

Unit III: Literature Search, Analysis & Ethics

Dr. Ayush Kumar Agrawal

SoCS
UPES

Managing References & Citations in Computer Science Research

Warm-Up

Question:

- When writing your project report, how do you keep track of references?
- Do you copy-paste links, maintain a Word file, or use a tool?

Purpose: To realize that systematic reference management saves time.

Why References Matter

- Acknowledge the original source of ideas.
- Avoid plagiarism and academic misconduct.
- Help readers verify and explore further.
- Strengthen the credibility of your own work.
- Provide a scholarly context for your research.

Referencing Styles in CS

- **IEEE:** Numbered citations [1], [2].
- **APA:** Author–Year format (Smith, 2020).
- **ACM:** Author–Year or Numeric, depending on journal.
- **Chicago / Harvard:** More common in social sciences.
- In CS, IEEE and ACM are most widely used.

Example (IEEE): “Deep learning has shown progress in NLP [5].”

Tools for Reference Management

- **Mendeley:** Free, easy to use, PDF annotation.
- **Zotero:** Browser-based, integrates with Word/LibreOffice.
- **EndNote:** Powerful, widely used in academia.
- **BibTeX (LaTeX):** Essential for thesis and technical writing.
- **RefWorks / Citavi:** Institutional subscription tools.

Demo (Instructor-Led)

- Show how to add a paper to Mendeley/Zotero.
- Demonstrate automatic citation insertion in Word/LaTeX.
- Export bibliography in IEEE/APA style instantly.

Students realize manual referencing is slow and error-prone.

Activity: Reference Formatting

Task:

- Instructor provides details of one paper: Smith, J. (2021). “AI for Cybersecurity.” *Journal of Computer Science*, 45(3), 120–135.
- Students write it in:
 - IEEE style
 - APA style

Compare answers and discuss differences.

Citations in Research Writing

- **Direct Citation:** “According to Smith (2021). . .”
- **Indirect Citation:** Summarizing in your own words.
- **Multiple Citations:** (Smith, 2020; Gupta, 2021).
- **Self-Citation:** Acceptable but should not be excessive.

Citation Ethics

- Avoid “citation padding” (adding irrelevant sources).
- Don't rely only on one author/group.
- Cite both classical and recent works.
- Always cite the original source, not just secondary ones.
- Acknowledge others fairly to maintain academic integrity.

Quick Quiz

Which of these is **plagiarism**?

- ① Citing a paper and paraphrasing in your own words.
- ② Copying text from a paper without citation.
- ③ Quoting a sentence with citation and quotation marks.
- ④ Summarizing multiple papers with proper references.

Quick Quiz

Which of these is **plagiarism**?

- ① Citing a paper and paraphrasing in your own words.
- ② Copying text from a paper without citation.
- ③ Quoting a sentence with citation and quotation marks.
- ④ Summarizing multiple papers with proper references.

Answer: Option 2.

Common Mistakes in References

- Mixing citation styles in one paper.
- Missing references in the bibliography.
- Listing sources not actually cited in the text.
- Incorrect formatting (author order, year, page numbers).
- Over-citing trivial facts.

Best Practices

- Choose the required style (IEEE/APA/ACM) before writing.
- Use a reference management tool.
- Keep notes while reading papers.
- Regularly update your reference library.
- Double-check formatting before submission.

Summary

- References = foundation of academic credibility.
- Styles: IEEE, APA, ACM → CS mainly uses IEEE/ACM.
- Tools: Mendeley, Zotero, EndNote, BibTeX.
- Citations must be accurate, ethical, and consistent.
- Good reference management saves time and prevents plagiarism.

Plagiarism & Paraphrasing in Computer Science Research

Warm-Up

Question:

- If you copy one paragraph from a paper into your thesis without citation, is it plagiarism?
- What if you change 2–3 words but keep the same structure?
- What if you cite properly?

Purpose: To highlight misconceptions about plagiarism.

What is Plagiarism?

- Using someone else's ideas, words, or work without proper acknowledgment.
- Can be intentional or accidental.
- Considered academic misconduct and can lead to rejection, penalties, or blacklisting.
- Applies to text, figures, code, data, and even ideas.

Types of Plagiarism

- **Direct/Verbatim:** Copy-pasting text word-for-word.
- **Mosaic:** Mixing copied phrases with own words.
- **Self-Plagiarism:** Reusing your own previous work without citation.
- **Paraphrasing Plagiarism:** Changing words but keeping same structure.
- **Accidental:** Forgetting to cite or improper citation.

Examples of Plagiarism in CS

- Copying code snippets from GitHub without credit.
- Using diagrams from another paper without citation.
- Submitting the same project report in two different courses.
- Publishing a conference paper, then resubmitting it as a new journal paper.

Tools for Plagiarism Detection

- **Turnitin:** Widely used in universities.
- **iThenticate:** For journal submissions.
- **PlagScan, Grammarly Premium:** Popular alternatives.
- **Code Similarity Checkers:** Moss, JPlag for source code.

Journals usually reject papers with similarity index above 15–20%.

What is Paraphrasing?

- Restating someone else's idea in your own words.
- Should change both words and structure.
- Must still include citation.
- Shows understanding of the original work instead of blind copying.

Good vs. Poor Paraphrasing

Original: “Machine learning models require large labeled datasets for training.”

Poor Paraphrasing: “Machine learning systems need big labeled datasets for training.” (Plagiarism – structure unchanged).

Good Paraphrasing: “To achieve effective learning, most ML approaches rely on extensive annotated datasets [?].” (New structure + citation).

Activity: Paraphrase This

Task: Original text: “Cloud computing improves scalability but raises security concerns.”

- Students write a paraphrased version in their own words.
- Share 2–3 answers with the class.
- Compare which versions are acceptable.

Academic Writing Ethics

- Always cite original sources, even if paraphrased.
- Use quotation marks for direct quotes.
- Do not “patchwork” multiple sentences from different papers.
- Acknowledge collaborators and prior work honestly.
- Remember: Originality = academic integrity.

Quick Quiz

Which of these is acceptable?

- ① Copying two lines from a paper with citation, no quotation marks.
- ② Paraphrasing an idea with citation.
- ③ Reusing your own published paper without reference.
- ④ Copy-pasting a figure without credit.

Quick Quiz

Which of these is acceptable?

- ① Copying two lines from a paper with citation, no quotation marks.
- ② Paraphrasing an idea with citation.
- ③ Reusing your own published paper without reference.
- ④ Copy-pasting a figure without credit.

Answer: Option 2.

Common Mistakes to Avoid

- Thinking citation alone = no plagiarism (bad paraphrasing still counts).
- Using synonyms but keeping the same sentence structure.
- Forgetting to cite figures, datasets, or code.
- Assuming self-plagiarism is not plagiarism.
- Submitting assignments with “copy-paste” from Wikipedia.

Summary

- Plagiarism = using others' work without credit (text, figures, code, data).
- Types: direct, mosaic, self-plagiarism, paraphrasing, accidental.
- Detection tools: Turnitin, iThenticate, Moss (code).
- Paraphrasing = reword + restructure + cite.
- Academic integrity requires honesty and originality.

Research Ethics in Computer Science

Warm-Up

Question:

- If a researcher manipulates results to make their algorithm look better, is it ethical?
- What if a student submits the same project to two different courses?
- What if a company collects user data without consent?

Purpose: To realize that ethics applies to people, data, and results.

What is Research Ethics?

- Principles guiding responsible conduct in research.
- Ensures honesty, transparency, and fairness.
- Protects research subjects, communities, and the scientific record.
- In CS: extends to software, data, algorithms, and human interaction.

Core Ethical Principles

- **Honesty:** No fabrication, falsification, or misrepresentation.
- **Objectivity:** Avoid bias in design, data, and interpretation.
- **Integrity:** Follow professional and institutional guidelines.
- **Respect:** Protect human participants and their data.
- **Accountability:** Be responsible for one's work and contributions.

Ethical Issues in CS Research

- **Human Subjects:** Experiments involving users must ensure consent and privacy.
- **Data Handling:** Secure storage, anonymization, and legal compliance.
- **Authorship:** Credit only those who contributed significantly.
- **Publication Ethics:** No duplicate submissions or fake peer reviews.
- **Bias in AI:** Algorithms must avoid discrimination.

Case Study 1: Facebook Emotion Study (2014)

- Facebook altered news feeds of 700,000 users without consent.
- Goal: Study effect of positive/negative posts on emotions.
- Outcome: Massive criticism for violating informed consent.
- Lesson: Large-scale CS experiments also need ethical approval.

Case Study 2: Google AI Ethics Controversy

- In 2020, Google faced backlash after firing researchers who published work on risks of large AI models.
- Raised questions about corporate influence on academic freedom.
- Lesson: Research must remain independent and transparent.

Activity: Spot the Ethical Breach

Scenario: A PhD student trains an AI model using medical records but does not anonymize patient names.

- Is this ethical?
- What principles are being violated?
- How should it be corrected?

Ethics in Data Usage

- Always obtain informed consent where human data is involved.
- Follow data protection laws (GDPR, HIPAA, Indian IT Act).
- Ensure anonymization of personal identifiers.
- Avoid using datasets with unclear or illegal provenance.
- Provide transparency on how data is collected and used.

Authorship Ethics

- Only those who made significant contributions qualify as authors.
- Supervisors should not demand authorship without involvement.
- Acknowledge collaborators, funding agencies, and tools.
- Ghost authorship or gift authorship are unethical.
- Follow journal/conference guidelines strictly.

Quick Quiz

Which of these is an ethical practice?

- ① Submitting the same paper to two conferences.
- ② Adding a friend's name as co-author without contribution.
- ③ Collecting survey data after obtaining informed consent.
- ④ Modifying results to make an algorithm look better.

Quick Quiz

Which of these is an ethical practice?

- ① Submitting the same paper to two conferences.
- ② Adding a friend's name as co-author without contribution.
- ③ Collecting survey data after obtaining informed consent.
- ④ Modifying results to make an algorithm look better.

Answer: Option 3.

Common Ethical Violations

- Plagiarism or self-plagiarism.
- Data fabrication or falsification.
- Duplicate publication.
- Ignoring informed consent/privacy.
- Unjustified or fake authorship.

Summary

- Research ethics ensure honesty, fairness, and accountability.
- In CS: covers data, human subjects, algorithms, and authorship.
- Case studies (Facebook, Google AI) show real consequences of violations.
- Good ethics = trust, credibility, and sustainable research.

IRB, Data Privacy Norms & Ethical Research Workshops

Warm-Up

Question:

- Before launching a survey on mental health apps, should a CS researcher need permission?
- Who checks if the research is safe for participants?

Purpose: To introduce IRB and privacy norms in CS research.

What is IRB?

- IRB = Institutional Review Board (also called Ethics Committee).
- Approves, monitors, and reviews research involving humans.
- Ensures research meets ethical standards.
- Protects participants' rights, safety, and privacy.

When Do You Need IRB Approval?

- Surveys, interviews, or experiments with human participants.
- Studies involving sensitive datasets (health, finance, biometrics).
- User studies in HCI, software usability, or gaming research.
- Any project where identifiable data is collected.

Key Principles of IRB

- **Informed Consent:** Participants must know risks/benefits.
- **Confidentiality:** Protect personal information.
- **Minimal Risk:** Avoid harm or discomfort.
- **Right to Withdraw:** Participants can leave anytime.
- **Fair Selection:** No bias in participant recruitment.

Data Privacy Norms in CS

- **GDPR (EU):** General Data Protection Regulation.
- **HIPAA (US):** Health data protection.
- **Indian IT Act, 2000 & DPDP Act, 2023.**
- Key Principles:
 - Collect only necessary data.
 - Ensure storage security.
 - Anonymize or pseudonymize data.
 - Share only with consent.

Case Study 1: Cambridge Analytica Scandal

- Data from 87 million Facebook users harvested without consent.
- Used for political profiling and manipulation.
- Violated trust, privacy, and ethical standards.
- Led to major reforms in data protection laws.

Case Study 2: Strava Heatmap Leak

- Fitness app released global heatmaps of running routes.
- Accidentally revealed secret military base locations.
- Lesson: Even anonymized data can expose sensitive information.

Activity: Privacy Check

Task:

- You are designing a mobile app to track student study habits.
- What data should you collect?
- How will you anonymize it?
- How will you get informed consent?

Workshops on Research Ethics

- Universities often hold workshops on:
 - Ethical data collection.
 - Responsible authorship.
 - Handling plagiarism.
 - Using IRB portals for project approval.
- Students learn by practice → preparing consent forms, anonymizing data, writing ethical statements.

Quick Quiz

Which of these needs IRB approval?

- ① Analyzing open-source GitHub repositories.
- ② Conducting a survey on stress among university students.
- ③ Building a compiler optimization algorithm.
- ④ Studying blockchain transaction logs (public data).

Quick Quiz

Which of these needs IRB approval?

- ① Analyzing open-source GitHub repositories.
- ② Conducting a survey on stress among university students.
- ③ Building a compiler optimization algorithm.
- ④ Studying blockchain transaction logs (public data).

Answer: Option 2.

Best Practices for Ethical CS Research

- Always check if IRB approval is needed before starting.
- Keep participant data anonymous and secure.
- Be transparent about data use and sharing.
- Document all ethical safeguards in publications.
- Participate in ethics workshops and training regularly.

Summary

- IRB ensures human-centered research is ethical and safe.
- Data privacy norms (GDPR, HIPAA, Indian DPDP) guide data handling.
- Case studies (Cambridge Analytica, Strava) show real-world risks.
- Ethics workshops prepare students for responsible research.
- Ethical compliance builds trust and credibility in CS research.