# Testing Discussion

CS 485/698: AI-Assisted SE

# Today's Agenda

- Team meeting (~10 minutes)
  - Sprint retro for P5:
    - Has your testing so far uncovered any serious bugs?
    - How do you know that your tests are actually testing the right things?
- Discussion of red team/blue team testing exercise (~10 minutes)
- A5 discussion (rest of class)

# Test Suite Quality Discussion

- Offensive teams: what was the best bug-finding strategy?
- Defensive teams: how useful were coverage metrics? Mutation testing?
- Do we think that automated test suite quality metrics will ever allow us to be confident that code is well-tested without actually reading the test cases ourselves? Why or why not?
- Other interesting observations?

# A5 discussion

- On the next slide, there is a list of groups (one group per row). Sit with your group near the big number
- Each group has ***at most one*** person who picked each reflection question
- We'll discuss the questions in order:
  - Discuss with your group for ~5 minutes
  - Share anything particularly interesting with the whole class
    - Priority: anything that will help other teams with P5

# A5 discussion: groups

| | | | | | | |
|---|---|---|---|---|---|---|
| **Group 1:** | Dhyani Soni | Aiden Barrera | Brandon Howe | Eric Perez | Justin Carreno | Swechcha Ambati |
| **Group 2:** | Krishi Shah | Tirell Spence | Paulo Bellame | Mark Youssef | Engy Masoud | Thomas Kolb |
| **Group 3:** | Saanvi Elaty | James Marciano | Declan Blanchard | Balaji Shashipreeth, Racherla | Luke Hill | |
| **Group 4:** | Jossie Zamora | James Mullins | Elvis Valcarcel | Isabel Patrisso | Haroon Aftab | Avanish Kulkarni |
| **Group 5:** | Fardeen Iqbal | Ashraf Aldekaim | Youssef Masoud | Allen Cabrera | Vishesh Raju | |
| **Group 6:** | Nafisa Ahmed | Jonathan Martinez | Aryan Modi | Xun Song | Roaa Elsayed | |
| **Group 7:** | Marcus Hilario | Salma Ghazi | Victor Jimenez | Zhirong Zhang | Alexander Tochtchev | |

# Question 1

How can you verify that the tests generated by the LLM are sufficient? What do you need in order to feel confident in the test cases (and why)? When do you stop testing?

# Question 2

Do you believe it is effective to use AI to test code generated by an AI? If hallucinations can happen with the original writing of the code, will that be compounded by the writing of the tests? How involved do you think human software developers need to be and why?

# Question 3

How do you get the LLM to keep tests up to date as you make changes to the program code? What would you add to your prompt to ensure the code and tests are updated together? Test this prompt and tell us how it went. Make a change to that prompt that results in an improvement and tell us how it went.

# Question 4

Imagine that you are the VP of Engineering for a software startup. You need to define the company's policy on testing. Write the policy on how your software developers should approach automated testing, when to test, how much to test, how many tests should be written, how to involve AI, etc. Explain the rationale for your policy to the CEO. Remember that you are responsible for reliably managing your customers' data, so they don't sue your fledgling company out of business.

# Question 5

Are LLMs better at generating some kinds of tests (e.g., unit tests, integration tests, ...) than others? Why or why not?

# Question 6

Which test suite quality metrics (if any) are good proxies for determining whether LLM-generated tests are good enough? Why or why not?

# Wrapup and Reminders

- P5 due Sunday
- Graduate students: start thinking about A7 topics
  - sign up is now available (grad students only)
- Undergrads: A7 will be an extra credit assignment. Once the grad students have picked their topics, I will open it to you
  - Grad students: pick a topic by April 8