

FOSE1025 — Scientific Computing

Week 10 Lecture 1: Ethics and Reproducibility

Diego Mollá

Department of Computer Science
Macquarie University

FOSE1025 2023H1

Programme

- 1 Ethics
- 2 Reproducibility

Reading

- These notes
- Readings listed in iLearn — Week 10

Programme

- 1 Ethics
 - Ethics in Science
 - Ethical Concerns with Data
- 2 Reproducibility

Programme

- 1 Ethics
 - Ethics in Science
 - Ethical Concerns with Data
- 2 Reproducibility


Ethics in Life

<https://www.scu.edu/mobi/resources-tools/blog-posts/ethics-in-life-and-business/ethics-in-life-and-business.html>



But What is Ethics?

<https://www.dictionary.com>

ethical [eth-i-kuhl] [SHOW IPA](#) 

SEE SYNONYMS FOR *ethical* ON THESAURUS.COM

adjective

- 1 pertaining to or dealing with morals or the principles of morality; pertaining to right and wrong in conduct.
- 2 being in accordance with the rules or standards for right conduct or practice, especially the standards of a profession:
It was not considered ethical for physicians to advertise.
- 3 (of drugs) sold only upon medical prescription.

Academic Integrity at Macquarie University

iLearn's Academic Integrity Module:

<https://ilearn.mq.edu.au/course/view.php?id=11590>



Ethics, integrity and policies at Macquarie University

<https://www.mq.edu.au/research/ethics-integrity-and-policies>

[STUDY](#)[RESEARCH](#)[PARTNER](#)[ABOUT](#)

Ethics, integrity and policies

RESEARCH

[Study](#)

Research

[Research Expertise](#)[PhD and research degrees](#)[Research centres, groups and facilities](#)[Research funding and grant opportunities](#)[Commercialisation and Innovation](#)[Research Collaboration and Partnerships](#)[Ethics, integrity and policies](#)

Ethics, Integrity, and Policies

Macquarie University is committed to maintaining the highest standards of research ethics and integrity, and has thus embedded a culture of professional and responsible research practice across all disciplines and faculties. The Research Ethics and Integrity teams support and promote responsible research practices through the provision of resources, guidance and education.



Research Integrity

At Macquarie, we aim to promote and support responsible research practices by providing resources and guidance to our researchers.



Human Research Ethics

All research with human participants must adhere to the National Statement on Ethical Conduct in Human Research (2007).



Animal Ethics

We are strongly committed to supporting the care and welfare of animals; and the ethical conduct of people who work with animals for scientific research purposes.

Key Ethical issues related to Scientific Research

Data Fabrication: Create or manipulate data to fit someone's purpose.

Plagiarism: Pretend that someone else's ideas are ours. Not acknowledging the source.

Impact in Society: Fail to consider the possibly negative impact of one's research.

Data Fabrication: The Case of Diederik Stapel (2004-2011)

https://en.wikipedia.org/wiki/Diederik_Stapel, accessed 4 May 2023

- Diederik Stapel is a Dutch former professor of social psychology at Tilburg University.
- During 2004-2011, he was involved in multiple cases of data fabrication.
- He manipulated data and fabricated entire experiments.
- It took several years to uncover the fraud because of several reasons:
 - The prestige of the researcher.
 - Insufficient clarity in the manuscripts as to how the data were collected.
 - Data used in the experiments were not made available.
 - As was usually the case in the field.

Different Forms of Fabricating Data

Besides the obvious fraud of creating data that does not exist, sometimes there is a fine line between fraud and poor research practice. For example:

- Manipulate the data so that it fits our expectations.
- **Cherrypicking** samples from our data.

<https://freshspectrum.com/toon-cherry-picking-data/>

This is not data analysis...



Plagiarism

- An essential part of science is to advance work made by others.



“If I have seen further than others, it is by standing upon the shoulders of giants”
(Isaac Newton)

- ... but the work by others needs to be acknowledged.
- Failing to acknowledge others' work can be plagiarism.
- Self-plagiarism is also plagiarism.

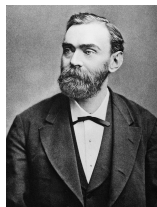
Impact in Society

- A common ethical problem with research and development in science and technology is not to stop and consider its possible (negative) impact in society.
- This is easier said than done. Often the implications of research in society are only known after the damage is done.
- Not everything is black and white: the impact can be both positive and negative.
- But stopping and thinking about these impacts can help improve the positive impact and diminish the negative impact.
- Several researchers have expressed regret for the research they undertook in their past.

The Case of Alfred Nobel

<https://www.britannica.com/biography/Alfred-Nobel>

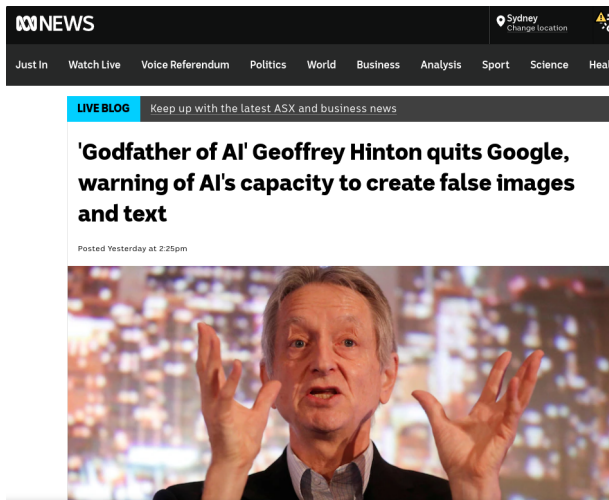
- Alfred Nobel invented the Dynamite.
- Dynamite was used for civilian use, e.g. mining . . .
- . . . But it also had military uses.
- It has been said that Alfred Nobel funded the Nobel prize as a means to prevent negative reputation because of his invention of dynamite.



(Alfred Nobel)

More Recently, in May 2023 . . .

<https://www.abc.net.au/news/2023-05-03/geoffrey-hinton-godfather-of-ai-quits-google-with-danger-warning/102297868>, viewed 4 May 2023



Programme

- 1 Ethics
 - Ethics in Science
 - Ethical Concerns with Data
- 2 Reproducibility

Key Ethical Issues with Data

<https://www.linkedin.com/learning/people-analytics/ethical-considerations>

Privacy: Make sure that private data keeps private.

Security: Protect your data, avoid unauthorised access.

Fairness: Avoid bias in data, avoid promoting bias when using the data.

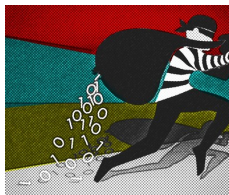
Disclosure and Consent: The users who provided the data need to know what you are using the data for, and give consent.

Privacy

- When data are collected from people, it needs to be **anonymised**.
- In anonymised data, references to private information are deleted or modified.
 - Names
 - Addresses
 - Passport numbers
 - ...
- Anonymisation is not easy: Currently there's so much data publicly available that sometimes it becomes possible to identify users and their habits even after the data have been anonymised.
 - Read, for example, this discussion about de-anonymisation of data from the Netflix prize:
<https://www.wired.com/2007/12/why-anonymous-data-sometimes-isnt/>

Security

- When gathering data with personal information, keep it secure.
 - Do not make it publicly available.
 - You may need to add password protection.
- Some organisations legislate **where** you can save the data.
 - When saving private data in the cloud, check the data policies of the provider of your cloud services.



<https://www.flickr.com/photos/111692634@N04/15855653380>

Fairness

- When data are collected, the data might be **biased**.
- That means that the data does not represent the real situation.
- E.g. collecting data from Twitter and assuming that Twitter users represent the entire population.
 - Only some kind of people use Twitter.
- Use of biased data might lead to:
 - Wrong business decisions (and lose money).
 - Being accused of bias or racism (and face a lawsuit).
<https://www.independent.co.uk/life-style/gadgets-and-tech/news/bing-image-search-microsoft-jews-racist-hitler-nazis-a8579596.html>

Consent of Data

<https://www.linkedin.com/learning/data-fluency-exploring-and-describing-data/data-ethics>

Regulations like GDPR (Europe — General Data Protection Regulation) and APP (Australian Privacy Principles) establish the need for consent when gathering data.

- **Consent:** When collecting data from people, consent from these people is needed.
- **Informed consent:** People need to know how the data will be used before they give consent.
 - Data collected for one purpose cannot be re-purposed.
- **Voluntary consent:** People have the choice to give consent or not.

Programme

1 Ethics

2 Reproducibility

- About Reproducibility
- Asking Questions for Help

Programme

1 Ethics

2 Reproducibility

- About Reproducibility
- Asking Questions for Help

Why Reproducibility?

- When you conduct science, you need to make sure that others can reproduce what you did.
 - If others can reproduce what you did, then your claims are more likely to be taken as valid.
- When you report a problem, you need to make sure that others can reproduce your problem.
 - Otherwise they may not be able to help you.
 - It is said that the first step to solve the problem is to be able to formulate the question.

Why Reproducibility?

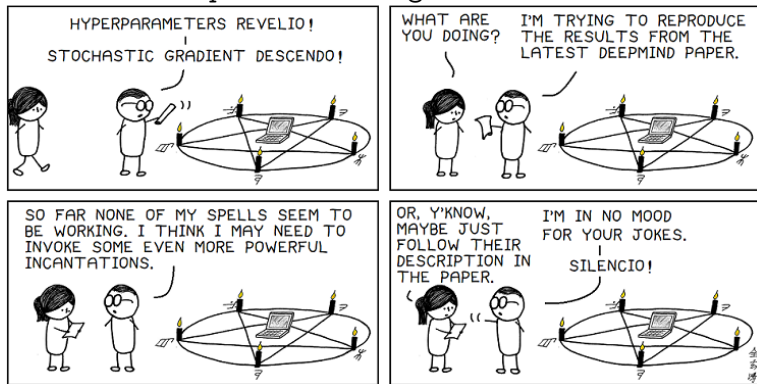
- When you conduct science, you need to make sure that others can reproduce what you did.
 - If others can reproduce what you did, then your claims are more likely to be taken as valid.
- When you report a problem, you need to make sure that others can reproduce your problem.
 - Otherwise they may not be able to help you.
 - It is said that the first step to solve the problem is to be able to formulate the question.

But What is Reproducibility?

- Basically, reproducibility means that someone else should be able to do the same as you did by following your instructions.
- When the experiments are performed with computers, there is some discussion/disagreement about what does “following your instructions” means:
 - 1 I can re-implement what you did after I read your report.
 - 2 I can run the code you wrote.
- The employability modules (“Achiever” and “Communicator”) touch item 1.
- In the lectures of this unit, we have focused on the use of scripts (MATLAB) to cover item 2.

Writing for Reproducibility

<https://abstrusegoose.com/588>



Failure to Replicate

https://www.sciencedirect.com/science/article/pii/S026240792200625X?via%3Dihub

https://www.newscientist.com/article/

mg25433810-400-the-replication-crisis-has-spread-through-science-can-it-be-fixed/

NEWSLETTERS

Sign up to read our regular email newsletters

NewScientist

News Podcasts Video Technology Space Physics Health More  Shop Courses Events

The replication crisis has spread through science – can it be fixed?

It started in psychology, but now findings in many scientific fields are proving impossible to replicate. Here's what researchers are doing to restore science's reputation



HUMANS 6 April 2022

By **Clare Wilson**

Programme

1 Ethics

2 Reproducibility

- About Reproducibility
- Asking Questions for Help

When do I Ask?

- Before asking a question, check if someone has asked it before!
- Know the most popular Q/A forums in your field.
 - For example, <https://stackoverflow.com/> is a very popular forum for asking questions related to programming.
- Know how to search the web for solutions to your problem.
 - For example, if you have a computer error, do a web search using the error message.

Is the Answer Somewhere in iLearn?

Follow these steps, in this order:

- 1 Look at the Announcement sections.
- 2 Look at the FAQ (if available).
- 3 Search the Discussion Forums (e.g. use the search box).

The screenshot shows the iLearn interface for the course FOSE1025 Scientific Computing. The top navigation bar is red and contains the Macquarie University logo, a 'Help Me' button, the user's name 'Diego Molla-Allod Student', and the 'iLearn' logo. Below the navigation bar, the course title 'FOSE1025 Scientific Computing' is displayed. A breadcrumb trail shows the path: 'My home / My units / FOSE1025_FHFYR_2020_ALL_U / General / General Discussion Forum'. A search bar is located on the right side of the breadcrumb trail, and it is highlighted with a red circle. Below the breadcrumb trail, the 'General Discussion Forum' section is visible, featuring a gear icon and a description: 'Check the General Discussion Forum on a regular basis for weekly communications by your tutor. This forum also allows you to post any questions you might have in relation to this unit.' On the right side, there is an 'Activities' sidebar with a list of links: 'Assignments', 'Dialogues', 'Forums', 'Quizzes', 'Resources', and 'Turnitin V2'.

How do I Ask?

From <https://stackoverflow.com/help/how-to-ask>

- 1 Choose the right title.
- 2 Describe your problem.
- 3 Help others reproduce your problem.
- 4 Proof-read before posting.

Choose the Right Title and Description

1. Choose the right title

- Pretend that you're talking to a busy colleague.
- Try to sum up your entire question in one sentence.
- Spelling, grammar, and punctuation are important.

2. Describe your problem

- Don't just say "it doesn't work".
- Explain how you encountered your problem.
- Explain what you did to try to solve the problem.
 - Otherwise people may give a solution that you have tried already.

3. Help Others Reproduce your Problem

- <https://stackoverflow.com/help/minimal-reproducible-example>
- Do not just post your entire program.
- Do specify the specific version of your program (and sometimes also the version of your operating system).
- Write **minimal code/instructions** that reproduces your problem.
 - Sometimes, when you are writing this minimal code, you may find the answer by yourself.
 - **Minimal**: Must be the smallest code possible that reproduces the problem.
 - But must be readable!
 - **Complete**: Must be detailed enough to reproduce the problem.
 - **Reproducible**: Test your code/instructions yourself to make sure that it reproduces the problem.

Take-home Messages

- Have a general awareness of ethical issues when conducting scientific research.
 - Data fabrication
 - Plagiarism
 - Impact in society
- Explain the key ethical issues related to data.
 - Privacy
 - Security
 - Fairness
 - Disclosure and Consent
- Write instructions that can be reproducible.
- Write error messages and reports that can be reproducible.
- Document your work so that someone else can reproduce it.

What's Next

- No lectures on Weeks 11-13.
 - There will be activities related to employability skills.
 - These will be listed in iLearn.
- There are SGTAs until week 12.
- In-class test 3 at the time of your scheduled SGTA of week 12.
 - 30% unit assessment weight.
- Wed 17 May: Submit the project.
 - 30% unit assessment weight.
- Reproducibility project:
 - 10% unit assessment weight.
 - Week 12: Submit phase 1 (reproducibility document, 7%).
 - Week 13: Submit phase 2 (peer assessment, 3%).