

Running Lecture Outline: Interdisciplinary Research Methods

Fall 2024

Contents

1	27-AUG-24	2
1.1	Miscellaneous	2
1.2	Faculty Intro	2
1.3	Syllabus	2
1.4	Scientific vs Engineering Research	2
2	29-AUG-24	3
2.1	Miscellaneous	3
2.2	Research, Proposals, Funding	3
2.3	Industrial research	3
2.4	Academic research	3
3	3-SEP-24	4
3.1	Critical Analysis of Papers	4
3.2	Assessing Contributions	4
3.3	Reading a paper	4
4	05-SEP-24	5
4.1	Miscellaneous	5
4.2	Critical Analysis of Papers, continued	5
4.3	Research Dissemination	5
5	10-SEP-24	6
5.1	Open access preprints	6
5.2	Predatory Journals and Conferences	6
5.3	Nonscholarly technical documents	7
5.4	Journal Bibliometrics	7
5.5	How are journals and conferences organized?	7
5.6	Types of review	7
5.7	Literature Review	7
6	12-SEP-24	8
6.1	Literature Review and Presentations	8
6.2	Generative AI	9
7	17-SEP-24	9
7.1	Generative AI, Continued	9

1 27-AUG-24

1.1 Miscellaneous

This class is in Gleason, starts at 9:30 AM Tuesdays and Thursdays.

No idea who the prof is. He looks cool though. Makes a parallax error joke with the clock. Quite funny, tbh.

This is a skill-building course. For BME+ChemE, MIE, and ECE. This is mainly for understanding what it takes to understand how to do things for academia.

1.2 Faculty Intro

There's Dr. Vinay Abhyankar (BME), Dr Andres Kwsasinki (ECE), and Dr. Risa Robinson (MIE).

1.3 Syllabus

Syllabus is in myCourses. Highlights include:

- Goals are to make them strong Engineering researchers, focus on Interdisciplinary and trans-disciplinary research skills, with **three themes**: research methods, statistics, and conducting ethical research.
- Interdisciplinary STEM research needs communication skills, lit review, and critical evaluation of research, publication process, research proposals and funding, and to prep for quals.
- Research stats involves how to use data well, and analyze data well. This looks like good AP stat. Also how to recognize cherry-picked data.

Grading wise, attendance counts for 10%, Interdisciplinary Research Methods is 40%, and Research Stats and Conducting Ethical Research are both 25%. COME ON TIME. Check myCourses for Assignments. For weeks 1 – 5, we have mock quals, that are graded. This will be hard. But it's very valuable.

There's also the CITI modules stuff. That's the ethics. We have a statistics exam on November 26th.

1.4 Scientific vs Engineering Research

So, what exactly is research?

Definition 1 (Research). *Research is the intellectual process to create new knowledge.*

This is essentially language building. Measurements prove or disprove our understanding of such theory or data.

Definition 2 (Inductive inference). *Inductive inference moves from observations to a more general conclusion or hypothesis.*

This is the scientific method/research.

Definition 3 (Deductive inference). *Deductive inference is more specific knowledge derived from more general principles.*

On the other hand, engineering research is new knowledge in the form of a new entity of tangible existence.

So, ultimately, what is scientific method/research?

Definition 4. (Scientific Method) *The Scientific method is the principles and procedures for the systematic pursuit of knowledge involving recognition and formulation of a proven collection of data through observation and experiment, and the formulation and testing of hypothesis.*

Only a sith (or a mathematician) deals with absolutes. For that reason, we use statistics, because we're not sure.

Engineering research contrasts from this philosophy because it deems to *create something new*.

2 29-AUG-24

2.1 Miscellaneous

Generally, don't use ChatGPT to write a paper, obviously. He doesn't dispute that they are important tools, so the use is fine, but the intellectual generation of content is still your brain. Also, disclose how you are using those tools. If you are using AI to do some code debugging, for instance, please let them know.

There is now a five minute discussion on the validity of this argument. I am not notating it, because I find it more of a diatribe than actual course content.

2.2 Research, Proposals, Funding

There are three major questions in this lecture:

Remark 1 (Question). *Where is research conducted in the US?*

Remark 2 (Question). *What is a research grant proposal?*

Remark 3 (Question). *How are research grants funded?*

Research in the US is in government, academia, and industry. Mainly industry, because of size. There are three major Research and development categories: Basic Research, which exists to acquire more research on underlying foundations of phenomena and facts, with no immediate application, applied research, which is primarily towards a specific, practical aim, and experimental development, which draws on knowledge gained from research, and is directed to producing new products and processes or improving existing products and processes.

Another way of categorizing this is by the *Technology Readiness Levels*. They are a set of nine levels that define the level of maturity of the research is. Level 1 is basic research level, while Level 9 shows "Actual system flight proven through successful mission operations", while the other ones are somewhere in between. The full set of levels are in the PPT. It was originally conceptualized by NASA, which explains why it says "flight". You can make the list analogously based on your own research.

Governmental research is conducted in a myriad of loci. They include national laboratories, which are "federally funded research and development centers". There is also National Institutes, (think NIH, NIST, USDA, DoD, etc.). TLDR: Government is large.

2.3 Industrial research

Industrial research is done by companies. We think big companies (see NVIDIA, Tesla, Google), but small companies do a lot too. Some industries form consortia to collaborate when conducting research, as well as develop roadmaps and standards. This includes semiconductor device fabrication, telecommunication standards, and airlines.

2.4 Academic research

We begin with four questions:

Remark 4 (Question). *What are the purpose and goals?*

Remark 5 (Question). *Is it important or necessary?*

Remark 6 (Question). *Who does academic research?*

Remark 7 (Question). *Who or what funds academic research?*

The primary use of universities is to educate. The goals for PhDs are to educate the next generation of researchers. There are differences in basic, applied, and experimental research. But why are PhDs necessary?

Professor goes into a diatribe of Bell Labs prof discussing how industry focuses on applied research, and unbalancing the "three legs" of research.

Who does academic research? You, me "Prof referring to themselves", postdocs. Who are we funded by? Obviously by university endowments, gifts, foundations, industry contracts, and government contracts.

Funding varies. There are generally proposals, which vary significantly even within one funding agency, each having many different types of grants ranging from 10k to over 10M, sometimes requiring multiple PIs at multiple institutions. These grant proposals are often reviewed by both peers as well as agency program officers and managers.

3 3-SEP-24

3.1 Critical Analysis of Papers

There are some pragmatic reasons to learn how to better analyze papers. One, this will be on your qualifying exam. But on a more personal level, it makes you a better reviewer. It also makes you more *able to learn*. The courses you are going to be taking will always be one step behind your actual research. Therefore, you need to do a critical analysis of recent papers to make decisions. There exists a duality between ability to better analyse papers and ability to better write papers. You have to learn from others, and you have to better communicate.

Writing paper is *not creative literature*. (Personal opinion: it should be). But the way you communicate research should stay in the lane of established good papers, recognizing the structure of good papers, and ultimately fundamentally writing like the greats.

You also get critical analysis skills. Critical has a different connotation here than in standard language. Here, *critical* means an act of objective judgment - that is, understanding both the strengths and weaknesses of a subject, including the justification of such in an objective manner.

There is fundamentally one question to answer when performing critical analysis: **What is the main contribution of the paper and how is this contribution justified and validated?** That is, what is the new knowledge generated from the paper? And how do you know whether or not it is accurate?

Critical comes from Greek *kritikós*, “of or for judging, able to discern”. There are various levels of analysis you can do, but not all are equally important, meaning that you should avoid superficiality. For instance, if a citation is old, that may be a reason to be critical, but not as strong as seeing a major issue of the paper itself. Obviously, you need to check for rigor and correctness, and that the paper ties up loose ends and gets rid of handwaving.

You should instead look for innovative and significant contribution, including looking for something new. You should make sure the contribution is not trivial. (The definition of “trivial” is hotly contested, though). You want to maybe stop and think if the triviality and obviousness of this claim is true. First of all, the paper was published. Therefore, they didn’t think it was trivial. You may be missing something.

You also need to validate such claims and contributions. No justification is no contributions at all. This is the Sagan Standard. When looking for a good presentation, make sure that it includes items made in the major community. It needs to be succinct, clear to read, with *good style*, which means that there needs to be concrete and clear cause-and-effect chain reasoning. We minimize superlatives, because we’re technical, and trying to communicate new knowledge. We’re not marketing experts. We also want to have a good choice of notation and no unnecessary buzzwords. We also generally want to be self-contained and not rushed. I guess that means don’t procrastinate. Ha.

3.2 Assessing Contributions

To assess contributions, *State of the Art should not be the only criteria*. Just because research is not state-of-the-art, that does not make it bad! This is the equivalent of low-hanging fruit argument. There are different aspects that make new research valuable. But that is not the only criteria! For instance, does the approach show promise? Does the research help you better understand some phenomenon? Does this help develop new research that can be built further on? Is there a mathematical contribution that can support further scientific and engineering research?

Professor then goes into a small diatribe regarding his time at TI.

The anatomy of a technical article varies by disciplines, but generally they include Titles, Abstracts, Intro, Problem Description, Set of Claims, Methodologies, Description of data, Set of Results, and Conclusions.

There is not a fool-proof method, but papers should generally take this path.

Now, there is now people trying to discuss negative results. TLDR: it is hard.

3.3 Reading a paper

If you want to critical analysis of a paper, it’s not really done sequentially! You start off by reading a title. You make a mental note of how the title conveys information on the subject of the paper, and then read the abstract and perhaps the conclusion. There should be:

- One to two sentences as motivation
- One to two sentences defining the Problem
- One to two sentences discussing main contribution
- One to two sentences giving evidence validating the claim

- Are there unsubstantiated claims?

I did parliamentary debate, I'm actually going to use parliamentary debate flow.

Then, you go to the introductory section. Start by checking the motivation: Why do we care? The professor spends time looking at references to prior works in the paper, checking for thoroughness, currency, relevancy, and to check details of claimed contributions.

Next, look at assumptions and conditions. Are they valid? Are they realistic? Do they narrow the scope of the research? Are they justified?

Then, check if the research problem is set up well. Is there ambiguity when analyzing the research elements?

Then, read the method and results. Obviously, check for how warranted the method is and if it is correct. Then, check if the results actually validate the claimed contributions, and whether or not those comparisons are fair. Then, we check if the results analyzed and discussed are done in depth, and if they are fair. Finally, we check if correct methods used in collecting, presenting, and analyzing results, and if sufficient information presented to independently reproduce these results.

Now, we look at a couple of graphs and whether or not there is validity in them. This is present in the powerpoint, so I am omitting the discussion here. Also, check for consistency of units.

4 05-SEP-24

4.1 Miscellaneous

Just reiterating what has happened in the email, the quick message is don't use electronics. Don't use electronics in class. Unless you're peer notetaker. I'm peer notetaker, so it's fine.

4.2 Critical Analysis of Papers, continued

Using a quad chart is a helpful way of presenting information! A quad chart organizes and presents information in the most essential form. It is also something you can use to organize your own thoughts. This method of organizing is direct and clear, which is very helpful. In fact, DoD basically requires it for research proposals. Think of it like an executive summary. It's called a *quad chart* because text is divided into four quadrants. An example is now shown. It is in the PPT, so I am not including it here.

Notably, there is four quadrants. This example shows objective, approach, key milestones, and graphical information in the top right. This forces you to be smart about what you include, as well as conciseness. Be reasonable with font sizes.

Another method is by splitting objective and motivation, contributions, methodology, and results.

Remark 8. *We will have an assignment soon concerning literature review, but as mentioned previously in this part of the course, will be incremental practice for the mock qualifying exam. This will discuss the qualifying exam. There will also be a quad chart on such.*

4.3 Research Dissemination

Research dissemination, in short, is how to publish your research. There are conference presentations as well as journal presentations. Every reference as of now is for peer-reviewed publications. (The implication is that it's actually vetted by experts in the field). Conference publications are generally considered as a venue for disseminating the latest research, including works-in-progress. This has a less careful refereeing process, and means that the manuscript cannot be revised and resubmitted, which means that there is immediate accept/reject based on reviewer scores as opposed to the other way round.

On the other hand, Journal publications are seen more as an archival medium, which implies research maturity, and generally has more details than a conference paper.

Note that this is **very** dependent on discipline. Prestige matters on discipline. For instance, publications to Systems and Control may be useful, but IROS conference papers may be sexier.

At the end of the day, acceptance is often made by score. They accept up to some percent of the top scorers. Typically, this is unfair.

There are also journal versions of a conference paper. This exists. Lol. This is a conversation that happens at the editorial world. When presented with a work of good quality, should we say this is unacceptable? We want to communicate the dissemination of this knowledge

There are different classifications of journal papers, here's a general gist:

- Transactions Paper, which is the typical full-length paper, which can either submitted through an open call, or a call for special issues.
- Letters are same in the concept to a journal, but they are generally shorter, with a quicker review process.
- Review Articles, which are a comprehensive summary of some topic with commentary added. I'm controls guy, so I think of the Ames paper on Control Theory and Applications.
- Tutorials, which are also a way to teach a somewhat new area to the research community. Think review articles that can become textbooks.
- There's also edited book chapters. Think a textbook, but each chapter is written by a leader in the field. This is closer to a chapter in the textbook than a paper.
- There are textbooks.

Now, conference papers are different. Here's some examples, roughly delineated by size (smallest to largest):

- Workshops are generally very focused on a topic, which are typically half-day or full day. There is usually only one concurrent session.
- Symposia are focused on an area of research, distributed over a few days, which may have more than one concurrent session.
- Conferences are focused on a broad area of research, or maybe, even a whole discipline. This is essentially equivalent to multiple concurrent symposia. This is like a "track".

There is also Open-Access repositories of preprints, meaning they haven't been peer-reviewed yet. They are generally in the arXiv. We love the arXiv. Obviously, if you're citing something from the arXiv, make sure there exists a peer-reviewed version of such.

5 10-SEP-24

5.1 Open access preprints

We continue our discussion with talking about open-access preprints, which include arXiv and bioRxiv. Now, the professor shows an example of a paper: *GPT-fabricated scientific papers on Google Scholar* to argue that a lot of AI papers do make it through peer review! This includes at indexed journals, but often in non-indexed journals as well. He then argues:

1. Be careful. Papers appearing in arXiv need not be perfect, and that peer review is still a decent stopgap, but not infallible.

Also, when publishing in peer-reviewed journals, you sign off copyright. This means that you pay a higher publication fee. There are two sides to this: one, does there exist scientific vetting of the information? Also, do the publishers profit? There are a majority of cases that even when you publish, the information is still paywalled. At RIT, this is transparent, because the library pays for these fees. But if you are at home and accessing papers, it may be more difficult.

This is obviously not a good thing for broad dissemination of ideas. Now, there is a push for making federally funded research freely funded without delay. Note that the Biden Administration has made policy guidelines to do this in August 2022, which at the time of writing is 2 years ago. The professor is a big proponent of open-access publications.

5.2 Predatory Journals and Conferences

Definition 5. *Predatory journals and publishers are entities that prioritize self interest at the expense of scholarship (Nature)*

These are scams. The interest is collecting publication fees from the authors, with a fake review process. These are increasingly common. The professor stated that has been receiving these claims to move conference publications to extended journals at their "reputable" journal, and then ask for publication fees. They are not that easy to identify. Check the editorial board, track record, where papers are published, and the sponsoring organization. Do you recognize the people involved? Are they well-regarded in the broad context?

5.3 Nonscholarly technical documents

There are white papers, technical reports, standards, and patents. You will learn a lot more about patents and standards in ENGR-702. They often provide good information for research, but they are not really peer-reviewed, but there is implicit evaluation of content. Standards are often competed by universities and companies, which require performance results, but you should consider to be peer-reviewed.

5.4 Journal Bibliometrics

The most famous information is defined by Clarivate (formerly Thomson-Reuters) called the *Journal Citation Report*. There is the **Journal Impact Factor**. This is calculated as follows:

$$2021JIF = \frac{2021\text{citations from journals, proceedings, and books to a journal in 2019 and 2020}}{\text{Number of articles and reviews(citable items published in 2019 and 2020)}}$$

This is ripe for abuse. There are very reputable journals that ask you to cite 3 – 4 papers from the same journal in the paper to be accepted, which games the IF higher. There is also the Eigenfactor Score, which is not influenced by journal self-citation. This is a bit better.

Eitherways, take these measures with a big grain of salt. You cannot take it as an absolute for what good research is, and specifically based on the research you are doing. IF is dependent on practices within discipline. Citations are a game that you play. In ECE, there is a discipline called information theory, which may have 10 citations. ML may have 5 to 6 times that. And they are technically the same discipline.

Now, a quick discussion on how citations in papers themselves could be unethically sourced.

Metrics should be then evaluated critically, and don't necessarily represent the whole picture. They should be relative to a discipline. But what is a "good score?" JCR basically does a web scrape of all journals and compiles a list for ranking. In ECE, a really good score is like 8-ish but it is super discipline dependent.

For conferences, acceptance ratio is generally used as a metric, which also needs to be used with care, and can be skeped by number of submissions. This is also skewed a lot because of COVID.

5.5 How are journals and conferences organized?

There is an editor-in-chief, which selects editors, makes policy, and chooses special issues. They also assign accepted papers to issues, and acts as the arbiter for disputes between authors, reviewers, editors, and does initial filtering of submissions, and assigns papers to area editors or to editors.

The area editor is generally a more narrow EiC, while Editors are assigned reviewers and makes accept/reject decisions. No one is compensated for this.

There is now a powerpoint slide (page 19) that diagramatically goes through the review process.

5.6 Types of review

There is either single-blind or double-blind review. In single-blind review, author identity is known but reviewer identity is not known. Double-blind review has both the author and reviewer identity unknown. Open peer review is what the presenter prefers. Everyone knows everybody, and everyone roasts everyone. This is preferable because conscious decision-making need not be defined by who you are!

Reviewers are expected to demonstrate expertise and objectivity, and to behave as a peer of the authors. They are also expected to behave professionally, to provide specific criticism, constructive criticism and act in representation of future readers.

Communication etiquette: give a cover letter for journal submission, and explain the rationale for actions taken to address comments and clearly describe major changes made. Similarly, for reviewers, communicate promptly if issues arise, and be professional. You want to to treat others the way you want to be treated.

The last discussion is on anatomy of conferences. The powerpoint slide (slide 23) covers all relevant information.

The next section is on with Literature Review and Presentations.

5.7 Literature Review

If I have seen farther, it is by standing on the shoulders of giants. (Isaac Newton)

So what is a literature review? Lit review is a really important part of the research workflow, and new knowledge is created from existing knowledge. There are some preliminary goals of literature review. One is to find the state of the art of what is already done. This also allows you to identify a research gap, including simplifying assumptions, why the study is needed, and what general study is needed. simplifying assumptions are useful, because it allows us to see what areas need further complexity, and what we need to work on in the future.

The literature review process is a 9-step program :). There are a good number of stages, and we generally start by identifying the research question. But we'll talk about that on Thursday.

6 12-SEP-24

6.1 Literature Review and Presentations

Remark 9 (Homework Reminder). *We have an assignment for Literature Review. Due next week. Check myCourses for details.*

As I stated earlier, it is a 9 step process. Here are the steps:

1. Identify Research Question
2. Keyword Expansion
3. Search
4. Review Findings
5. Categorize/Tag
6. Refine Search Terms
7. Thoroughly Digest Important Articles
8. Synthesize/Outline
9. Write

However, this process is cyclic! You revise and repeat, and you take notes on your own. However, it shouldn't be a simplified table of oversimplified contributions and how those papers seem to relate to each other.

It is also the addition of commentary, the synthesis of ideas that needs to be present.

You start with a research question. What do you want to learn more about? I encourage you to actually write that question down.

Then we do keyword expansion, in order to search, and to narrow/broaden the scope of the paper, as well as provide search term synonyms/antonyms, which makes the literary information somewhat unclear in the future. The ultimate goal is *identification*. That is, what are the search terms necessary to find relevant papers?

Next, use that key word to actually search. Then, you seek to find the "citation train", or a list of papers that contribute into the list of papers in the list.

In that citation train, you want to find the article you found, go forward in time, and then find what cited it? We then want you to go backward in time to find the *seminal paper*, or the paper that inspired this inquiry.

Remark 10. *Citation train is a bit of a misnomer. It's more of a food web than a food chain.*

Both the RIT website and Google Scholar work well for such searching. You can view citation path. This is very useful.

Remark 11. *Creating a citation manager account is so valuable! One example of such a citation manager is Zotero, Mendeley, EndNote. RIT library has a webpage dedicated to this. LaTeX uses bibTex for all of this. I enjoy using ancient technology, so I also use LaTeX. Add sources as you go.*

Then, you review your findings. To do so, you create a bullet list of main points.

Remark 12. *However, DO NOT BE DISTRACTED BY INTERESTING CITATIONS. YOU ARE NOT A RABBIT. DON'T GO DOWN RABBITHOLES.*

This should be a quick process - one hour per paper. Then, you organize your papers. This is generally a personal path. The best way to categorize papers is the best way that works best for you! Adding keywords makes your life easier. Don't tag everything as miscellaneous, though. That's quite cringe.

Then, you should refine the search term to be more specific on what you are looking for. Remember how I said this process is iterative? That also means you iterate/refine your search terms. I would also suggest to set up google scholar alerts to continue searches as time goes on.

You then outline/synthesize all of these sources, and find how these sources necessarily relate to each other, including finding the consensus and disagreements in a topic. This means that you also have a lot of the writing already done in the long run.

Now, Prof shows Overleaf, and .bibTex. Love it.

6.2 Generative AI

GEN AI is the technology behind tools like ChatGPT, Gemini, claude.ai, and various others. For LLMs, words in a text are represented as vectors so that words with similar meaning are vectors close to each other. These are then input into an ANN engineered and trained for pattern recognition in input text, with the pattern generating text mimicing text in natural language, following the input text. Essentially, they are ANNs trained to mimic human conversation.

We engineer these neural networks so that they can learn by recognizing patterns in text. This can lead to *hallucinations*.

His opinion on how this works: this landscape feels like the Internet Search Engine Boom.

Here are some you should look at:

elicit.org does GEN AI specialized for literature review. However, it cannot search behind paywalls! They are generally trying to search behind paywalls.

There is also researchrabbit.ai. It seems to be an AI assisted tool to add papers, explore similar papers, and various other information systems.

Claude.ai is also a group he recommends. You also need peripheral skills, figure interpreters, etc. Also, consider looking at Illuminate. The key point is: they are all *new*.

7 17-SEP-24

7.1 Generative AI, Continued

We continue a quick debrief on AI. Use them with extreme caution. We are young. We can't catch nuisances in previous answers by agents. Generally, avoid the path of least effort, and don't let use of tools impede learning. The only way to learn is through hard work. It's still good to experiment, but run those conversations with your advisor.

Now, he talks about the paper: *The AI Scientist*. There have been other items, like OpenAI o1. There's another one, called SciAgents. I'm tired of AI. The prof sees GenAI more like a personal assistant, which we already have in writing and coding. He sees the next step to be an interlocutor, or an AI collaborator to generate technical, practical ideas. This would be called a "cognitor" according to ChatGPT which is hilarious.

The next 20 – 30 minutes shows a conversation between the professor and Claude regarding contributions. Note that AI systems are still quite mid-tier at working with figures.