Academic Plagiarism, Ethics, Reproducibility & Related Topics...

Outline for Today

- Eight "What If" Ethics Scenarios
  - Plagiarism in Courses / Education
  - Plagiarism in Research / Publications
  - Data Reporting & Reproducibility
  - Software License / Intellectual Property
- Formal Policies / References
  - NSF Reviewer Conflict of Interest
  - ACM/IEEE Ethics for Engineers & Algorithm Transparency
  - Ethics for Journalism / Data Visualization
  - IRB / FERPA / HIPAA
  - Accessibility
  - Inclusivity

Outline for Today

- Eight "What If" Ethics Scenarios
  - Plagiarism in Courses / Education
  - Plagiarism in Research / Publications
  - Data Reporting & Reproducibility
  - Software License / Intellectual Property
- Formal Policies / References
  - NSF Reviewer Conflict of Interest
  - ACM/IEEE Ethics for Engineers & Algorithm Transparency
  - Ethics for Journalism / Data Visualization
  - IRB / FERPA / HIPAA
  - Accessibility
  - Inclusivity

[1/8]

Sally has been swamped with TA work. She didn't have time to do her homework for a required theory class. Note: Her research focus is not theory!

A classmate sends her a website with solutions to all problems from the course textbook. Everyone has the link. She copy-pastes from the solutions, carefully fixing a few errors she finds in the math, and submits the assignment.

- Does the syllabus permit referencing online solutions?
- Does the syllabus permit peer collaboration or teamwork on the homework?
- What if the assignments are only for studying and 100% of her grade will be based on test scores?
- Who does this hurt?

Read the syllabus. It's probably NOT ok! She's hurting herself, and her classmates.

- What should you do as a TA if you notice probable plagiarism or academic dishonesty by your students?
- What are the appropriate consequences?
- Does it matter if it's a first offense or a repeat offense?
- Does it matter if it's an undergraduate student or a graduate student?

Report this to the course instructor.

Tony is writing a paper after months of work developing a complex networked parallel simulation. His results are solid -- it's a clear improvement over the work of Smith et al.

However, he's having trouble writing the paper introduction. He's always hated writing.

He re-reads the introduction to Smith et al. and likes how they phrased the problem statement. He pastes a couple paragraphs from their introduction into his Latex document. He plans to rephrase those sentences before submission...

- How many words do you need to change in a sentence in order to make it not plagiarism?
- Is it enough to cite the original paper in the bibliography?
- Is it enough to put a reference [Smith et al. 2015] at the end of the mostly borrowed paragraph?
- Can you put the entire paragraph as a block quote?

These are all bad ideas.

- Is it ok if the original author is:
  - one of your advisor's close friends?
  - one of your advisor's former students?
- Is it ok if the original paper is in another language and you translated it?

These are all bad ideas.

- Do you have to write an introduction? Can you just put "See [Smith 2015]" and jump straight to technical stuff?

You need to describe the problem in your own words.

Justine is writing a review article for her research qualifier. She's learning state-of-the-art research that will be the foundation for her thesis. And her advisor introduced her to the star researchers in the field at recent conferences. They plan to submit the review article for publication later this year.

She's hoping to include diagrams and data tables from the original articles. Her fellow grad student says it's fine to reprint the original material in a review article.

- Is it ok if you:
    - cite the original author?
    - have permission from the author to reuse their figures/tables?
    - have permission from the publisher of the book/journal/proceedings?
- Should you get the author's/publisher's permission in writing? Getting explicit permission in writing is best.
- Is it ok to redraw/recreate the figure? Or run your own simulation to produce and plot a similar dataset?
  This is an acceptable solution. It may be the best plan.
- Should you trust your officemate, a senior graduate student, who says it's "fine"?
- Should you trust your advisor who says it's "fine"?
  You must be responsible for your own academic ethics and decisions.

[4/8]

Felix is presenting the selected paper at his research group's reading seminar next week. It's on a topic just outside of his current research, but someone suggested it might be a direction he should explore and consider for his future work.

He's found the slides online from the original author's conference presentation. Can he use those slides for his presentation to the group?

- Does it matter if he does or does not modify the slides?

  It's generally ok, and encouraged, to re-use academic materials for educational purposes. Researchers put materials online to disseminate their work!

- Does it matter if the seminar materials are archived on a private Dropbox for students who miss the meeting?
- What if the seminar materials are published on a public website on a university server?

  Re-publishing materials is generally not ok though. Be careful.

- Should he contact the authors for permission?

  Generally not necessary for classroom or private seminar use. Make sure you cite the author and publication/source of the materials.

  If you're presenting as part of a conference tutorial or other public forum, please do ask for author permission.

[5/8]

Bernadette has written a half dozen conference/journal papers with her advisor and a couple other grad students. She's first author on most of the papers and can claim one or more technical contributions from each of the papers.

She's ready to write and defend her PhD thesis. Her advisor has approved her thesis outline -- four key chapters of contributions that will draw heavily from these papers.

Can she start writing by copy-pasting from the Latex source of these papers into these thesis chapters?

- Who wrote the text for the original papers?
  - What if Bernadette wrote the first draft, her advisor gave limited edits/feedback, and Bernadette did the revisions? `Yes, this is ok. (see below)`
  - What if it was a group effort where everyone contributed paragraphs that were shuffled and edited to fit into the page limit? What if no one remembers who wrote any specific sentence? `Should rewrite.`
  - What if Bernadette wasn't involved in writing one of these papers, but her algorithm was used (and saved the day), so she was thanked with a co-author spot? `Should definitely rewrite.`
- What are RPI's rules for citing your own work in your thesis? `Specific format for footnote citation at start of every chapter.`

[6/8]

Chris is frustrated with his research, but is pressured by his advisor to prepare a paper for an upcoming conference. He's run his simulation against the standard set of 10 benchmarks.

One benchmark does not compile (he doesn't have time to debug it right now). Another benchmark runs 2X slower under his system. But the remaining benchmarks show a 5-10% performance improvement with his system. However, he later finds a bug in his implementation that will negatively impact a few of his results. And unfortunately, he still uses a hard-coded mystery constant in the code he cannot explain.

How should he write up his results?

- How should you handle:
  - Missing data? (e.g. did not compile) No, this is data fabrication!
    Is it ok to make an educated guess for the value?
  - Erroneous data? (e.g., flaw in testing environment)
    Is it ok to publish the collected data without documentation
    of the collection flaws? No, this is data falsification!
  - Bad data? (e.g. outliers that don't support your hypothesis)
    Is it ok to simply omit them? No, this is data falsification!

- ACM SIG Review form: "Can the work be reproduced from
  the information in the paper? Are all important algorithmic
  or system details discussed adequately? Are the
  limitations and drawbacks of the work clear?"

- *Your RA funding source may require you to complete CITI
  training on the responsible conduct of research.*

[7/8]

Keith just returned from a machine learning summer
internship at Softbookzon. He's motivated to finish his
degree, get a respectable salary, and use his skills to make
positive contributions to both industry and society!

Over the summer he wrote software to target ads to users
based on private and proprietary data. He sometimes worked
on his personal laptop (no one explicitly said he couldn't) and
he never deleted the data.

That data would be helpful right now as he develops a
machine learning algorithm for his research (unrelated to
Softbookzon or advertising). Can he use this data?

- What if he only uses the data for development and debugging? No, this is not ok.
- What if he only publishes performance scaling results based on the data? Of course he'd never publish any details of the dataset! No, this is not ok.

- What if the data includes personal medical records?
- What if the data includes educational records?
- What if the data includes financial information?

  NO! This is NOT OK! See HIPPA! And FERPA!

- What if he puts his manager as a co-author on the paper?

  No, this is not relevant or ok!

- *You should always get approval before collecting or using private user data in your research. See your Institutional Review Board (IRB) for application information.*

[8/8]

Allison is polishing her resume by getting a masters in computer science before she and her roommate (an MBA) plan to launch a software-hardware hybrid security startup. Allison will be the tech lead for the project, initially writing code, but eventually transitioning that work to a team of new developers they plan to hire.

If Allison studies the source code from the Linux kernel in one of her classes, will this cause any problems?

What if Allison's research group attends a presentation by a company working on a related topic and the company asks all attendees to sign a non-disclosure agreement?

- What is open-source vs. copyright vs. copyleft?
- Can a closed-source software product produced and sold by a for-profit company include, use, distribute, and/or be based on open-source software?

  There are many different open source licenses!

- Is a developer tainted if they have:
  - done similar work at another private company? What's a non-compete clause?
  - seen similar source code in open-source or proprietary software?
  - signed a non-disclosure agreement in order to attend a company presentation?

    Always fully read paperwork before signing. Consult a lawyer.

- Can a tainted person serve as project manager to a group of untainted developers?

  Hmm.. Consult a lawyer?

Outline for Today

- Eight "What If" Ethics Scenarios
  - Plagiarism in Courses / Education
  - Plagiarism in Research / Publications
  - Data Reporting & Reproducibility
  - Software License / Intellectual Property
- Formal Policies / References
  - NSF Reviewer Conflict of Interest
  - ACM/IEEE Ethics for Engineers & Algorithm Transparency
  - Ethics for Journalism / Data Visualization
  - IRB / FERPA / HIPAA
  - Accessibility
  - Inclusivity

National Science Foundation (NSF) Reviewer Conflict of Interest (COI)

https://www.nsf.gov/pubs/forms/nsf1230p.pdf

- Family (spouse, child, sibling, parent)
- Business or Professional partnership
- Thesis advisor or thesis student (forever)
- Collaboration on book/paper/report
  in last 48 months
- Co-editor of journal in last 24 months
- Close personal friendship

Similar rules for conference / journal reviewing

ACM/IEEE Code of Ethics for Software Engineering

https://ethics.acm.org/code-of-ethics/software-engineering-code/

In accordance with their commitment to the health, safety and welfare of the public, software engineers shall adhere to the following Eight Principles:

1. PUBLIC — Software engineers shall act consistently with the public interest.
2. CLIENT AND EMPLOYER — Software engineers shall act in a manner that is in the best interests of their client and employer consistent with the public interest.
3. PRODUCT — Software engineers shall ensure that their products and related modifications meet the highest professional standards possible.
4. JUDGMENT — Software engineers shall maintain integrity and independence in their professional judgment.
5. MANAGEMENT — Software engineering managers and leaders shall subscribe to and promote an ethical approach to the management of software development and maintenance.
6. PROFESSION — Software engineers shall advance the integrity and reputation of the profession consistent with the public interest.
7. COLLEAGUES — Software engineers shall be fair to and supportive of their colleagues.
8. SELF — Software engineers shall participate in lifelong learning regarding the practice of their profession and shall promote an ethical approach to the practice of the profession.

ACM Statement on Algorithmic Transparency and Accountability

https://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf

Computer algorithms are widely employed throughout our economy and society to make decisions that have far-reaching impacts, including their applications for education, access to credit, healthcare, and employment. The ubiquity of algorithms in our everyday lives is an important reason to focus on addressing challenges associated with the design and technical aspects of algorithms and preventing bias from the onset.

1. Awareness
2. Access and redress
3. Accountability
4. Explanation
5. Data Provenance
6. Auditability
7. Validation and Testing

Society of Professional Journalists' Code of Ethics

https://www.spj.org/ethicscode.asp

- Fact check, cite sources, disclose conflicts of interest
- Sources: question motives, anonymous if appropriate, avoid reimbursement for news
- Avoid stereotypes, misrepresentation, distortion, re-enactments, surreptitious/undercover work, pressure from advertisers & special interests
- Be compassionate, sensitive, invite dialog, admit mistakes

Examples of Questionable Photographic Editing:
http://people.csail.mit.edu/fredo/tmp/FredoDurand_EthicsComputationalPhotography-small.pdf

Visual.ly's Code of Ethics for Data Visualization Professionals

- Data analysis is important

- Clearly state assumptions

- Too narrow focus or omission can lead to bias

- Incorrect analysis must be avoided

- Be open to criticism, learn from past work

- Be aware of colorblindness, cultural meaning

- Be open to criticism

A Hippocratic Oath

Jason Moore also suggested a hippocratic oath for visualization. This version is slightly edited from his original, and I guess some more work could be done on it. But I think it's a great start.

> " I shall not use visualization to intentionally hide or confuse the truth which it is intended to portray. I will respect the great power visualization has in garnering wisdom and misleading the uninformed. I accept this responsibility willfully and without reservation, and promise to defend this oath against all enemies, both domestic and foreign.

Privacy & Visualization

- Most visualization computation *assumes* unrestricted access to data
  - How do we do visualization computation with partial information?
  - How do you design hardware/software visualization systems to ensure data security?

- **Quasi-identifiers & Doxing/doxxing (document tracing):**
  "Internet-based practice of researching and publishing personally identifiable information about an individual. … searching publicly available databases and social media, ... cyber-vigilantism, hacktivism and cyber-bullying." (Wikipedia)

- Who would potentially benefit from unlimited access to visualization of the complete data?
  - Scientific discovery
  - Improve healthcare
  - Improve education
- What data has privacy concerns?
  - Corporate secrets
  - Health records
  - Academic records
  - Personal finances
  - Personal location

FERPA - The Family Educational Rights and Privacy Act
https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

- Students/parents can inspect & review educational records
- Students/parents can request a correction to their record
- Schools may disclose, without consent, "directory" information
  - @RPI: name, address, photographs, phone #, e-mail, date/location of birth, major field of study, academic load, participation in officially recognized activities and sports, weight and height of members of athletic teams, dates of attendance, degrees, honors and awards received, class year in school, and most recent previous educational institution attended
- *However, schools must allow students/parents to opt out of directory information disclosure*
- Students/parents must be regularly informed about their rights

Health Insurance Portability and Accountability Act (HIPAA)
https://www.hhs.gov/hipaa/for-professionals/index.html

- Long Title: "An Act To amend the Internal Revenue Code of 1986 to improve portability and continuity of health insurance coverage in the group and individual markets, to combat waste, fraud, and abuse in health insurance and health care delivery, to promote the use of medical savings accounts, to improve access to long-term care services and coverage, to simplify the administration of health insurance, and for other purposes."
- Unintended negative outcomes
  - Reduced retrospective chart-based research (responses dropped from 96% to 34% in one study on heart-attack followup surveys)
  - Legalistic details on privacy preservation techniques has made informed consent forms even longer and less user-friendly
  - Stiff penalties for violations, lead doctors to withhold information (even sometimes from people who have rights to see it!)
  - Expensive to implement
  - Requires training healthcare providers

Institutional Review Board (IRB)

- FDA requires all institutions conducting human subjects research have an IRB
  - Researcher training
  - Research plan review & tracking
- Privacy, Confidentiality, Anonymity, and Informed Consent
- Reduce risk (physical/mental/privacy) to the participants engaged in research

---

Institutional Review Board
Rensselaer Polytechnic Institute

Informed Consent Form

I understand that Barbara Cutler, who is a professor of Computer Science at Rensselaer Polytechnic Institute, wishes to interview me as part of the research project on a new Spatially Augmented Reality (SAR) system for education and entertainment applied to games. I understand that she will be making her best possible effort to guarantee me every possible protection, including the following:

1. I am under no obligation to be participate in the study or to be interviewed if I do not wish to do so.

2. I am not obligated to perform any of the game play exercises or answer any of the questions. I may decline to answer any or all of the questions, and I may terminate the study or interview at any point, without giving any reason.

3. Participants for this study will be compensated for their time in the form of a gift certificate at the rate of $10 per hour. This compensation is not contingent upon the subject completing the entire study and will be prorated if the subject withdraws.

4. I will be identified by a randomly assigned ID number that is used only for this study. All recordings and game state files will be labeled with this ID. All information and data relating to the user study will be protected to secure confidentiality. All electronic files will be stored on password protected computers. All paper forms will be stored in a locked office. The correspondence between the ID number and my name will be recorded by Barbara Cutler and be accessible only by her. This correspondence will be destroyed once analysis of the data is complete, within 1 year after participation in the study.

5. If there is anything that I do not wish to have quoted, or any game state files that I do not want made public, I may say at any point during or after the interview what I wish to have kept off the record and it will not be quoted or used in a publication.

6. I understand that if Barbara Cutler decides to use any portions of this interview or any examples of my game play in subsequent publications, that she will send me a copy of the portions of the interview and any game play, including any quotations and paraphrases that she decides to use, for my editing and written approval. I will have the right to edit the material and I will receive a copy of the final publication. She will only use the material that I have approved and the use of all material will be anonymous. I may also change my mind at any point up to and including the review of any quotations and paraphrases and game play that might be used.

7. Based on reading this form (check one):
   _____ I agree to be interviewed.
   _____ I do not agree to be interviewed.

8. The basic camera-projection Spatially Augmented Reality (SAR) setup has been described to me and I have been warned not to look directly at the projector lenses. Standing close to the projector (30cm) and looking directly into the projector bulb for 2 seconds or longer may cause permanent eye damage.

_____   _____   _____
Name of Participant   Signature   Date

Accessibility

Rensselaer Polytechnic Institute is committed to providing access to our educational programs and services for those with disabilities. This includes physical access as well as access to Rensselaer websites and electronic information.

- https://studenthealth.rpi.edu/disabilityservices
- https://info.rpi.edu/statement-of-accessibility
- https://submitty.org/developer/
  interface_design_style_guide/web_accessibility

Inclusivity in the Classroom
Inclusivity in Academia
Inclusivity in the Workplace

https://www.opensocietyfoundations.org/explainers/value-inclusive-education
https://info.rpi.edu/diversity
https://president.rpi.edu/news/memo/11/22/2019/our-commitment-inclusivity

- Different and diverse learning/working side-by-side
- Value different and unique contributions
- Everyone can feel safe and have sense of belonging
- Separate is not equal
- Inclusive curriculum - themes and contributions rom or relevant to marginalized groups