UNIVERSITY OF MARYLAND

COLLEGE OF COMPUTER, MATHEMATICAL, AND NATURAL SCIENCES
*Department of Computer Science*

# CMSC818I Advanced Topics in Computer Systems; Large Language Models, Security, and Privacy

## Course Information

**Term:** Fall 2024
**Course Number:** CMSC 818I
**Credits:** 3
**Classroom**: CSI 3117
**Course Dates:** Aug 26 - Dec 13, 2024
**Course Times:** TuTh 9:30am - 10:45am
**Qualifying Status:** Yes

**Professor:** Yizheng Chen yzchen@umd.edu
**Instructor Office Hours:** By appointment https://calendly.com/c1z/30min
**TA:** Yanjun Fu yanjunfu@umd.edu **OH:** Wednesday 2-3pm IRB 5112 Desk #22
**TA:** Khalid Saifullah khalids@umd.edu **OH:** Monday 2-3pm IRB 2119

### Course Description

This course will cover advanced topics in Large Language Models (LLMs) for cybersecurity, as well as security and privacy issues of LLMs. We will also explore topics related to AI Agents, in particular LLM Software Agents that can use computers. We will examine the latest research on capabilities and risks of Software Agents. Most lectures will consist of paper presentations and discussions. This course will prepare students to do research in the intersection of LLM, Security, and Privacy. The bulk of the grade will be based on a course project.

### Helpful Resources

- ELMS: https://elms.umd.edu/
- Course Website: https://surrealyz.github.io/classes/llmsec-fall24/llmsec.html
- Sign-up spreadsheet: https://docs.google.com/spreadsheets/d/1iBBBXVsSlXg8FiJpNnQdasjlfhdI80AyZl9tmi2ho48/edit?usp=sharing

### Course Structure

This course is an **in-person** graduate seminar where we will read and discuss academic papers on the topics of Large Language Models, Security, and Privacy together. As we dive into these topics together, my goal as instructor is not to impart knowledge in a one-directional way, but rather to foster a fun and safe intellectual environment where we are all engaged with the material and learning from one another. To better serve this goal, this seminar will adopt a "role-playing" structure, wherein each class session, every member of the class will adopt a different "role" in analyzing the paper. (This is an adaptation of Alec Jacobson and Colin Raffel's approach described here.) In addition to these roles, each class session will feature one or two main student presenters who will prepare a short 10-15 minute (total) presentation as a jumping off point for discussion.

The role-playing structure will have two components: written response to ELMS and in-class panel. On ELMS, for each paper, you are expected to post a short written response or any questions you might have from the perspective of a role (due by 11:59pm the day before each lecture). Your written responses are expected to play different roles by the following counts throughout the semester:

| Scientific Peer Reviewer | Archaeologist | Academic Researcher | Industry Practitioner | Hacker | Private Investigator | Social Impact Assessor |
|---|---|---|---|---|---|---|
| 3 | 3 | 3 | 3 | 3 | 3 | 2 |

In the classroom setting, we will have a panel of students with different roles to initiate the discussions, and each student is expected to play at least three different roles (presenter counts) in the entire semester.

The final component of the course will be a class project, which you may work on individually or in a group of up to three students. The final projects will involve a mid-term project report, final conference-style writeup (along with any deliverables), and in-class lightning talks (last two class sessions). In the past, strong final projects have led to workshop/conference publications, so feel free to reach out to meet with me for feedback if you are interested in submitting your work for publication.

The weekly schedule is as follows:

| Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
|---|---|---|---|---|---|---|
| | **11:59pm:** Reading response due. | **9:30am - 10:45am:** Attend in-person session (CSI 3117) prepared for your "role." | **11:59pm:** Reading response due. | **9:30am - 10:45am:** Attend in-person session (CSI 3117) prepared for your "role." | | |

## Role Playing

For each class session that you are not the main presenter, you will sign up for one of several possible different roles to play during our in-class discussion. The description of the roles are here:

**Scientific Peer Reviewer.** The paper has not been published yet and is currently submitted to a top conference where you've been assigned as a peer reviewer. Complete a full review of the paper answering all prompts of the official review form of the top venue in this research area. This includes recommending whether to accept or reject the paper.

**Archaeologist.** This paper was found buried under ground in the desert. You're an archeologist who must determine where this paper sits in the context of previous and subsequent work. Find and report on one older paper cited within the current paper that substantially influenced the current paper and one newer paper that cites this current paper.

**Academic Researcher.** You're a researcher who is working on a new project in this area. Propose an imaginary follow-up project not just based on the current but only possible due to the existence and success of the current paper.

**Industry Practitioner.** You work at a company or organization developing an application or product of your choice (that has not already been suggested in a prior session). Bring a convincing pitch for why you should be paid to implement the method in the paper, and discuss at least one positive and negative impact of this application.

**Hacker.** You're a hacker who needs a demo of this paper ASAP. Modify the implementation of the paper to make it run on a small dataset or toy problem. Prepare to share the core code of the algorithm to the class and demo your implementation. Do not simply download and run an existing implementation – though you are welcome to use (and give credit to) an existing implementation for "backbone" code.

**Private Investigator.** You are a detective who needs to run a background check on one of the paper's authors. Where have they worked? What did they study? What previous projects might have led to working on this one? What motivated them to work on this project? If you are a panelist, feel free to contact the authors, but remember to be courteous, polite, and on-topic.

**Social Impact Assessor.** Identify how this paper self-assesses its (likely positive) impact on the world. Have any additional positive social impacts left out? What are possible negative social impacts that were overlooked or omitted? Examples of social impact factors in Computer Science: https://perma.cc/K22T-5DFU. A book on AI Safety: Introduction to AI Safety, Ethics, and Society https://www.aisafetybook.com/

## In-Person Sessions and COVID Safety

CMSC818I will meet on Tuesdays and Thursdays IN PERSON from 9:30 am - 10:45 am in CSI 3117. This is not a hybrid class, and there is no option to enroll in this class remotely. In-class participation and engagement is expected, and constitutes 15% of your final grade. However, our top priority is to promote a safe and healthy learning environment. If you are feeling sick or test positive for COVID, do not come to class. Please let me and the TAs know if you are unable to attend for this reason; you will not be penalized. University policy no longer requires the wearing of masks in classroom settings; however, if you are experiencing mild symptoms or have a known recent covid exposure, you are strongly encouraged to wear a mask for the health and safety of your classmates! By default, class sessions will not be recorded.

## Tips for Success

1. **Participate.** Discussions and group work are a critical part of the course. You can learn a great deal from discussing ideas and perspectives with your peers and professor. Participation can also help you articulate your thoughts and develop critical thinking skills.
2. **Manage your time.** Make time for your readings and preparation for participation in discussions each week. Give yourself plenty of time to complete assignments. Don't wait until the last minute to start your final project.

3. **Login regularly.** Log in to ELMS-Canvas several times a week to view announcements and posts. You may need to log in multiple times a day when group submissions are due.
4. **Do not fall behind.** This class moves at a quick pace and each week builds on the previous. It will be hard to keep up with the course content if you fall behind in the pre-work or post-work.
5. **Use ELMS-Canvas notification settings.** Canvas ELMS-Canvas can ensure you receive timely notifications in your email or via text. Be sure to enable announcements to be sent instantly or daily.
6. **Ask for help if needed.** If you need help with ELMS-Canvas or other technology, IT Support. If you are struggling with a course concept, reach out to me, and your classmates, for support.

## Assignments

### Presentation and Discussion Panel

- Sign up to be either a main presenter or one of the seven roles for at least three class sessions. If a single session is shared between two main presenters, please coordinate with the other student presenter. We will do our best to divide up presentations and roles evenly among students. Please bear in mind that being a "main presenter" does not mean you need to prepare to present for the full 75 minutes. Due to the role-based discussion format, the role of the main presenter is to set the stage for discussion with a 10-15 minute walk-through of the paper(s).
- Come to class prepared to discuss the paper from the perspective of your role, along with any additional comments or questions you wish to share.
- Participate in warm-up activities.
- If you are feeling any symptoms, have recently tested positive for COVID, or been exposed to someone who recently tested positive, DO NOT COME TO CLASS. You will NOT be penalized for missing class due to these circumstances. Please let the me and the TAs know

### Reading Response

- You are required to submit a short reading response no later than 11:59pm on the day before the class. You will need to choose a role for your response, even if you do not sign up for the presentation or the panel in the corresponding lecture. Your response should summarize your findings in the role you have adopted for that paper, along with any other comments or perspective you wish to add. Each response could include two questions for the discussion session.

### Exam

- For Fall 2024, CMSC818I is a qualifying course in Computer Science. This means that a minimum of 30% of the grade will be based on exams. To satisfy this requirement, the course will have one take-home midterm exam worth 30% of your final grade. No collaboration or external help is permitted for the exam.

### Final Project

- Work as an individual or in small groups of two or three. If working in a group, an additional statement describing what each individual group member contributed to the project will be required.
- More details and suggested topics will be provided in the semester.

# Grading Structure

## Breakdown

| Assignment | Percentage % |
|---|---|
| Homework (Reading Responses) | 10% |
| Paper Presentation and In-Class Panel | 15% |
| Midterm Exam | 30% |
| Midterm Project Report | 20% |
| Final Project Report | 25% |

## Academic Integrity

The University's Code of Academic Integrity is designed to ensure that the principles of academic honesty and integrity are upheld. In accordance with this code, CMNS does not tolerate academic dishonesty. Please ensure that you fully understand this code and its implications because all acts of academic dishonesty will be dealt with in accordance with the provisions of this code. All students are expected to adhere to this Code.

# Deadlines and Excused Absences

## Late Policy

All reading responses are due 11:59:59pm Eastern Time of the day before the class; all project reports are due 11:59:59pm on the due date. We do not accept any late submissions to the reading responses. For project reports, you are entitled to a late bank of 3 days, however you would like to use it among the midterm or final project reports. If you use up your late bank and submit the report even later, every day will incur an exponential penalty of 50% per day for the assignment grade. For example, if you submit the final project report 4 days late, using 3 days from the late bank, the maximum grade you would receive for the final project report would be half of 25%, i.e., 12.5%.

## Excused Absences

You are required to take the midterm exam at the scheduled time. There are several excused absences from an exam: illness, religious observation, participation in required university activities, or a family or personal emergency. We will work with you to make sure that you have a fair amount of time to make up for excused absences. The best way that we can help is if we know about absences as well in advance as possible.

- Provide a request for absence in writing.
- Provide appropriate documentation (to the instructor) that shows the absence qualifies as excused.

- Provide as much advance notice as is possible, safe, and appropriate.

Please note that, because exams are considered "**Major Scheduled Grading Events**," a self-signed note may not be sufficient: For medical absences, you must furnish documentation from the health care professional who treated you, which must verify the timeframe that the student was unable to meet academic responsibilities. In addition, it must contain the name and phone number of the medical service provider to be used if verification is needed. No diagnostic information will ever be requested.

💡 Please submit all doctors notes and requests for extensions and absences directly to the instructor, and not to a TA.

It is the University's policy to provide accommodations for students with religious observances conflicting with exams. You must inform the instructor prior to the end of the first two weeks of the class if you have a religious observation that conflicts with an exam.

For **missed exams due to excused absences**, the instructor will arrange a makeup exam. If you might miss an exam for any other reason other than those above, you must contact the instructor in advance to discuss the circumstances. We are not obligated to offer a substitute assignment or to provide a makeup exam unless the failure to perform was due to an excused absence.

**The policies for excused absences do not apply to project assignments.** Projects will be assigned with sufficient time to be completed by students who have a reasonable understanding of the necessary material and begin promptly.

## Use of external resources, including LLMs (e.g., ChatGPT)

If you use external sources, you must cite them. Anything you quote must be appropriately indicated (with quotation marks or block quotes), with citations. Your submission must not be substantially quotations — you must demonstrate independent thought. We do not specify a citation format, as long as it is clear.

Any response from Large Language Models (LLMs)—such as ChatGPT, GitHub Copilot, and Google Bard—must be treated as any other external reference: indicate what you are quoting or paraphrasing, and cite the LLM, including the prompt or prompts used. An LLM cannot be the sole source of information; so doing will result in a zero for the assignment: If you are going to use an LLM, you must also include supporting citations.

Please note that LLMs provide unreliable information, regardless of how convincingly they do so. If you are going to use an LLM as a research tool in your assignments, you must ensure that the information is correct and addresses the actual question asked.

## Campus Policies

It is our shared responsibility to know and abide by the University of Maryland's policies that relate to all courses, which include topics like:

- Academic integrity
- Student and instructor conduct
- Accessibility and accommodations

- Attendance and excused absences
- Grades and appeals
- Copyright and intellectual property

Please visit https://gradschool.umd.edu/course-related-policies for the Graduate School's full list of campus-wide policies and follow up with me if you have questions.

## Accessibility and Disability

Students who have been certified by the Accessibility and Disability Service as needing any type of special accommodations must see the instructor **as soon as possible during the schedule adjustment period (the first two weeks of class)**. Please provide ADS's letter of accommodation and any other relevant paperwork to the instructor at that time. All arrangements for exam accommodations as a result of disability must be made and arranged with the instructor **at least five business days prior to the exam date**; later requests (including retroactive ones) will be refused.

## Course Evaluations

Please submit a course evaluation through CourseEvalUM in order to help faculty and administrators improve teaching and learning at Maryland. All information submitted to CourseEvalUM is confidential. Campus will notify you when CourseEvalUM is open for you to complete your evaluations for fall semester courses. Please go directly to the Course Eval UM website to complete your evaluations. By completing all of your evaluations each semester, you will have the privilege of accessing through Testudo, the evaluation reports for the thousands of courses for which 70% or more students submitted their evaluations.

## Information Subject to Change

Although every effort has been made to be complete and accurate, unforeseen circumstances arising during the semester could require the adjustment of any material given here. Consequently, given due notice to students, the instructor reserves the right to change any information on this syllabus or in other course materials.