Machine Learning in Production
Midterm 1, Spring 2025

Claire Le Goues and Austin Henley

Name: _____

Andrew ID: _____

Instructions:

- Including this cover sheet and the scenario, your exam should have **8** pages. Make sure you are not missing any pages. *You may detach the last page and recycle it after the exam.*
- All questions in this midterm refer to the scenario on Page 8. Answers are graded in the context of the scenario; **generic answers that do not relate to the scenario will not receive full credit.**
- The exam has a maximum score of **65** points. The point value of each problem is indicated. We designed the exam anticipating approximately one minute per point.
- **Please write legibly**. We are unlikely to be able to grade your solution if we can't read it.
- We give an amount of space commensurate with what we expect you to need for each question. We use horizontal lines to suggest where to not use the full page. You may exceed those limits if it is clear where to find the rest of your answer. However, we strongly recommend writing concise, careful answers; short and specific is much better than long, vague, or rambling. However, **do NOT write anything you want us to grade on the back of pages.** We will scan the exam and will not look at the back sides.
- This is a **closed book exam**; no books or electronics allowed. You may refer to 6 sheets of notes (handwritten or typed, both sides).

Question 1: Goals and Telemetry [15 points]

All questions in this exam relate to the scenario on the last page. You may detach the last page if you like. Your first task is to explore and document goals for the CMU Student Copilot project and identify how you can measure success.

(a) [6 points] Give one plausible *user goal*; one plausible *organizational goal*; and a plausible *key performance indicator* for that organizational goal for the CMU Student Copilot project that you are trying to support.

User goal:

Organizational goal:

Key Performance Indicator:

(b) [9 points] You plan to evaluate how the model for Scholarly Activity Detection does *in production*. In particular, you would like to see *how often the model detects the student is in class when they actually are not.* Design a measure and suggest what data to collect and how to operationalize the measure with telemetry. The measure can be an approximation, but must be plausible within the realism of the scenario.

Measure:

Data to collect (what and how):

Operationalization:

Question 2: Model and Data Quality [15 points]

(a) [5 points] Early experiments with machine learning for Scholarly Activity Detection create results that seem too good to be true. Name one pitfall in model accuracy evaluation that could cause overly positive accuracy results in evaluation in this scenario and describe how you would detect or avoid the problem. Your answer must demonstrate an understanding of the pitfall and relate to the scenario.

(b) [5 points] You consider whether the CMU Student Copilot could be improved by adding additional *testing.* Describe one specific test, either for the data or monitoring of the system, that may help us identify issues with the Scholarly Activity Detection feature.

Type of test: ☐ *data test* ☐ *monitoring test*

(c) [5 points] Provide a plausible concrete example of *concept drift* in the scenario (i.e., changes in decision boundaries, not in data distributions) that may degrade the accuracy of your Scholarly Activity Detection model in production over time.

Question 3: Trade-offs [12 points]

You are considering how to deploy the models of the CMU Student Copilot project (see the scenario on the last page), whether on the device, in the cloud, or on the student's laptop.

(a) [6 points] Identify and roughly rank two qualities that are important for the decision in this scenario and one quality of little importance (no measure required for any of them). Provide a brief justification of why they are important or not important:

Quality 1 (most important):

Quality 2 (second most important)**:**

Quality 3 (low importance):

Justification:

(writing below this line is allowed but discouraged)

(b) [6 points] Make a recommendation with a brief justification of how/where to deploy the model for Scholarly Activity Detection, considering the tradeoffs between the qualities. Refer explicitly to the important qualities identified previously and underline them in your text. Your answer must relate to the scenario. If you are missing information to make that decision, describe what information you would need and how you would make a recommendation with it.

(writing below this line is allowed but discouraged)

Question 4: Risks and Mitigation [18 points]

To plan for mistakes you try to better understand the requirements and risks of the product (see the scenario on the last page). For this question, you focus on the following requirement:

> *"The voice assistant feature should not activate, for any reason, during lectures, exams, or recitations."*

(a) [3 points] Classify the following parts of the CMU Student Copilot system into world and machine entities (in the world vs machine sense from the reading and lecture):

- The voice of the student ☐ *world entity* ☐ *machine entity*
- The microphone sensor in the device ☐ *world entity* ☐ *machine entity*
- The historical data of recent verbal utterances ☐ *world entity* ☐ *machine entity*
- The connection between the device and wifi ☐ *world entity* ☐ *machine entity*
- The corpus of audio training data ☐ *world entity* ☐ *machine entity*
- The actual location of the student ☐ *world entity* ☐ *machine entity*

(b) [3 points] State one **software specification** that is necessary for the system to satisfy the above requirement.

(c) [4 points] State one **environmental assumption** that might be assumed in the system design to satisfy the above requirement, *but that may not actually be true* and may hence lead to failures to meet the requirement in the running system.

(writing below this line is allowed but discouraged)

(d) [8 points] If the model incorrectly predicts that the student is *not* in a lecture when they are, the requirement above may be violated. This would correspond to paths in a fault tree (you do not need to draw the tree). Describe a *mitigation* to make it less likely that the requirement will be violated even if the model prediction is wrong. *The mitigation should be at the system level, outside the ML component* (i.e., not just "train a more accurate model" or "use an ensemble model"). In addition, check the box corresponding to whether your mitigation would eliminate a basic event from the fault tree or add another basic event (with an AND or OR connection).

Mitigation description:

Update to a fault tree (check the option corresponding to your mitigation below, no further explanation needed):

- ☐ *eliminate basic event*
- ☐ *add basic event with an AND connection*
- ☐ *add basic event with an OR connection*

Question 5: Process [5 points]

Your team is discussing how to be more *agile* and adopt better engineering processes. They recently took *Machine Learning in Production,* and want to apply what they learned.

(1) [2 points] Briefly describe the rationale for using Git for this project. Describe a problem that might occur if your team decided to not use version control, such as Git.

(2) [3 points] Your team would like to use Docker as part of their production pipeline. What is either a benefit or a drawback of containerization in this context?

Type of answer: ☐ *benefit* ☐ *drawback*

Scenario: CMU Student Copilot

(The scenario is fictional. You may detach this page from the exam.)

CMU researchers have designed the *CMU Student Copilot* device. Each and every student enrolled at CMU is given one to maximize their student experience. The device is clipped to your shirt or jacket, and is always watching and listening as you engage in scholarly activities. It goes where you go. It listens to your lectures, watches you complete your assignments, knows your schedule, analyzes your emails, watches Canvas for updates, etc. All of this information is used to give feedback, help, and suggestions entirely custom for you. It is the ideal AI assistant for students.

Features include:

- An always-on camera and microphone that can follow along with lectures, see your computer screen, read documents and books in front of you, etc.
- An AI voice assistant that continuously checks all of your university-related accounts for information (e.g., Canvas, GitHub, Slack, email, Piazza, etc.) combined with Google search and OpenAI LLMs
- Detects what scholarly activity the student is engaged in (e.g., doing homework or sitting in lecture) using a proprietary *Scholarly Activity Detection* model and will provide assistance only when appropriate (e.g., no help on exams and disable voice commands during lectures)
- Automatically connects to university Wifi and tracks location using GPS
- The backend system is hosted on servers in a vault underneath CMU where your personal data is stored using state of the art encryption
- Long battery life with free swappable batteries found in most campus buildings

You have been recruited by President Farnam to join the CMU Student Copilot team to help take the project to the next level. Given the large investment by CMU into the project and the initial positive feedback, the university would like to find ways to monetize it. The team has a strict budget for paying developers and running servers for training and deploying models.