Machine Learning in Production
Midterm 2, Fall 2024

Christian Kaestner and Sherry Wu

Name: _____

Andrew ID: _____

Instructions:

- Including this cover sheet and the scenario, your exam should have **7** 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 7**. 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 **50** 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: Scaling and Operations [19 points]

All questions in this exam relate to the scenario on the last page. You may detach the last page if you like.

(a) [6 points] The typical input to the model (drone photos and other data) is usually around 30 megabytes, whereas the model itself is very large (about 500 gigabytes). The model is currently deployed as a service (written with Flask) that responds with analysis results immediately when a farmer requests the analysis for provided drone footage. You are considering whether to switch to batch processing, stream processing, or the lambda architecture. Within the realism of the scenario, please recommend one approach, and justify your choice. Underline the relevant qualities in your justification; your justification must demonstrate an understanding of the tradeoffs involved.

☐ Batch processing ☐ Stream processing ☐ Lambda architecture ☐ Keep as service

Justification:

(b) [4 points] Image data in the training dataset is never changed, only added. However, other training data about farm locations and crop yields sometimes need to be corrected, when the initial data reported to the government was not correct. This is tabular data of which less than 0.1% per year is corrected. Which of the following versioning strategies is the most *space efficient* for the tabular data in this scenario:

☐ Keeping copies ☐ Tracking deltas ☐ Recording offsets

Explain your answer:

(c) [4 points] You occasionally will need to update the model with new data or will want to experiment with different model versions in production. Therefore, you want to track which prediction was made with which version of the model, and what data the model was trained on. Explain how *MLFlow* (and the many similar tools) can help with this and what additional information needs to be tracked outside of MLFlow:

(d) [5 points] Deploying and updating a large foundation model can be tricky and the amount of MLOps tooling out there can be overwhelming. Provide an example of *prudent and deliberate technical debt* related to deploying the machine learning model that would be plausible in the context of the scenario. Your answer must explain why it is prudent and deliberate.

Example:

Why prudent:

Why deliberate:

(writing below this line is allowed but discouraged)

Question 2: Fairness [12 points]

While you do not handle information about people, you are still worried about whether the software supports small family farms and very large farms equally. Therefore you analyze the predictions your system does grouped by farm size.

(a) [4 points] Making more mistakes when predicting fertilizer needs for small farms than large farms creates:

☐ a harm of allocation ☐ a harm of representation ☐ both ☐ neither

Brief explanation:

(b) [6 points] Consider the source of the bias that could lead to lower accuracy for small farms. Select two of the following sources of biases discussed in class – *tainted labels; historical bias; skewed sample; limited features; proxies* – that may be responsible. For each selected source, briefly explain how it might have led to the different model behaviors for small vs. large farms in this scenario.

Select #1: ☐ *tainted labels* ☐ *historical bias* ☐ *skewed sample* ☐ *limited features* ☐ *proxies*

Explanation:

Select #2: ☐ *Tainted labels* ☐ *historical bias* ☐ *skewed sample* ☐ *limited features* ☐ *proxies*

Explanation:

(c) [2 points] Since the system is cloud-based you know what predictions you made for each customer. You have access to the last 1000 predictions for small farms and the last 5000 predictions for big farms. You can see how often your model identified crop stress indicators and how often your model recommended fertilizer. However, you do not have direct information about which predictions were good and feedback on what customers actually did is delayed and not yet available. Is there any notion of fairness that you could measure with this information you have now? (No explanation required)

☐ Anti-classification ☐ Group fairness ☐ Equalized odds ☐ None of these

Question 3: Explainability/Transparency [9 points]

In the scenario, farmers would benefit from explanations for better understanding when to trust the model suggestions to make informed decisions about fertilizer application.

(a) [5 points] Name a technique to provide *explanations of individual predictions* that could be used by a data scientist on your team to understand why the model made a given (faulty) prediction and explain briefly how the technique works (rough intuition is sufficient).

Technique's name:

Basic idea of how the technique works:

(b) [4 points] Would this method suggested in part (a) provide explanations that are appropriate for farmers in the scenario? Argue why/why not and justify your answer.

Question 4: Safety and Security [10 points]

Nobody in the team had previously thought about security and safety. You want to be responsible and consider common problems before they occur.

(a) [6 points] Within the context of the scenario, consider a potential *targeted poisoning attack* that an adversary (e.g., competitor or terrorist) could attempt and how the product could mitigate such attack. Answer the four prompts below. Your answer must be reasonably realistic in the scenario and demonstrate an understanding of targeted poisoning attacks.

Attacker goal (intended outcomes):

Which security property does the attack intends to undermine:

☐ Confidentiality ☐ Integrity ☐ Availability

Attack method (what the attacker would do):

Mitigation strategy (how the developers can prevent the attack/make it harder):

(b) [4 points] After some testing, you don't believe that you can ensure high robustness for the model before the initial release, but it may be possible to make the system safe regardless. Explain with an example relevant to the scenario how a system can have an unreliable component (i.e. model not robust) but still be safe. Your answer must demonstrate an understanding of the difference between safety and reliability.

Scenario: Precision Agriculture with Drones

You've just joined the *AgriVision AI* team at a well-resourced global agricultural organization, akin to John Deere or Bayer's Crop Science division. The team operates as an independent, startup-like unit within the company, focused on rapid innovation in AI and drone technology. AgriVision AI's mission is ambitious: to empower environmentally friendly farming practices, lower costs, improve soil quality, and reduce CO_2 emissions. The platform achieves this by identifying early signs of crop stress and reducing the use of fertilizers and pesticides through targeted, as-needed application – this is known as *precision agriculture*. The goal is to support both small-scale organic farmers and large commercial agricultural producers, fostering sustainability across the agricultural spectrum.

The team has already demonstrated feasibility of the approach, filed patents, and published several papers. By leveraging publicly available datasets (high-resolution satellite imagery from the USDA's National Agricultural Imagery Program, crop and soil health records through USDA public reports), data from customers, and partnerships with agriculture programs at universities, the team trained a powerful vision-based foundation model (like an LLM, but for images). The model is capable of detecting crop stress indicators like discoloration, uneven growth, and pest damage in aerial photos. When given a series of drone images (both historical and recent), alongside farm-specific data such as crop types, past soil test results, and yield history, the model predicts precise localized interventions for improving crop health and optimizing resource use. For example, after analyzing drone imagery and soil data, the system might recommend targeted actions like planting nitrogen-fixing crops in areas with depleted nitrogen or focusing irrigation efforts on moisture-deficient zones. Model evaluations are very promising. The company is excited about the *AgriVision AI* project, because it ties in with other drone projects by the parent organization for precisely applying fertilizer with drones.

Your current goal is to transform this promising research prototype into a scalable product. The business model is to lease drones and offer a subscription-based analysis service, where farmers upload drone imagery and receive actionable recommendations through a cloud-based platform. The system will run entirely in the cloud. The model will be regularly retrained based on more recent government data and based on data provided by customers (drone images and explicit feedback about whether they followed the suggested actions)

With its strong environmental mission and significant funding, the team has attracted several strong AI researchers who are experienced with state of the art vision and foundation models. However, embedded in a company rooted in hardware and science, the team lacks extensive software engineering expertise. Innovation and rapid prototyping have been prioritized over building robust, long-term infrastructure. The company has little prior experience with cloud-based subscription models for software.