CS7240 Modern Internet Research Methods

Course Syllabus for Fall 2026

Delivery: 100% Web-Based, Asynchronous

Instructor Information

Course Instructor: Dr. Maria Konte, mkonte@gatech.edu
Head TA: Anita Rao, arao338@gatech.edu

About This Course

Welcome! This is a research-oriented course that covers new developments in Internet measurement techniques, with an emphasis on topics related to reliability, freedom, and security of modern Internet platforms.

The goals of this course are to:

a. Explore new research topics in the modern Internet interdisciplinary research areas.
b. Familiarize and experiment with techniques, tools, platforms and datasets.
c. Develop new research ideas and deliver an academic research paper. Use the course material as a starting point to brainstorm new research ideas and select a topic of interest. Perform the entire cycle from selecting a research topic, focusing on a specific research question, following through (e.g., data collection and analysis, system design and evaluation, etc.), and finally delivering the results through an academic paper.

Areas covered through the course

The topics the course discusses span three areas:

1. **Measurement Techniques for Internet and Cybersecurity Analytics**: Techniques to map and study the Internet host population along with the services it offers, and how these techniques can be leveraged for different applications, e.g., Internet infrastructure resilience, or cybersecurity analytics.

   Topics covered:

   - **Identifying and studying the "live" Internet host population at scale**. Passive and active scanning techniques for reliably identify which IPv4 addresses correspond to active, reachable devices on the Internet.
   - **IP space utilization inference**: Given IPv4 exhaustion, how can we more precisely infer which portions of the address space are actively utilized versus

assigned-but-idle or completely unassigned? Accurate measurements have important implications for resource management and policy.

- **Service prediction across ports**: Techniques to predict what services (e.g., HTTP, SSH) are running on any port—not just standard ones—using machine learning models, improve our visibility about how services are deployed, even when they don't follow expected port selections.
- **Spoofed traffic detection.** Measurement techniques to detect spoofed traffic that might interfere with scanning campaigns, data collection and measurements.
- **Internet infrastructure hijacking.** Techniques that can be (ab)used to overtake control of the Internet infrastructure (DNS & BGP hijacking). Techniques to detect infrastructure hijacking.
- **The certificate ecosystem.** A brief look into Web management. A measurement approach to identify **stale or revoked certificates** and how they may be abused.
- **Internet voting** systems. An example technique to identify possible risks.

2. **Measurement Techniques for Studying Blocked Internet Access**: Techniques currently used to block access and measurement techniques used to detect when blocking is taking place.

   Topics covered:

   - What is censorship? Reviewing diverse **techniques used to enforce censorship**.
   - How are **censorship observatories designed.** An example observatory. Leveraging public observatories for longitudinal analysis.
   - Measurement techniques to **locate censorship devices.**
   - From rules to **machine learning based approaches** for detecting censorship **at scale**.

3. **Measurement Techniques for Understanding Abuse on Online Platforms**:

   Techniques used to map out the current landscape of abuse (toxicity, harassment, misinformation, etc.). Identify abusive behaviors and study abusive accounts on online platforms. Techniques that leverage social platforms for early warning.

   Topics covered:

   3A. Abuse on social platforms: hate, toxicity and harassment.

   - Classifying content related to hate, harassment and toxicity.
   - Detecting **toxic accounts and studying their trends and patterns.**

3B. How entities abuse online platforms.

- Studying **sock puppet accounts.**
- **Online cybercrime** communities and approaches to understanding them.
- **Mining the GitHub platform** to identify suspicious repositories.

3C. False information spread on online platforms and the supporting Internet infrastructure.

- Identifying **false information** on online platforms.
- **Measuring relationships** among websites that spread misinformation.
- Studying the **ecosystem of Internet infrastructure that supports** misinformation websites.

4. **Sustainable research and ethics**: Finally, the course covers topics related to ethics guidelines when performing large-scale Internet measurements and discusses elements of sustainable research, such as transparency and reproducibility.

   Topics covered:

   - **Reproducibility and replicability** for experimental networking research
   - **Ethical frameworks** for Internet measurement studies

Research Project Format (or "Avenues to Approach a Research Topic")

The research areas are the areas that we cover with lectures-papers presentations. This list only serves as a starting point. The students are welcome and highly encouraged to branch out and explore from there, cutting across traditional boundaries.

Also, there are different "avenues" to explore a research question. The students are encouraged to shape their projects based on their **background, interests, and goals**.

Common approaches include:

- **Literature Review**: Conduct a systematic review of existing research – typically does not involve coding.
- **Survey Study**: Design and distribute a questionnaire and analyze the results to uncover trends or patterns.
- **Data Analysis**: Collect a dataset or use a public one to perform exploratory or in-depth analysis.
- **Learning Techniques**: Apply, evaluate, or even design an AI/ML method for a dataset related to a topic.
- **System Design**: Build and evaluate a system (e.g., a tool, pipeline, or framework) that tackles a specific problem.

- **Replication Study**: Reproduce and reassess results from a previously published paper—this could include publishing new datasets, re-running experiments, or testing under different conditions.
- **Prototype & White Paper**: Design and build a tool through a prototype that shows the core functionality, and write a white paper (typically short) that explains the main technical aspects.

Lectures Format

The instructor presents weekly prerecorded lectures, delivered across **15 modules**. Each module presents 2–4 research papers over 2–4 short videos. Modules can be completed in any order; a recommended pacing is given in the calendar. Each module has an associated quiz. In parallel, students work through milestones that guide them toward completion of their research project and paper.

Course Learning Objectives

The learning objectives of this course are to:

1. Describe the current state of research at the intersection of Internet measurements and cybersecurity — risks the Internet infrastructure faces (e.g., infrastructure hijacking), Internet censorship, abuse and entities on social platforms, web trust management, the false-information ecosystem, ethical guidelines, and principles of sustainable research.
2. Describe a plethora of passive and active measurement techniques, data collection, and analysis approaches for each of the above topics.
3. Demonstrate the ability to apply the learned techniques to different or new research questions.
4. Demonstrate the ability to put together a research project — from a broad idea, to a well-defined research question, to outlining and executing a research approach.
5. Demonstrate the ability to transfer a research project into an academic paper and deliver a presentation of the paper.

Course Materials

There are no required books for this course. The lectures are based on academic papers.

Office Hours

See Ed Discussions for details.

Course Assignments and Grading

- The main component of the course is a semester-long research project. See Canvas for rubrics.

Weight	Component
0%	Acknowledge course policies
1%	Research Interests questionnaire
14%	Weekly Check-in Log & Meetings
6%	Brainstorming I and II
5%	Research Problem and Related Work
1%	GT GitHub repo + GT Overleaf project setup
5%	Project Proposal
1%	Identify target conference
36%	Research Milestones (equally distributed across milestones)
5%	Lecture quizzes
4%	Final project code
6%	Final paper presentation (15–20 min recorded)
6%	Final paper (10–12 pages, academic format)
10%	Class roundtable discussions
+2%	Extra Credit I: results dissemination
+1%	Extra Credit II: course surveys
up to +10%	Extra Credit III: course contributions (artifacts)

- Assignment details
- **Acknowledge the course policies [0%]** — must be acknowledged. Reach out to the instructor with questions about class policies.
- **Research Interests questionnaire [1%]** — short questionnaire on Canvas, cross-posted to Ed. Helps students find peers with aligned interests, complementary skills, and compatible research goals for forming groups.
- **Weekly Check-in Log & Meetings [14%]** — check-ins support progress, identify challenges, and ensure accountability. Beginning in **Week 2**, each group meets with the instructional team every week. You **must meet with the professor the week after submitting your proposal and after each milestone**; in other weeks, any member of the course staff is fine. Each student submits the same brief weekly log

in two places: **Canvas** (for grading) and **Ed** (to share progress with classmates). The log summarizes individual contributions, challenges, and plans, and guides the weekly check-in meeting. Attendance, participation, and thoughtful responses are required for full credit.

- **Brainstorming assignments (I and II) [6%]** — two assignments that summarize the group's progress toward the proposal as the project idea is refined.
- **Research Problem and Related Work [5%]** — each group starts a running draft of the paper describing the main research problem, the paper's main contributions (typically three), and the related work section.
- **GT GitHub repository and Overleaf Project [1%]** — each group starts a GT GitHub repo and a GT Overleaf project and shares the links with the instructional team.
- **Project Proposal [5%]** — each group expands its running draft to include all sections of the paper in skeleton format, explaining problem and approach in detail, plus a schedule distributing work across the semester and among team members. Each group splits work into milestones (e.g., data collection, data analysis, system design, implementation, evaluation). **After the proposal deadline, each group is set** in terms of team members, research problem, and approach — no further changes.
- **Identify the target conference [1%]** — submit the URL of the target conference, along with deadlines, formatting, or submission type.
- **Research Milestones [36%, equally distributed]** — teams submit a PDF of the project's current state based on the milestones stated in the proposal.
- **Lecture Quizzes [5%]** — one quiz per module, open book and open resources. All quizzes due by the end-of-term date in the calendar.
- **Final Project code [4%]** — code hosted on GT GitHub throughout the course.
- **Final Paper presentation [6%]** — recorded 15–20 minute presentation.
- **Final Paper [6%]** — 10–12 page report in academic style format, hosted on GT Overleaf throughout the course.
- **Class Roundtable Discussions [10%]** — in the first two weeks, share research interests on the Ed Discussion forum. From **Week 3** onward, at least one student per week gives a 5–10 minute presentation on the problem they are working on (e.g., their specific problem, or a paper they recently read). Non-presenting students comment constructively with follow-up questions. Each student presents once during the semester; the instructor coordinates the schedule.

- **Extra Credit I: Results dissemination [+2%]** — e.g., (1) set up a page on sites.gatech.edu with title and abstract, (2) submit the presentation link to the OMSCS conference and showcase, (3) create a public GitHub repo hosting code and run instructions, (4) other ideas (e.g., an interactive environment where others can run your scripts) — reach out to the instructors.
- **Extra Credit II: Course surveys [+1%]** — three feedback surveys sent out over the semester.
- **Extra Credit III: Course Contributions [up to +10%, at course staff's discretion]** — artifacts beneficial to other researchers or future course iterations (data collection scripts, documented processes for PACE, benchmarks, well-motivated alternative research directions). Submit with a 1–2 page summary describing the artifact, its purpose, and potential value. Awarded on quality, originality, and usefulness. Exceptional contributions may be acknowledged in a Course Contributions section of the syllabus.

Forming a Team, Team Size, and Proposed Work

- **Why groups?** Writing an academic paper differs from an end-of-semester class report and benefits greatly from teamwork. Almost no academic paper has a single author: the full research cycle is rewarding but time-consuming, and brainstorming and problem-solving with others makes the paper better and improves the chances of acceptance at top venues.
- **How groups form:** Students fill out the research questionnaire and cross-post to Ed, then form groups with peers who have aligned interests, complementary skills, and compatible research goals.
- **Team size: 2–4 students.**
- **Communication:** Groups may coordinate on any platform. All students are added to a dedicated channel on Georgia Tech's Slack workspace, plus a dedicated Ed thread for their project.

Grading scale: A 90–100 · B 80–89 · C 70–79 · D 60–69 · F 0–59.

Course Prerequisites

Geared toward students interested in pursuing a research project and writing an academic paper. The course intersects Internet protocols, computer networks, cybersecurity, and data analysis. Prior coursework in systems, ML/AI, data visualization, data structures, algorithms, or computer architecture is a plus. **The course will not cover undergraduate material** typically taught in undergraduate networking, cybersecurity, or data analysis courses. Students are expected to code in **Python (or a language of their**

choice) at an intermediate level — comfortable with OOP, data structures, control structures, testing, and debugging.

In lieu of a readiness questionnaire, prospective students should be comfortable with and/or passionate about:

- Reading and understanding "An Open Platform to Teach How the Internet Practically Works."
- Defining their own research questions/ideas — working with open-ended projects rather than predefined assignments.
- Student-led projects requiring autonomy and ownership of the work and pace.
- Projects requiring coding skills, plus technical writing and presentation skills.
- Receiving peer-to-peer feedback.
- Working in a group of students with multidisciplinary backgrounds.

Course Calendar (Fall 2026)

Unless otherwise specified, all deliverables and check-in logs are due **Sundays at 11:59 PM AOE**. The initial post for class roundtable discussions is due **Thursday 11:59 PM AOE**, with responses due Sunday 11:59 PM AOE; the presentation schedule is posted on Ed Discussion. Modules may be completed in any order — the pacing below is a recommendation. Each module has an associated quiz, and all quizzes are due at the end of the semester.

Wk	Dates	Module(s)	Key Deliverables
1	August 24–30	1	Research Interests Questionnaire · Brainstorming I · Course Policy Acknowledgement
2	August 31–Sept 6	2–3	Brainstorming II · Group Formation
3	Sept 7–13	4	Research Problem & Related Work · GitHub and Overleaf Setup
4	Sept 14–20	5–6	**Project Proposal**
5	Sept 21–27	7	Identify Target Conferences · Meet with Professor
6	Sept 28-Oct 4	8	Course Survey I
7	Oct 5–11	9	**Milestone I**
8	Oct 12-18	10	
9	Oct 19–25	11	Course Survey II · Meet with Professor
10	Oct 26-Nov 1	12	**Milestone II**
11	Nov 2–8	13	
12	Nov 9-15	14	Course Survey III · Meet with Professor
13	Nov 16–22	15	**Milestone III**
14	Nov 23–29	—	
15	Nov 30–Dec 6	—	**Final paper, code, and presentation** · CIOS Survey · Extra Credit: Results dissemination

End-of-term: Dec 7 (Mon) final day of instruction · Dec 17 (Thu) quizzes due / end of term · Dec 21 (Mon) grade submission deadline · Dec 22 (Tue) final grades available after 6:00 PM ET.

> **Spring/Fall vs. Summer:** Spring/Fall runs 15 weeks with three milestones (Wks 7/10/13). Summer runs 11 instructional weeks with **two milestones (Wks 7 and 9) — no Milestone III**; Milestone II is the final milestone and absorbs the near-final-paper requirements.

Module Summaries

#	Title
1	Course Overview and Reading, Writing & Presenting Papers
2	Surveying the Internet Address Space (I)
3	Surveying the Internet Address Space (II)
4	Overtaking the Internet Infrastructure Control
5	Internet Censorship [I]
6	Internet Censorship [II]
7	Web and Trust Management
8	Measurements & Voting Systems
9	Abuse on Social Platforms
10	Entities on Social Platforms
11	False Information on Web & Social Media
12	The False Information Ecosystem
13	Online Platforms as a Vantage Point to Study Cybercrime Communities
14	Ethics in Internet Measurements
15	Sustainable Research: Transparency & Reproducibility

Module 1: Course Intro & Crash Review — "How the Internet Works" and Reading, Writing, and Presenting Papers

Walkthrough of the course map, the topics covered across the modules, and the learning objectives. A quick review of Internet protocols and how the Internet works via an open platform that offers opportunities for experimentation. Then a technique for reading academic papers, the basic "ingredients" of a well-written paper, and writing and presentation tips.

Topics: What the course is about and its learning goals · overview of course topics and techniques · how each topic is approached · reviewing how the Internet works · structuring an academic paper · delivering a paper presentation.

Readings:

- An Open Platform to Teach How the Internet Practically Works [CCR 2020]
- How to Read a Paper
- LaTeX template
- Writing tips
- How to Give a Great Research Talk (Microsoft Research, 2016)
- How to Write a Great Research Paper (Microsoft Research, 2016)

Module 2: Surveying the Internet Address Space (I)

Active measurement scanning techniques across multiple Internet protocols, plus data collection and analysis to survey the Internet for liveness (Internet host visibility). Covers a proposed taxonomy of liveness spanning responses across multiple protocols, key findings from longitudinal studies covering multiple probe types and replies, and a technique that predicts Internet services across all ports (a learning framework first trained on a small sample to learn patterns between services across ports).

Topics: Internet scanning as a key measurement technique and its practical applications · probing techniques to test host responsiveness · a method to predict hosted services across ports · good Internet citizenship — practical and ethical considerations for designing scanning tools and experiments.

Readings:

- Scanning the Internet for Liveness [ACM CCR, 2018]
- Predicting IPv4 Services Across All Ports [SIGCOMM, 2022]

Module 3: Surveying the Internet Address Space (II)

Passive measurement techniques to survey the Internet for "meaningful usage" — how many of the allocated IPv4 addresses are meaningfully used, and identifying classes of usage (used, routed but unused, unrouted and assigned, available). The techniques analyze captured network traffic from multiple vantage points. Also covers detecting and dealing with traffic that has a forged (spoofed) source IP address, which can complicate data analysis.

Topics: A technique for inferring IP address utilization · classifying IP address space by usage · a technique to identify spoofed traffic that might interfere with probing measurements.

Readings:

- Lost in Space: Improving Inference of IPv4 Address Space Utilization [JSAC, 2016]
- Detection, Classification, and Analysis of Inter-Domain Traffic with Spoofed Source IP Addresses [IMC, 2017]

Module 4: Overtaking the Internet Infrastructure Control

Techniques used by cyber actors to overtake (hijack) Internet infrastructure — DNS and BGP-based hijacking — how they work, their flavors, and how to detect them. Also a survey conducted among network operators about their awareness, the defenses they use in practice, and their willingness to adopt new approaches.

Topics: How cyber actors can overtake (hijack) DNS and BGP-based Internet infrastructure · example approaches to detect hijacking.

Readings:

- Retroactive Identification of Targeted DNS Infrastructure Hijacking [IMC, 2022]
- Profiling BGP Serial Hijackers: Capturing Persistent Misbehavior in the Global Routing Table [IMC, 2019]

Module 5: Internet Censorship I

Internet censorship, reviewing the blocking methods and censoring devices used by censors worldwide. Measurement techniques and datasets for understanding emerging censorship techniques: (1) geoblocking, where users from particular countries or regions are denied access to particular servers, and (2) Twitter throttling, where users experience significant slowdown as a means of discouragement. Also measurement approaches for locating censoring devices and their operations across multiple countries and networks.

Topics: What is Internet censorship — an overview of content blocking methods and approaches · a technique to identify and locate censorship devices.

Readings:

- Network Measurement Methods for Locating and Examining Censorship Devices [CoNEXT, 2022]

Module 6: Internet Censorship II

An Internet measurement platform specialized for censorship research — Censored Planet. Covers its architecture, ongoing deployments, data collection and analysis techniques used to study censorship over multiple geographic regions and networks, and the resulting longitudinal studies.

Topics: What are Internet censorship observatories? An example observatory and how it is designed · a technique to detect censorship at scale, from rule-based to ML-based censorship detection.

Readings:

- Censored Planet: An Internet-wide, Longitudinal Censorship Observatory [CCS, 2020]
- Augmenting Rule-based DNS Censorship Detection at Scale with Machine Learning [ACM SIGKDD, 2023]

Module 7: Web and Trust Management

The certificate authority ecosystem, via measurement techniques based on collecting and analyzing multiple datasets: certificate snapshots, IPv4 scans, popular domains (e.g., Alexa), certificate logs, and others. Also how previously issued certificates are revoked and the associated risks of stale certificates.

Topics: How certificate issuance works · a measurement approach to identify stale or revoked certificates and how they can be abused.

Readings:

- Stale TLS Certificates: Investigating Precarious Third-Party Access to Valid TLS Keys [IMC, 2023]
- Towards a Complete View of the Certificate Ecosystem [IMC, 2016]
- An End-to-End Measurement of Certificate Revocation in the Web's PKI [IMC, 2015]

Module 8: Measurements & Voting Systems

Online voting systems and their associated risks. Making remote voter participation easier and available to more people is an advantage, but online ballot delivery introduces new risks that could even alter major election results. Reviews the OmniBallot platform and the associated client-side risks.

Topics: *(The Summer 2026 syllabus repeats Module 7's topic bullets here — likely a copy/paste artifact in the source document.)*

Readings:

- Security Analysis of the Democracy Live Online Voting System [USENIX, 2021] (initial public version)
- Can Voters Detect Malicious Manipulation of Ballot Marking Devices? [Oakland, 2020]

Module 9: The Landscape of Abuse on Social Platforms

The landscape of abuse and threats on social platforms, focusing on toxic content and the behaviors of online toxic accounts related to hate and harassment, and how they may relate to other emerging online threats. Studying such accounts' behaviors via an example study on Reddit toxic accounts — from data collection to trends and patterns. Also how to approach designing a toxic content classifier and the challenges involved.

Topics: An overview of abuse on social platforms with a focus on toxicity, hate, and harassment · a method to study behaviors, trends, and patterns of toxic accounts · example strategies to identify toxic content.

Readings:

- SoK: Hate, Harassment, and the Changing Landscape of Online Abuse [Oakland, 2021]
- Understanding the Behaviors of Toxic Accounts on Reddit [WWW, 2023]
- Designing Toxic Content Classification for a Diversity of Perspectives [USENIX SOUPS, 2021]

Module 10: Entities on Social Platforms

Abusive accounts on social platforms across a wide spectrum of abuse (spam, fake engagement, illegal content, violence, terrorism). Starts with a taxonomy of these accounts and the challenges of detecting them at scale, plus an overview of existing detection approaches. Then focuses on holistic approaches that leverage the network structure of the social networks on which accounts operate — capturing the overall behavior of an account by creating embeddings that account for network structure, properties, and behaviors of neighboring accounts. Covers how to design such a system, extract features, and evaluate. Finally, a study leveraging social network structure to detect multiple identities in discussion communities.

Topics: A taxonomy of abusive accounts and the spectrum of their behaviors · a technique to study sock puppet accounts · an example technique to detect abusive accounts at scale.

Readings:

- An Army of Me: Sockpuppets in Online Discussion Communities [WWW, 2017]
- Deep Entity Classification: Abusive Account Detection for Online Social Networks [USENIX, 2021]

Module 11: False Information on the Web and Social Platforms

The landscape of false information on the web and social platforms — mechanisms, actors, rationale, and characteristics — plus detection approaches. Also how to identify false information campaigns that spread across multiple platforms (including cross-references in text and videos), via an example study covering data collection, feature extraction, and system evaluation.

Topics: An overview of false information on the web and social platforms · an example technique to detect cross-platform spread.

Readings:

- False Information on Web and Social Media: A Survey [Book chapter, CRC Press, 2018]
- Cross-Platform Multimodal Misinformation: Taxonomy, Characteristics and Detection for Textual Posts and Videos [ICWSM, 2022]

Module 12: The False Information Ecosystem

The ecosystem that supports false information. First, the web hosting ecosystem supporting false-information websites (e.g., narratives and conspiracy theories), including a web crawling technique to collect data, how to form a graph from the collected data, and how to analyze that graph. Then, the Internet infrastructure supporting false-information websites — domain registrars, web hosting, and email providers — via an example study showing how measurements identify and characterize that infrastructure.

Topics: A technique to identify Internet infrastructure (domain registrars, email providers, advertising partners) that supports false-information websites · a graph-based technique to understand the underlying relationships between news websites and false-information websites.

Readings:

- On the Infrastructure Providers that Support Misinformation Websites [ICWSM, 2022]
- No Calm in the Storm: Investigating QAnon Website Relationships [ICWSM, 2022]

Module 13: Online Social Platforms as a Vantage Point to Study Online Cybercrime Communities

Opportunities to leverage online social platforms as vantage points for detecting emerging threats. Looks closely at two types of platforms: (1) GitHub repositories, and (2) channels of online chatter (tweets, blogs, underground forums). Examines how to mine each platform to identify emerging threats — data collection, preprocessing, feature engineering, designing a learning system, and evaluating the proposed framework.

Topics: Online cybercrime communities and how different disciplines approach understanding them · mining online platforms · an example technique to mine GitHub to identify potentially suspicious repositories.

Readings:

- SourceFinder: Finding Malware Source-Code from Publicly Available Repositories [USENIX RAID, 2020]
- The Art of Cybercrime Community Research [ACM Computing Surveys, 2024]

Module 14: Ethics in Internet Measurements

Ethical challenges and risks when collecting, analyzing, sharing, or publishing network data. Many research projects advancing our understanding of an ever-expanding list of topics rely on collecting and analyzing sensitive network data. Reviews the ethics issues and principles: informed consent, human rights, releasing and using shared data, hacking, analysis techniques, ethical review, and Research Ethics Boards (REBs). Two case studies (malware exploitation, classified materials) highlight the challenges and risks in practice.

Topics: Overview of ethical challenges and risks faced by measurement studies, including sensitive Internet data · overview of ethics norms: informed consent, human rights, releasing and using shared data, hacking, analysis techniques, ethical review, and REBs.

Readings:

- Ethical Issues in Research Using Datasets of Illicit Origin [IMC, 2017]
- Understanding the Ethical Frameworks of Internet Measurement Studies [NDSS, 2023]

Module 15: Sustainable Research — The Importance of Transparency, Reproducibility & Replicability

The importance of reproducibility in scientific research: it is crucial for accelerating the transition between science and technology and for establishing the trustworthiness of results. Reviews how ACM defines the cornerstones of sustainable research — repeatability, reproducibility, replicability — their goals and principles, and the challenges to achieving them. Focuses on best practices recommended by academics and practitioners, and the importance of good documentation and other practices through a case study of a large-scale web measurement.

Topics: Goals and challenges of reproducible research · ACM definitions: repeatability, reproducibility, replicability · best practices for scientific Internet research.

Readings:

- The Dagstuhl Beginners Guide to Reproducibility for Experimental Networking Research [ACM CCR, 2019]
- Encouraging Reproducibility in Scientific Research of the Internet [NSF Seminar, 2018]
- Reproducibility and Replicability of Web Measurement Studies [WWW, 2022]

Course Policies

Late submissions & extensions. Students are expected to complete work on time. In case of an emergency, reach out to the TA team through a **private Ed Stem post** so a plan can be made to make up the work, depending on the type of emergency and its impact.

Plagiarism & academic integrity. Students must follow the Georgia Tech Honor Code, including the Graduate Addendum. All incidents of suspected dishonesty are reported to and handled by the Office of Student Integrity. If in doubt whether an action is allowed, ask the instructor/TAs.

Additional course-specific policies:

- **Project overlap.** If your MIRM project is the same or similar to a project you are working on this semester or have worked on previously (other class projects, 8903s, etc.), you must communicate with the instructional team to carve out a piece appropriate and specific to MIRM for the current semester.
- **ACM authorship & AI use.** Follow the ACM policy on authorship and the ACM FAQ on the use of AI.
- **Required acknowledgement.** Per ACM guidelines for Acknowledgements, acknowledge the MIRM course for providing guidance for your paper. The following statement is sufficient:

  "We would like to thank the Georgia Tech OMSCS-8803-O23 Modern Internet Research Methods course instructional team for supervising the first draft of this paper."

  This is pre-filled in `paper-template/paper.tex`.
- **Project visibility.** To inform prospective students about the types of papers the MIRM instructional team works on, a high-level description of each semester's projects may be published on the course website.
- **Ed Discussion code of conduct.** Be respectful when interacting on Ed Discussion; review the student code of conduct.
- **Student resources & accommodations.** The Disability Services team collaborates with students to find creative solutions and reasonable accommodation. Contact the Office of Disability Services at (404) 894-2563 as soon as possible to discuss your needs and obtain an accommodation letter, and send a private Ed Stem message to "Instructors." **The TA team cannot provide accommodations or extensions without an accommodation letter, and accommodations cannot be applied retroactively.**

Communication Policy

Use **Ed Discussion** (available via the Canvas course site) for all communication with the instructional team.

Hosting Requirements

- **Code:** Georgia Tech GitHub
- **Paper draft:** Georgia Tech Overleaf
- Both links must be shared with the instructional team (Wk 3 deliverable).

Contributions

See the Extra Credit III assignment.

Subject to change at any time per instructor.