Problem Set 0

6.7930/HST.956

Due: Feb 10, 11:59pm on Canvas.

The CITI course required in this assignment takes several hours to complete and requires working through videos and associated quizzes, so be sure you allow sufficient time!

The purpose of this exercise is for you to get access to sensitive healthcare datasets. In it, we expect you to
1. Obtain CITI certification.
2. Apply for MIMIC access.

Once you complete both of these steps (details below), please zip your CITI completion report and submit it via Canvas. Do so before the deadline.

CITI certification

You will need to complete the CITI "Data or Specimens Only Research" course. If you have already taken the course through MIT and it does not expire before the end of the course (May 2025), you may skip these steps. **Note: completing the CITI course may take several hours.**

1. Register on the CITI site via the instructions on the page below:
   https://couhes.mit.edu/training-research-involving-human-subjects
   Note that there are mechanisms for both MIT and non-MIT personnel to make sure your CITI site is affiliated with MIT which is important in ensuring you can obtain access.
2. Register on the CITI program website, selecting "Massachusetts Institute of Technology (SSO)" as your organization affiliation (not "independent learner").
3. Select "View Courses" next to Massachusetts Institute of Technology.
4. If you are not able to add Massachusetts Institute of Technology as your affiliation, then add Massachusetts Institute of Technology Affiliates instead and select "View Courses" next to it.
5. Select "Add a Course", then answer questions (note that you will be taking IRB courses) :
   1. "Data or Specimens Only Research:"
   2. "I have not previously completed an approved Basic Course."
   3. "No"
   4. Questions 4 - 7: Skip (leave unanswered)
   5. If you added Massachusetts Institute of Technology Affiliates instead, select "No" for Questions 4, 5, and 7.

6. Complete the course and save a copy of your completion report. The completion report lists all modules completed, with dates and scores. *You should allow several hours to complete this course!*
7. You will submit this completion report twice: once to PhysioNet (for MIMIC access) and once to the course staff (via Canvas).

MIMIC Access

Please complete the steps required to get access to MIMIC data outlined here
https://mimic.physionet.org/gettingstarted/access/
Please use your MIT email address username as your PhysioNetWorks email (otherwise, please email us your PhysioNetWorks email). One of the steps will require you to fill out a Data Use Agreement (DUA) where you will be asked for:

1. Reference category. Choose Supervisor
2. A reference name. Write Peter Szolovits.
3. If asked for a phone number contact info, put 617-253-3476. If asked for an email, use psz@mit.edu. If asked for a job title, use "Professor"
4. In the research area box, provide at least **20 words** explaining the general research area for which the data will be used. (e.g. institution (*Massachusetts Institute of Technology*), course number (*HST.956 / 6.7930*), title (*Machine Learning for Healthcare*), term (*Spring 2025*), dataset (*MIMIC-III, MIMIC-IV*), etc.)
5. You *must* ensure your form is complete, or it will not be approved.

Please review the following regarding responsible use of large language models with MIMIC data:
https://physionet.org/news/post/gpt-responsible-use

When you are asked to submit your CITI course results, you must submit the full CITI completion report (which lists all modules completed, with dates and scores) rather than just the CITI certificate, which shows only course completion.