Practice Midterm

Christian Kaestner and Eunsuk Kang

Name: _____

Andrew ID: _____

Instructions:

- Not including this cover sheet and the appendix, your exam should have _ pages. Make sure you're not missing any pages. **Write your Andrew ID on every page.**
- The exam has a maximum score of _ points. The point value of each problem is indicated. We allocated approximately one point per minute.
- Clearly indicate and write your answers in the space provided for each problem. We cannot give you points for answers we cannot find or read. Write concise, careful answers; short and specific is much better than long, vague, or rambling.
- We give an amount of space commensurate with what we expect you to need for each question. If you need more space, use the last page and make it obvious in your first answer where to find the rest of the answer. Do NOT write answers on the back of pages (we scan the exam for grading and won't see those).
- You may NOT use calculators, cell phones, laptops, or any other electronic or wireless devices, nor consult with your friends or colleagues. **You may consult handwritten or printed notes and books**.
- Good luck!

Scenario

For the practice midterm, consider the following scenario (or any other scenario discussed in class):

You work at a large established social media company, as a software engineer on a video chat system. Concerned by recent media reports about inappropriate behavior on the platform, especially targeted at children, you work on a system to detect the age of participants and to detect inappropriate content (especially nudity and graphic violence).

The chat application is linked to the social media platform and allows text and video chats between two and up to 10 participants. In the current system, video is exchanged between participants directly (peer to peer) and not stored.

You work with a group of data analysts, researchers trained in machine learning. They have developed a prototype with which they attempt to identify participants and map them to their social media profile -- which is linked to their age. In addition, they built two classifiers to detect nudity and violence in still images. In a separate project, the data analysts have created a model to check whether the behavior of a user on the platform is consistent with the stated age in the user's profile. They assure you that all of these models have achieved pretty good accuracy on their held-out evaluation data.

Question: Qualities

For the detector that checks whether the age in the profile (likely) corresponds to the user's real age, the data analytics team is developing more and more sophisticated models. They use all kinds of data sources, including past interactions with the social media site (postings, viewing, ...), current and past profile pictures, geo-location on the users' phones and IP addresses, tracking cookies on other sites, and so forth.

[8 points] You are getting a bit worried having to operate and maintain that system. Name two engineering concerns that the data analysis team may not be aware of that will be relevant for you and your team and justify why that concern is important:

- Concern 1:

- Justification

- Concern 2:

- Justification:

[10 points] Beyond accuracy of the models, name three qualities that are relevant in this specific scenario and each provide a description of how you would measure that quality:

- Quality 1

- Why relevant:

- How to measure

- Quality 2
- Why relevant:
- How to measure

- Quality 3
- Why relevant:
- How to measure

Question: Architecture

You consider two options:

- Design A: You perform the recognition of participants and of nudity and violence on the smartphones of the participating users
- Design B: You perform the recognition on your own cloud infrastructure, but to do this you give up on the peer to peer communication and instead route all video traffic through your cloud.

[4 points] Identify two qualities for which design A is better and justify why it is better in that design. Make sure the answer and justification is grounded in the scenario.

[4 points] Identify two qualities for which design B is better and justify why it is better in that design. Make sure the answer and justification is grounded in the scenario.

[4 points] In the next team meeting you need to argue for one of the two designs. Which ones would you argue for and why? Make sure the answer and justification is grounded in the scenario.

Question: Model Quality

You plan to integrate the model that detects violence in still images in the infrastructure. You would like to independently assess its model quality.

- [2 points] Give an example of a false positive:

- [2 points] Give an example of a false negative:

- [4 points] In the scenario, is recall or precision more important? Justify your answer.

[5 points] Models can make mistakes in often unpredictable ways. Briefly describe how you would use the model's result in the product and how you handle possible mistakes:

[2 points] You plan to design telemetry to be able to measure the model quality *in production*. In the context of the scenario, suggest a realistic way to assess model quality:

- Describe what data you would gather:

- Describe how you would determine model quality with that data:

- How do you avoid that your approach collects an overwhelmingly large amount of data?

Question: A/B testing

You want to test the system before you deploy it. You decide to test it in production with an A/B experiment. Specifically, you want to see whether users that use the new system will spend significantly less time chatting than those without the system. For your A/B test, you need to decide for how many users you want to activate the system.

[3 points] In this specific scenario, what's the risk of using too many users?

[3 points] In this specific scenario, what's the risk of using too few users?

Question: Testing

[6 points] After observing some issues with models provided, you plan to test the entire pipeline used to learn the models.

Name two characteristics of good unit tests:

Suggest and briefly explain a measure how you can measure adequacy of your test suite:

Question: Measurement [6p]

[6 points] Give an example for each of the following types of measurements, and state their measurement scales (i.e., nominal, ordinal, interval, and ratio):

- Organizational objectives:

- Leading indicators:

- User outcomes:

Question: Requirements and Risks [20p]

Consider the following requirement of the system: *Users under the age of 18 should never be exposed to content depicting violence or nudity on this video chat platform.*

[2 points] What are environmental entities and software components in this system?

[5 points] What are environmental assumptions and software specifications that are necessary to establish the above requirement?

[10 points] Construct a fault tree that describes how the system may fail to satisfy the above requirement (i.e., the top event in the tree should be “Underage users are exposed to inappropriate content”).

[3 points] Identify the minimal cut set(s) for the fault tree that you constructed in (3). Suggest a design recommendation to reduce the risk of failure for ONE of the minimal cut sets.