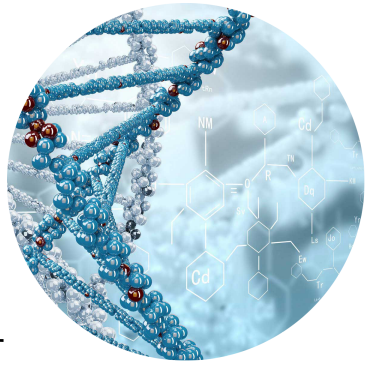
Manual for the Computational Data Science Lab at the University of Texas Arlington



This manual summarizes a set of resources, guidelines, and rules for successful education, research, and communication of Science among members of the Computational Data Science (CDS) Lab in the College of Science at The University of Texas at Austin.

For more information, contact:

Amir Shahmoradi ,
Principal Investigator (PI), CDS Lab,
Assistant Professor, Department of Physics,
Data Science Program, College of Science,
The University of Texas at Arlington
655 W Mitchell St | SEIR 365
Arlington, TX 76010-1765
a.shahmoradi@uta.edu
Tel: (817) 272 9047
shahmoradi.org
cdslab.org

Contents

1 Introduction

The Computational Data Science (CDS) Lab at The University of Texas at Arlington (UTA) carries out research in a wide range of scientific disciplines, including **Biomedicine**, **Biophysics**, **Astrophysics**, **Mathematical Modeling**, and **Scientific Software Development**, all of which require intensive usage and development of Computational and Data Science methodologies and algorithms.

While the research carried out in the CDS Lab is purely theoretical / computational, we collaborate frequently and welcome new collaborations with experimental groups and clinicians across the UTA campus, DFW Metroplex, as well as nationally / internationally.

The CDS Lab is proudly a member of the College of Science at The University of Texas at Arlington and is heavily involved in the development of the Data Science Program as part of the UTA's Strategic Plan | 2020.

This manual is written for the current, new, or prospective members of the CDS Lab to help them communicate, collaborate, and carry out research at the highest levels of efficiency and scientific integrity possible.

2 Current and Past Members

For information about current and past members, as well as the Lab's PI, see https://cdslab.org/people/.

3 Communication

3.1 Instant Communication

- **Slack:** https://cdslaborg.slack.com
  Slack is the primary tool for instant communication among members of the CDS Lab. All lab members should download the Slack Application on their mobile device, and use it for instant communication with individual or all members of the lab, including the Lab's PI (Amir). Non-urgent lengthy time-consuming issues can be discussed via email or in person.

3.2 Other Communications

The CDS Lab has webpages on multiple media platforms, including,

- **Medium:** https://medium.com/cdslab
  This is a social platform that we mostly use for communicating medium-to-long pieces of scientific research or educational material with the world, prepared by the lab members in the form of blog-posts. All lab members are welcome and encouraged to contribute posts to this page. Medium is a highly-popular platform for advertising your work to the entire world. Writing an elegant readable article on this page that could attract public's attention, would significantly boost your profile as an expert in your field.
- **Facebook:** https://fb.me/cdslab
  Facebook is mostly used for communicating the lab news and articles with the Lab members as well as public community on Facebook. You can follow the page by clicking on the link provided above.
- **Twitter:** https://twitter.com/cdslaborg
  Similar to Facebook, Twitter is mostly used for communicating the lab news and articles with the

Lab members as well as the public community on Twitter. You can follow the page by clicking on the link provided above.

- **Instagram:** https://www.instagram.com/cdslaborg
  Similar to Facebook and Twitter, Instagram is mostly used for communicating the lab news and articles with the Lab members as well as the public community on Instagram. Obviously, any news that appears on Instagram has to be followed by an illustration or photo that would serve as a visual explanation of the news.

- **GitHub:** https://github.com/cdslaborg
  GitHub is a highly popular code repository that we use for keeping track of changes in our scientific codes and data, as well as for sharing our codes and data processing pipelines with the entire world. This act of sharing, once it is complete, is permanent and will be there for as long as GitHub exists. As a member of a data-intensive lab, you are expected to have a GitHub account. There are also other, almost equally-capable online code repositories available, such as Bitbucket. However, at the moment GitHub is the primary code repository that we use in our lab.

- **Dropbox:** https://dropbox.com/
  Like GitHub, we often use Dropbox as a way of communicating all digital information such as code, data, articles, manuscripts automatically. However, unlike Git and GitHub which are used for professional version controlling and sharing of our codes and pipelines with the entire world, this mode of communication (by Dropbox) is solely used for our own internal communications.

  If you don't have a Dropbox account yet, contact the Lab's PI (Amir) as soon as possible to send you an invitation email.

4 Expectations and Responsibilities

4.1 Everyone

As a member who has just joined the CDS Lab or, is a current member, you are expected to,

- read, fill out, and sign the CDS Lab questionnaire/checklist and return it to the Lab's PI (Amir)before starting your work in the lab.
- be happy while you are in the lab (and in general, in everyday life). If there is a reason to be unhappy – in particular, about your presence in the CDS Lab – then you should bring it up to your mentor (e.g., the Lab's PI (Amir)) and ask for their advice/help.
- talk to other lab members, effectively and respectfully, and seek help from them whenever needed.
- **be on time for appointments and meetings, whether it is lab meetings, or meetings with the Lab's PI (Amir) or other lab members**. Every lab member's time and schedule is precious and everyone else in the lab should respect and recognize this simple fact. If you cannot make it to a meeting or appointment, you should inform the Lab's PI (Amir) and other people involved in the meeting well in advance.
- schedule frequent regular meetings with your mentor/supervisor, who is the Lab's PI (Amir), in most cases.

- attend the weekly lab meetings, unless a legitimate reason for your absence is provided to the Lab's PI (Amir) well in advance of the meeting.
- be already fluent upon joining the CDS Lab or, be willing to quickly become fluent in **Scientific Programming** and **Data Analysis**. Computation and Data are two of the three fundamental pillars of the CDS Lab as explicitly mentioned in the lab name. The third pillar is the hypotheses/-models that we develop to solve research problems.

  Everyone in the lab is therefore, expected to be well-versed in at-least one high-level programming language such as **MATLAB**, **Python**, or **R**) and be preferably well-versed in at-least one high-performance compiled programming language (such as **Fortran**, **C**, or **C++**).

  Although the use of **Microsoft Excel** is not prohibited in the lab, it is highly unprofessional for a computational and data-science lab member to solely rely on Excel for all sorts of data manipulations and computations needed for their project.
- be well-familiar with the foundations of Probability Theory or at least, be highly motivated to learn it quickly upon joining the lab. **Probability Theory is the language of Scientific Inference** and it is an essential component of all projects in the CDS Lab.
- already be or be willing to quickly become well-versed in **LaTeX** typesetting language upon joining the lab. LaTeX is the principal typesetting system used for drafting reports and manuscripts in our lab and virtually every research lab across the world. Nevertheless, it is fine to use Microsoft Word documents for temporary quick notes or letters. But any document that is rather permanent and sophisticated, or lengthy must be absolutely typed and prepared in LaTeX .
- be present on campus on most working days of the week, if not all days and weekends. If you are **not** going to be present on campus, you must inform your supervisor/mentor – in most cases, the Lab's PI (Amir) – about your anticipated absence. If your absence is not due to a sickness or an emergency, then **you must remain responsive to all email inquiries** by you supervisor during your absence period, unless a communication plan with your supervisor is set up in advance.

  If you are sick, please do not show up at work; Stay home, inform the Lab's PI (Amir), and rest until you recover.
- use your assigned space and work hours **absolutely and solely** for the purpose of advancing the project to which you are assigned. Note that in academia, graduate students, postdocs, and staff scientists are given more flexibility in their work hours, **provided** they make **sufficient progress** on their projects.

  This ï¿½academic flexibility in work hours does not extend to other *administrative* paid positions because people in such positions are responsible for maintaining a consistent workflow and structure for the rest of the members and components of the lab to function properly.
- use your assigned space in SEIR building, or be prepared to lose it to someone else in the lab or outside the lab. Space is an asset in the newly-constructed SEIR building on the UTA campus. Currently the CDS Lab has two offices (SEIR 337, 338) assigned to its members. SEIR 337 is shared with the students of Dr. Leili Shahriyari from the Department of Mathematics.

- document every request you have and provide a written request to the Lab's PI (Amir) for any approval needed. It is important for you to document every communication that we have in the lab, from simple approval requests, to long-time scientific discussions.

4.2 The Principal Investigator

The CDS Lab Principal Investigator (PI): Amir Shahmoradi is responsible to,

- provide long-term vision for the lab.
- help individual members of the lab to develop vision for their own research, academic life, and long-term career plans for as long as they are members of the lab. This will be done through continuous mentorship and directions provided by the Lab's PI (Amir) to lab members.
- actively seek external funds to support research and members of the lab.
- support you in your personal growth and life by giving you ï¿½flexibility in working hours and environment, and encouraging you to do things other than science.
- set up regular meetings with you, go through your work, read your drafts and manuscripts, and talk to about science and your research progress.
- get together with you sometimes for a coffee, lunch, or some social fun event.

4.3 Postdoctoral Fellows

A **postdoc** is a researcher with an earned PhD degree who works under the supervision of a mentor as part of a larger research group. A postdoctoral position is meant to prepare young researchers to become principal investigators or junior faculty members or prepare them toward a research-oriented industry career path. Therefore, as a postdoc in the CDS Lab, you are expected to,

- be regularly present on campus and communicate effectively and regularly with the Lab's PI (Amir).
- be able to **conduct research rather independently**, either on a pre-specified project or one of your own-designed creative ideas.
- be able to **prepare and write scientific papers**, independently of your supervisor, based on the research you perform in the lab.
- be able to take on senior responsibilities such as **mentoring students**, **grant writing**, and **teaching** junior lab members.
- be able to move toward being and behaving more like a Principal Investigator, for example, by **giving invited and contributed talks** at conferences regularly and, by **maintaining and establishing new collaborations** with researchers in other labs or institutes.

4.4 Graduate Students

Graduate students, in particular PhD students, are the backbone of any academic research group. As a graduate student in CDS Lab, you are expected to,

- **know the scientific literature of your own research topic better than your supervisor**. Of course, this is not a knowledge that a new PhD students could achieve in a few days or weeks upon joining the lab. However, as a rule of thumb, a successful graduate student (i.e., a student who has high chances of begin hired by industry or academia upon graduation) should have a knowledge of the relevant research literature that is comparable to their supervisor's knowledge after 1-2 year(s) of work, certainly by the time of their dissertation defense. This can be achieved only by continuously reading the relevant research articles on a daily basis over the course of your PhD, almost **one paper a day**.
- seek out and apply for college-wide, campus-wide, state-wide, and nationwide awards and fellowships (including travel awards to conferences, etc).
- realize that **a typical successful graduate student works at least 10 hours per day, seven days of the week**. This includes studying for the courses and doing research.
- realize that sometimes even 70 hours of work per week may not be enough to achieve a goal or meet a deadline, and that you may have to pull an all-nighter, perhaps for days, to meet the deadlines.
- realize that there also times to leave work early and live like others.

4.5 Undergraduate Students

As an undergraduate student joining the CDS Lab, you are expected to,

- realize that the most important aspect of your undergraduate academic life is your performance in the courses in which you are enrolled.
- enroll in an independent study course with the Lab's PI (Amir) to receive credit for your time in the lab. As an undergraduate student, you should also plan on producing an annotated bibliography of 5-10 articles on your selected research topic and making a 15-minute or longer presentation of your studies/research at lab meeting, sometime during the semester.
- be **utterly reliable and willing to help with whatever projects need you**. At a bare minimum, reliability includes,
  1. showing up on time,
  2. sending your weekly schedule to your mentor (e.g., the Lab's PI (Amir)) and,
  3. making sure that all of your work is accurate (double-check everything in your computations/analyses).
- use your first semester in the lab as an opportunity to see whether continuing in the lab is a good fit for you. After your first semester, you will meet with the Lab's PI (Amir) to discuss whether you should continue in the lab.

5 Employees

If you join the CDS Lab as an official employee of the University of Texas at Arlington or, at some point get a paid appointment in the CDS Lab, then you should keep in mind the following important notes and expectations:

- Paid employees, whether full-time or part-time, whether student or postdoc, are expected to use their time and resources that are provided to them absolutely and solely to support the projects to which they are assigned. **Failure to do so, could lead to the termination of the employee's contract**.
- **Hours of work:** The CDS Lab maintains the same hours of work as other organizations and entities within UTA: 8:00AM-5:00PM, with 1 hour at 12PM-1PM for lunch. Given that we are working in an academic environment, which is more flexible than industry, it is fine to slightly shift your regular work hours to better adjust it to your own lifestyle. However, every paid employee in the lab must work at-least 8 hours. This is the minimum number of work hours per day that is expected from every regular UTA employee.

  To maintain full-time benefits eligibility, your work-hours per week must not drop below 40; if on a given week you work substantially less than 40 hours, your accrued vacation time will be used to make up the difference (per UTA Human Resources (**HR**) policy).
- **Time-Off and Vacation-Leave:** Requests for time-off and breaks from work or for vacation should be made at least two weeks in advance, in a written email to the Lab's PI (Amir). In that case, you are responsible for making sure you that have sufficient vacation hours to cover any time off; otherwise, you will not be paid.

  Per UTA HR policy, every *full-time* employee of UTA accrues 8 hours of vacation-leave per month, with a maximum of 180 hours of vacation-leave, only if they been employed by the State of Texas (including UTA) for less than 2 years. This number is 9 hours per month with a maximum of 244 hours of vacation-leave for employees with at least 2 but less than 5 years of continuous employment status within the State Government of Texas. For more years of employment, consult UTA HR.
- **Sick-Leave:** Sick time should be also requested over email to the Lab's PI (Amir). Per UTA HR policy, full-time employees accrue sick leave at the rate of eight hours per month. At any point during your employment, if you use more than the total number of sick-leave hours you have accrued, you will need to provide medical verification for illness.

6 Code of Conduct

6.1 Essential Policies

The CDS lab, and the university, is an environment that must be free of harassment and discrimination. All lab members are expected to abide by the UTA policies on discrimination and harassment, which you can (and must) read about here. The lab is committed to ensuring a safe, friendly, and accepting environment for everybody. We will not tolerate any verbal or physical harassment or discrimination on

the basis of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, or religion.

We will not tolerate intimidation, stalking, following, unwanted photography or video recording, sustained disruption of talks or other events, inappropriate physical contact, and unwelcome sexual attention. Finally, it should go without saying that lewd language and behavior have no place in the lab, including any lab outings. If you notice someone being harassed, or are harassed yourself, tell the lab PI immediately. If the PI is the cause of your concern, then reach out to the department chair or another trusted faculty member in the department.

6.2 Scientific Integrity

Research Misconduct

The CDS lab and UTA is committed to ensuring research integrity, and we take a hard line on research misconduct. We will not tolerate fabrication, falsification, or plagiarism. A big problem is why people feel the need to engage in misconduct in the first place, and that's a discussion that we can have. If you are feeling pressured to succeed (publish a lot, publish in high impact journals), you should reach out to the PI (Amir) and we can talk about it – but this pressure is something we all face and is never an excuse to fabricate, falsify, or plagiarize.

Also, think about the goal of science and why you are here: you're here to arrive at the truth, to get as close as we can to facts about nature. Not only is research misconduct doing you a disservice, it's also a disservice to your own goals and to the scientific field. Also, it risks your entire career. It is never right and never worth it. Don't do it.

Reproducible Research

If you gave someone else your raw data, they should be able to reproduce your results exactly. This is critical, especially in a computational data science lab, because if they can't reproduce your results, it suggests that one (or both) of you has made errors in the analysis, and that your results are not trustworthy. Reproducible research is an essential part of science, and an expectation for all projects in the lab.

For results to be reproducible, the analysis pipeline must be organized and well documented. To meet these goals, you should take extensive notes on each step of your analysis pipeline. This means writing down how you did things every step of the way (and the order that you did things), from any pre-processing of the data, to running models, to statistical tests.

It's also worth mentioning that you should take detailed notes on your experimental design as well. Additionally, your code should also be commented, and commented clearly. We all know what it is like to sit down, quickly write a bunch of code to run an analysis without taking time to comment it, and then having no idea what we did a few months down the road. **Comment your code so that every step is understandable by an outsider**.

Finally, it is highly encouraged that you use some form of version control (e.g., Git in combination with GitHub) to keep track of what code changes you made and when you made them, as well as sharing code with others. The CDS lab's GitHub page is https://github.com/cdslaborg.

Reproducibility is related to **replicability**, which refers to whether your results can be obtained again with a different data set. That is, if someone ran your study again (with a different dataset or independent methodology), do they get the same results? If someone ran a conceptually similar study, do they get the same results? Science grows and builds on replicable results – one-off findings don't mean anything. Our goal is to produce research that is both reproducible and replicable.

Authorship

Authorship credit should reflect the individual's contribution to the study. An author is considered anyone involved with initial research design, data collection and analysis, manuscript drafting, and final approval. However, the following do not necessarily qualify for authorship: providing funding or resources, mentorship, or contributing research but not helping with the publication itself. The primary author assumes responsibility for the publication, making sure that the data are accurate, that all deserving authors have been credited, that all authors have given their approval to the final draft; and handles responses to inquiries after the manuscript is published.

At the start of a new project, the student or post-doc taking on the lead role can expect to be first author (talk to the PI about it if you aren't sure). The PI, Amir, will typically be the last author, unless the project is primarily under the guidance of another PI and Amir is involved as a secondary PI – then Amir will be second to last and the main PI will be last. Students and post-docs who help over the course of the project may be added to the author list depending on their contribution, and their placement will be discussed with all parties involved in the paper.

If a student or post-doc takes on a project but subsequently hands it off to another student or post-doc, they will most likely lose first-authorship to that student or post-doc, unless co-first-authorship is appropriate. All of these issues will be discussed openly, and you should feel free to bring them up if you are not sure of your authorship status or want to challenge it.

Recommendation Letters

Letters of recommendation are extremely important for getting new positions and grants. You can count on Amir to write you a letter if you have been in the lab at least one year (it's hard to really know someone if they have only been around for a few months). Exceptions can be made if students or post-docs are applying for fellowships shortly after starting in the lab.

If you need a letter, notify Amir as soon as possible with the deadline, your most-recent CV, and any relevant instructions for the contents of the letter. If the letter is for a grant, also include your specific aims. If the letter is for a faculty position, also include your research and teaching statements. In some cases, especially if short notice is given, you may also be asked to submit a draft of a letter, which will be modified based on Amir's experience with you, made more glamorous (people are often too humble about

themselves), and edited to add anything you left out that Amir thinks is important. This will ensure that the letter contains all the information you need, and that it is submitted on time.