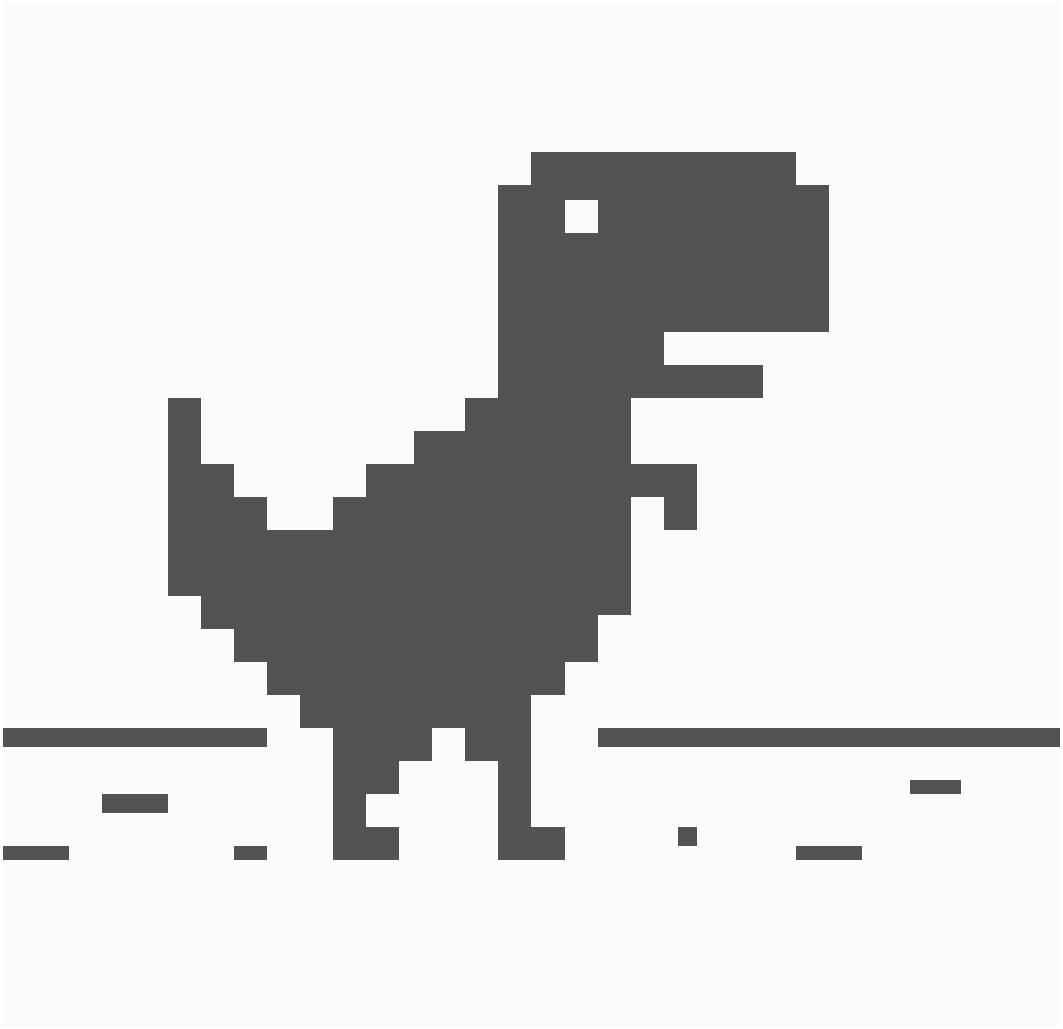
Roars Lab's Handbook



Roars' Members

George Mason University

Contents and Summary

Preface

Part I

Introduction

Chapter 1

Meeting

- Meeting is *mandatory* for all lab members and is in-person.
- We meet *every Thursday from 2pm to 3pm*. Typically, our meeting directly follows the Software Engineering seminar (Thurs from 1pm to 2pm).
- We meet in the main CS conference room (Room #4201 Nguyen Engineering Building).
- If you need to attend remotely, you can join via Zoom but you need to inform me ahead of time.

Part II

Writing

Oral and written communication is critical to research and science. You can create the most powerful systems with ground breaking results, but none of this matters if you cannot communicate or *"sell"* these to others.

Chapter 2

PUblications

2.1 Where do we publish?

We aim to submit our work to the highest quality, most prestigious—"tier 1"venues but also the best fit for a specific project. Such top venues (generally) provides top peer review, a strong community of researchers, and if our work is accepted, provides visibility, credibility, and the opportunity for the most impact.

Conferences Typically we aim for conferences considered by CSRankings and other well-known ones in SE and PL. Depending on the topics of the work, we may also consider other venues. For example, recently we also publish in Logics and AI/ML venues.

- SE venues: ICSE, FSE, ASE, and ISSTA.
- PL venues: POPL, PLDI, OOPSLA. We have never published at POPL but it is a good venue for more theoretical work.
- Logics venues: CAV and TACAS.
- AI/ML venues: NeurIPS, ICML, ICLR, AAAI.

Journals Journals are not the main publication venue in most CS fields. However, we sometimes submit extended versions of our conference papers to top SE journals such as TSE and TOSEM.

2.2 How to write research papers

- Always use LaTeX for writing papers. Do not use Word or Google Docs.
- The paper should *tell a story*. Humans are storytellers and story receivers.

- Be concise. Use the fewest words to communicate effectively. For example, change "In order to" to just "To": that's 2/3rds cost savings for the same information.
- Avoid passive tense
- Know the difference between /Such as/ and /like/
- Avoid long paragraphs
- Spell out numbers less than 10 (/three/ steps instead of /3/ steps).
- Use Eq. X instead of Equation X
- Don't put /etc/ in e.g., or such as
- New paragraph for new thought/idea. Sentences in the same paragraph should be connected.

2.3 Roar Lab Writing Style Guide

Writing Introductions

The Introduction is the most important part of any paper. It must both excite your reader and provide a high-level view of your entire paper. Indeed, by the end of the Intro, your reader should want to read more but also have sufficient understanding about the key takeaways that if they stopped, they would "largely get it."

The intro has 5-6 paragraphs:

2.4 Authorship order and inclusion criteria

- The first author is often the *lead* student on a project, the last author is usually the PI or senior author, and the remaining authors are ordered based on contribution.
- Establish authorship order early in the paper writing process. We do not want to have authorship disputes at the end of the process. Many conferences also require authorship order at submission time.
- Who will *not* be an author? If you do not explicitly fund the work and do not participate in writing or discussions about the paper/work, you should not be an author. This might acceptable at some places, but not in our lab.

2.4.1 Determining authorship

ACM's Criteria for Authorship Anyone listed as Author on an ACM submission must meet all the following criteria:

- They have made substantial intellectual contributions to some components of the original work described in the manuscript; and
- They have participated in drafting and/or revision of the manuscript and
- They are aware the manuscript has been submitted for publication; and
- They agree to be held accountable for any issues relating to correctness or integrity of the work

Chapter 3

Rebuttal

- Start with a *thank you note* to the reviewers for their time and feedback. Reviewing is a thankless job and reviewers often put in lots of time and effort to provide feedback. Even if the reviews are harsh and you disagree, do it anyway. It can make them more receptive to your points.

Chapter 4

Presentation

- **Number your slides**. This is useful when people want to ask questions (or give feedback) on specific slides (e.g., *"on slide 10, you said ..."*).
- **Every slide takes approximately 1 minutes**. Using this rule of thumb, you can estimate how many slides you need for your presentation.

Chapter 5

Data and Code Management

5.1 Data Availability

- We will *always* make the data available. So make sure prototypes, tools, results, etc are avail on Github.
- Github repo:
  - `README.md`: include a description of the project, how to install and use the tool.
  - For tools submitted to conferences, include scripts to generate the data and run the experiments.

Part III

Classes

- Independent Study
  - You need to be my student before you can register for independent study with me. I.e., unless I officially advise you, you cannot register for independent study with me.
  - You will typically get an A grade if you do the work satisfactorily. This is one of the reason why PhD students often get high GPAs, and also the reason why no one cares about your GPA after you got a PhD.
- Dissertation/Thesis Credits
  - Register for CS 998 (Dissertation Proposal) for ..
  - Register for CS 999 (Dissertation) for ...
  - If you perform satisfactorily in your research, you will get an **IP** (in progress) grade, which then converted to S (satisfactory) at the end of the semester when you advance to candidacy (for CS 998) or graduate (for CS 999).
  - If you do not perform satisfactorily, I will talk to you directly.
  - If I cannot even reach you, I will give you an **NC** (no credit) grade. This can happen, e.g., in a case I said I would be willing to give the student a try but the student never contacted me during the semester or made any progress.

Part IV

Reimbursement

Chapter 6

Conference Travel

Meals We are paid per diem for meals when we travel for conferences. The per diem rates depend on the city we are visiting. This means you don't need to save or submit receipts for meals. If you don't eat anything (please do not do this), you will still get the per diem amount.

The first day and the last day of travel, they will not pay the full per diem rate because you are not there for the full day.

Part V

Equipment and Resources

Chapter 7

Connecting to Wifi and VPN

7.1 Connecting to Wifi `Eduroam`

- https://its.gmu.edu/knowledge-base/eduroam/

Linux https://its.gmu.edu/knowledge-base/how-to-set-up-eduroam-on-linux/

- Select `eduroam` from the list of available wireless networks.
- Configure the following settings:
  - Wi-Fi security: WPA & WPA2 Enterprise
  - Authentication: Protected EAP (PEAP)
  - Anonymous: leave this blank
  - Domain: `gmu.edu`
  - Check No CA certificate is required
  - PEAP version: Automatic
  - Inner Authentication: MSCHAPv2
  - Username: Your FULL George Mason email address (e.g., `netid@gmu.edu`)
  - Password: Your Patriot Password
- Finally, click `Connect`. You should now be connected to `eduroam`.

7.2 Connecting to VPN

TBD

Chapter 8

Lab Servers

For all questions regarding lab servers, discuss in **Discord #servers channel**.

8.1 Server Specs

We have **four** servers (pictures).

1. **Boba**: ideal for AI-based projects, e.g., generative LLMs
   - NVIDIA DGX Spark
   - CPU: NVidia GB10 Chip, 20 core Arm (10 Cortex-X925 + 10 Cortex-A725 Arm)
   - Unified Memory/Graphic: 128GB RAM
   - OS: NVIDIA DGX OS (based on Ubuntu LTS)
2. **Pizza**: ideal for projects that can leverage many cores and specialized GPU.
   - Custom built
   - CPU: 32 cores, 64 threads, 3.6GHz (AMD Ryzen Threadripper PRO 5975WX)
   - Graphic: 2x Nvidia RTX 4090 24GB RAM
   - Memory: 128GB RAM
   - OS: Debian Stable

     This system is ideal for projects that can leverage many cores and specialized GPU.
3. **Sushi**
   - Lenovo ThinkStation P620

- CPU: 64 cores, 128 threads, 4.3GHz (AMD Ryzen Threadripper PRO 3995WX)
- Graphic: Nvidia RTX 5090 RTX 24GB RAM
- Memory: 128GB RAM
- OS: Debian Stable

4. **Taco**: ideal for experiments that need fast CPU (but not lots of core) and fast GPU.
   - Alienware Aurora
   - CPU: 16 cores, 24 threads, 5.20GHz (Intel 12th Gen Core i9-12900KF)
   - Graphic: Nvidia RTX 3080 Ti 12GB RAM
   - Memory: 64GB RAM
   - OS: Debian Stable

For Debian machines, **DO NOT** install anything other than Debian stable packages to these computers (e.g., **do not mix unstable or new repositories** to `apt` sources list). You can download and install anything in your home (local) directory. If in doubt, contact us.

8.2 Getting an Account

To get an account on the server, send a request and your **SSH public key** (for *password-less* log in).

8.3 Miscs

1. You can use an IDE such as `VSCode` to do remote development on these servers (just connect to the server via the builtin remote login method in VSCode). You can also use other editors that support remote development (e.g., `Emac` + `TRAMP`).

2. Do take advantage of the multicores these machines have. Most programming languages support multiprocessing (e.g., `Python`) and can significantly reduce program runtime.

3. Periodically check `top` and `htop` to make sure **you** are not taking too much CPU or memory.

4. Periodically check `/var/tmp` or `/tmp/` to make sure **you** are not taking too much space there.

8.4 Pictures

Fig. 8.1: **Sushi**

Fig. 8.2: **Sushi Cooling System**

Fig. 8.3: **Pizza**

Fig. 8.4: **Boba**

Chapter 9

Lab Website and Other Online Resources

- Our main Roars lab website is at https://roars.dev/. It is in plain HTML. All code is open source and available on GitHub (https://code.roars.dev).

Appendix A

LoRs

A.1 Asking

- If you want a letter from me, you should **ask** at least 2 weeks in advance (ideally a month).
  - **DO NOT** put my name in the application system without explicitly asking me first.
- If I don't know you well enough, I will let you know and advise you to find someone else. But if you insist, I will still write you a letter.
- You need to **waive your right** to view the letter. It is a confidential evaluation, and I will not write one if you insist on seeing what I write.
- You should provide me your CV and, if you'd like, your SOP, so I can determine whether I could complement some of the things you've said.
  - Feel free to provide me with any other information that you think is relevant, e.g., your grades, projects, research experience, awards, etc. I might not use them if I don't think they are relevant, but they might help if I miss something.

A.2 Writing

- I **will not ask you to draft a letter** for me. My reputation is on the line, and I do not want a student (or AI) writing a letter on my behalf.
- I **do not write negative letters**, i.e., I don't say bad things about you (even if you're bad). However, if I don't know you well then a neutral, short, or generic letter will not help your case given the competitiveness.

- It takes me about **an hour** for a strong letter for someone I know well, and about 10 minutes for a letter for someone I don't know well.
- In all letters, at the end I will include a short paragraph about **an area you need to improve**, regardless of whether you're super strong or weak. But I will phrase it in such a way that the graduate study environment or a good advisor can help you overcome it. It makes the letter more complete and not just full of praise.
- I **do not customize** the letter for specific schools.
  - You customize your SOP to explain why you want to apply to school `X` and work with professor `Y`. I do not need to explain why you want to go to school `X` or why you would be a good fit for professor `Y`. Note that if I know you and `Y` very well, then I might send a letter directly to `Y` to mention you.
- My letter will have the university logo and my (digital) signature.
  - However, I should note that when I read a LoR, I do not pay attention to whether it has a logo or signature.

A.3 Sending

- After you put my email into the application system, it will send me a request email with a unique URL to go to. It will also give me some deadline that is *likely* not the same as yours—usually later.
  - In some cases the systems don't even give me a deadline (various LoR request examples are shown in my `Demystify` book, e.g., Section 3.2.3: https://github.com/dynaroars/phd-cs-us/).
- I **do not mind if the student is anxious and sends me multiple reminders**. I will likely not respond to them. But I am not bothered by reminders. Over-communication is better than under.
- Most systems simply ask me to upload the letter—though a few have short questions like comparing the applicant to undergrads or grad students. As an adcom reviewer, I don't really pay attention to these comparisons—only the letter matters.
- I usually send my letter (in PDF) in **batch mode**, e.g., I just sent out 10+ letters all at once and cleared my inbox of these requests.

  - For each request, it took me less than 30 seconds to upload the PDF and hit `submit`. So do not worry that it is time-consuming or a burden; it's not.

- After I send my batch, I'll let you know I just sent everything and ask if I missed any.
  - I would appreciate if you share an online spreadsheet showing what schools you have asked me to write letters for so I know what to expect (and if I miss something).

A.4 Updating

- Let me know your progress and especially your outcomes, e.g., interviews, offers, and where you eventually go to. I'd love to hear these updates from you.
- This is a common courtesy. Though I probably won't remember or expect you to do this. However, if you do this and then years later reach out for another letter, I might not be able to help you.

Appendix B

Templates