[Mini Research Challenge Due 9/4](https://brightspace.vanderbilt.edu/d2l/le/news/442017/243997/view)

"Use Generative AI to produce 100 Python coding problems that compose an automated assessment that can be used to measure an aspect of Generative AI coding abilities."

Your output needs to be produced by ChatGPT (or Code Interpreter), Bard, or Claude. You can, but do not have to, use the API to do the generation.

Your output needs to be something that can be run as an automated benchmark that produces data that can be automatically analyzed (e.g., X did better than Y on the coding problems in these cases). Think about tagging, keywords, concepts, etc. to enhance your problems to make them more valuable.

What unique aspect of coding ability are you assessing? The ability to solve LeetCode problems has been thoroughly explored. You must do something more interesting than producing 100 coding problems with tests. What is new and different about your problems that will give us new insights? How will you create data that has a known ground truth label for each problem? Does ground truth even matter?

You will complete this project individually, although you will meet on 8/30 with a partner to brainstorm and discuss approaches.

Put all of your work into a GitHub repo with an excellent README.md file that describes in detail how to run your coding benchmark, what is new and novel about it, and what it assesses.

When you are done, turn it in here: <https://forms.gle/yJBRtPLJg1wTewQi9>