data is collected).
  - Algorithms can amplify bias at scale.

Where Bias Shows Up

- Collection
  - Sampling bias (who is included/excluded)
  - Measurement bias (how it's recorded)

- Labeling
  - Human judgments in tagging data (e.g., 'toxic' comments)

- Modeling
  - Algorithmic bias (unequal error rates across groups)

- Feedback loops
  - Recommendations reinforce what you already watch/click

Bias in the Real World

Examples

- Hiring filters that disadvantage certain groups
- Face recognition performing worse on some skin tones
- Credit scoring biases from historical lending patterns
- Predictive policing reinforcing over-policing

Reduce Harm (Practical)

- Use representative samples
- Measure performance across subgroups
- Document datasets and assumptions
- Human review for high-stakes decisions
- Transparency + appeal process

Ethical Use of Information

- Just because you can use data doesn't mean you should.

- Ethical Principles:
  - **Respect -** privacy and consent
  - **Responsibility -** minimize harm; consider consequences
  - **Fairness -** avoid discrimination; test impacts
  - **Accountability -** be transparent; own mistakes
  - **Integrity -** avoid plagiarism; cite sources; be honest about uncertainty

Ethics in Student Life (Real Scenarios)

- **Using AI tools** - disclose usage when required; verify outputs; don't fabricate sources.

- **Sharing content** - don't repost private images/screenshots without consent.

- **Research/projects** - don't scrape personal data without permission; anonymize data where possible.

- **Assignments** - cite properly; paraphrase responsibly; avoid plagiarism.

Ethics Checklist Before You Use Data

1. Do I have permission/consent to use it?

2. Is there sensitive or personal data?

3. Could this harm someone if leaked or misused?

4. Am I being transparent about sources and limitations?

5. Would I be comfortable if this was about me?

RECAP

- Data → information → knowledge

- Different data types need different handling and tools.

- Credibility is a skill. verify before sharing.

- Bias can exist in data and algorithms. measure and mitigate it.

- Ethics, respect privacy, cite sources, and use information responsibly.

Thank you