Name: Kerui Liu

Andrew ID: keruil

Course: Machine Learning in Production

Assignment 2

Stakeholders

Direct Stakeholders

1. Dashcam company (manufacturer)
2. Dashcam end users (drivers/vehicle owners)
3. Automotive manufacturers integrating dashcams
4. Non-profit child-safety organization

Indirect Stakeholders

5. Missing children and their families
6. Law enforcement authorities
7. Bystanders captured on dashcam video (privacy)
8. Insurance companies (use of dashcam data in claims)
9. Regulators (data protection, safety, consumer rights)
10. Technology contractors (AI vendor, cloud providers)

Goals

Organizational Goals (company)

Goal: Grow share of market using differentiated product with safety features.

Measure: (M) overall market share of following sectors; (D) quarterly sales report, industry, (O) complete market share growth percentages of the company vs. the company's top 5 competitors based on reported sales!

System Goals (dashcam platform)

Goal: Facilitate prompt and dependable dashcam video retrieval for missin children.

Measure: (M) average response time from query request to report readiness; (D) the system's distinguishing queries when submitted and timestamps are represented in reports; (O) median response time is calculated at a single query level.

User Goals

Drivers (end users): Safeguard privacy while providing information for searches.

Measure: (M) Number of opt-outs vs. opt-ins; (D) app configuration logs; (O) compute percentage of users consenting to participate.

Law Enforcement: Receive accurate leads on sightings.

Measure: (M) True positive vs. false positive sightings; (D) cross-check sightings with confirmed ground reports; (O) compute precision (%) of reported matches.

Families of Missing Children: Obtain faster location reports.

Measure: (M) Time from report issued to first sighting received; (D) alert and log data; (O) average response time per case.

Model Goals

Goal: Ensure strong recognition accuracy in diverse conditions (low light and distortion).

Measure: (M) Recognition accuracy; (D) benchmark dataset of varied real-world dashcam clips; (O) compute F1-score across conditions (day/night, clear/blurred).

Relations: Higher model accuracy → improves precision of sightings → raises law enforcement utility → increases family trust → improves user willingness to opt in → supports organizational differentiation.

Hazard Analysis

In order to predict and reduce risks, I utilized a rigorous hazard analysis process using a system-theoretic approach called STPA (System-Theoretic Process Analysis). The process began by identifying three important and distinct stakeholders: drivers (end users), law enforcement, and families of missing children. For each stakeholder, I described how these stakeholders use the system in their values or goals and listed the losses they may experience. I then converted these into direct risks or risks of a system, then into system requirements—driving requirements meaning on the right side of application. This approach allows a proactive and technically functional system to engage with the human factors and organizational issues present in implementation.

Step 1: Identify the values or goals of stakeholders.

Drivers, as the end-users, value privacy, safety, cost, and trust in the company. Law enforcement values data accuracy, timeliness to investigation, and manageable workloads. The families of missing children value rapid recovery, effective communication, and fairness in the outcome of detection.

Step 2: Identify potential losses.

Losses are directly correlated to violations of the values. For example, the drivers may lose control over their personal data when there is indiscriminate uploading of video. Law enforcement may experience wasted resources when false positive usage of a significant volume of video overwhelms the desks of investigators. The family of missing children, when mistaken for someone, goes emotional harm—the false hope generated from incorrect details presented as sightings.

Step 3: Convert losses into risks and requirements.

Through analyzing the losses, I identified specific risks and wrote requirements to mitigate them. This conversion turns the abstract values into tangible design constraints for the system.

Risks and Requirements by Stakeholder

Drivers (End-Users):

Risk: Unauthorized access to stored video.

Requirement: All local video storage must be end-to-end encrypted.

Risk: Excessive mobile data usage resulting in unexpected charges.

Requirement: Upload only brief, relevant snippets of video.

Risk: A false match generates a suspicion and harassment.

Requirement: A human authority must verify matches before releasing them.

Risk: Battery drain from processing.

Requirement: Implement adaptive processing modes to lessen the load when on battery.

Risk: Loss of trust in the manufacturer.

Requirement: Transparent opt-in consent and communication on relevant use of videos.

Law Enforcement

Risk: Too many false positives wasting time.

Requirement: Introduce thresholds and a verification process before reporting.

Risk: Missed detections resulting in lost opportunities.

Requirement: Retrain and validate the models regularly with new data.

Risk: Report delays making reports less useful.

Requirement: Have a reporting process that is fast and prioritized.

Risk: Evidence dismissed in court as inadmissible.

Requirement: Maintain tamper-proof logs and metadata to uphold the chain of custody.

Risk: Information overload.

Requirement: Provide contextual filters by time and location to narrow results.

Families of Missing Children

Risk: Failure to detect can delay recovery efforts.

Requirement: Create efficient pipelines for fast detection and reporting.

Risk: Misreports provide false hope.

Requirement: Authorities should verify reports before notifying families.

Risk: Demographic bias leads to ineffective results.

Requirement: Conduct fairness assessments and modify thresholds.

Risk: Lack of participation can decrease coverage.

Requirement: Increase participation in the dashcam network through collaboration.

Risk: Poor communication between families and agencies.

Requirement: Provide clear communication mechanisms with designated contacts.

The analysis determined that there are many interdependencies across requirements. For example, enhancing model fairness is beneficial for both law enforcement (less missed detections) and for families (equity of outcomes). The same is true for encryption and selective uploading that would serve drivers' privacy issues as well as law enforcement's need for solid evidence. It forms a base of further requirement decomposition.

Requirement Decomposition

From the requirements suited for analysis during hazard analysis, I have selected the following requirement for further analysis:

REQ: The system only uploads video snippets for active searches, protecting user privacy, and limits unnecessary data uploads.

The requirement is designed to address user-centric privacy concerns and costs while allowing law enforcement access only to relevant evidenced. Estimating the requirement will require outlining assumptions about the environment and the software specifications required to guarantee this behavior.

Environmental assumptions (ASM):

Connectivity: Dashcam users only occasionally connect their devices to the Internet through smartphone tethering, car Wi-Fi, or hotspot.

Accurate search parameters: Authorities provide correct, and accurate search parameters (ex. facial image of child, time-frame, geography).

Local capacity: Dashcams have adequate local storage to buffer video until relevant pieces can be filtered.

User consent: Users "opt-in" explicitly and consented to the upload of relevant portions of their data.

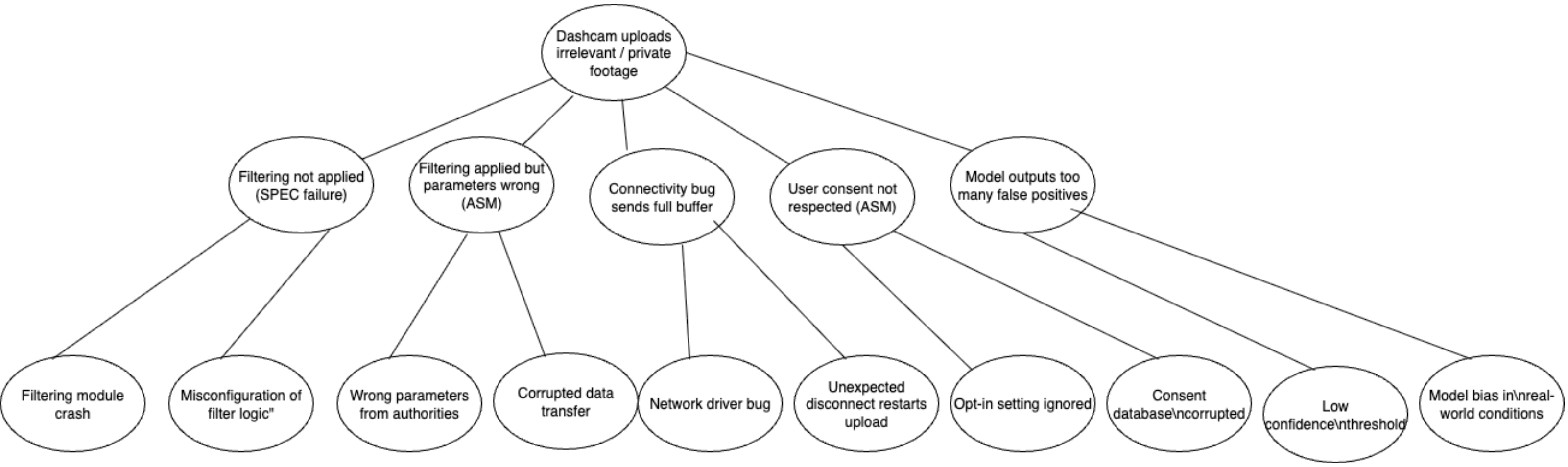
Stable power source: Cars provide a consistent power supply, creating a low risk of interruption during processing.

Software Specifications (SPEC):

The onboard software for the dashcam should utilize local filtering logic which contrasts recorded video to parameters defined by the user. Only videos that exceed basic confidence thresholds and were recorded in a specific timeframe or place should be composed and queued for upload to a server. Videos that do not meet those thresholds would remain stored and unavailable for access to other systems.

By merging these assumptions and specifications, data security under the requirements can be achieved. The assumptions define the necessary conditions in the world for the software to perform as necessary (e.g., consent and connectivity), while the specification defines what the software should do with those inputs (e.g., filtering and uploading selectively). If the system did not meet one of the assumptions or specifications, then the requirement could be violated, which is a good reason to think through potential faults and mitigation in subsequent steps.

Fault Tree (Risk Analysis)


Dashcam uploads irrelevant / private footage
Filtering not applied (SPEC failure)
Filtering applied but parameters wrong (ASM)
Connectivity bug sends full buffer
User consent not respected (ASM)
Model outputs too many false positives
Filtering module crash
Misconfiguration of filter logic"
Wrong parameters from authorities
Corrupted data transfer
Network driver bug
Unexpected disconnect restarts upload
Opt-in setting ignored
Consent database\ncorrupted
Low confidence\nthreshold
Model bias in\nreal-world conditions

Mitigations

There are two system-level approaches to lessen the chance of dashcams uploading privacy invasive footage or footage unrelated to driving.

Mitigation 1: Independent Pre-Upload Verification

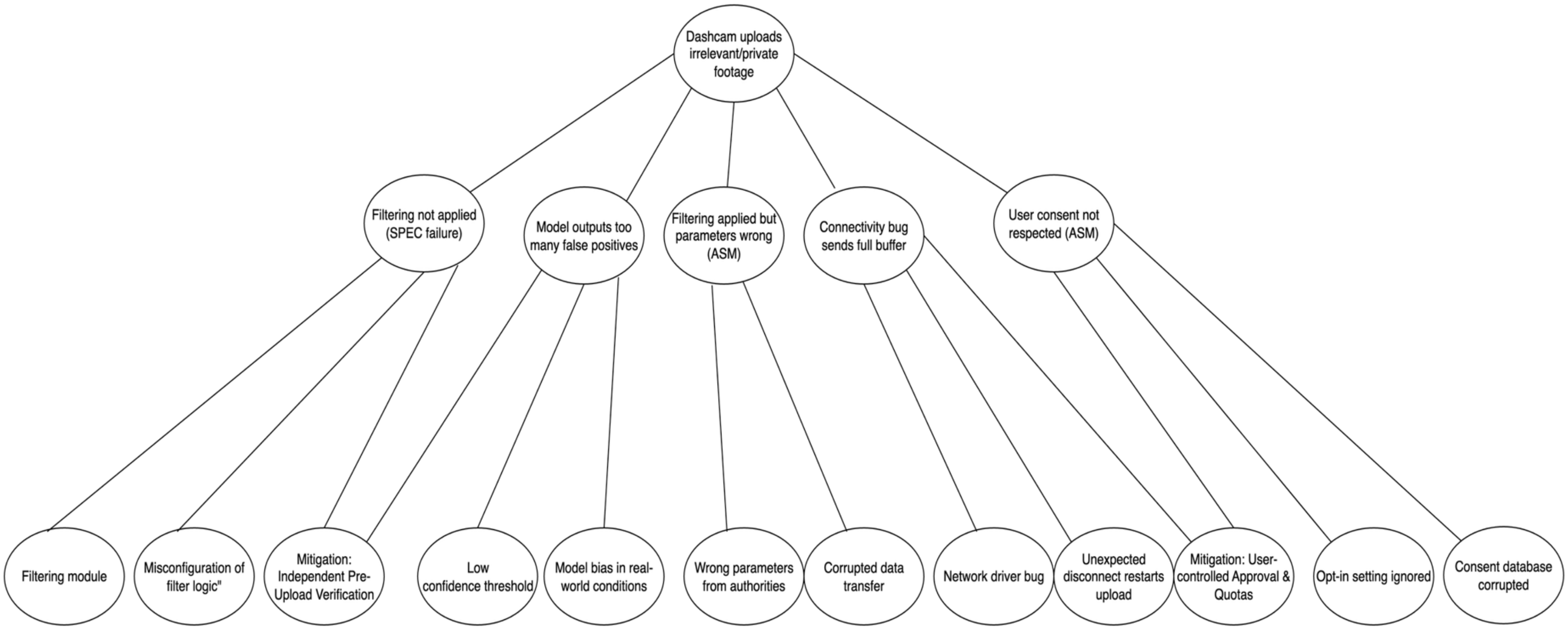
A lightweight verification module can check clips prior to uploading and verify pre-upload verification: confidence score, time/location, and clip duration. If the AI flags data as too excessive, or the filtering system fails, this independent verification module will suppress bulk uploads. This will prevent any clips, except the small percentage of the data that is validated as high confidence and relevant, from leaving the hard drive.

Mitigation 2: User-Controlled Approval and Quotas

A companion user interface will notify users prior to uploading data and provide them the option of approving or denying the upload. The user interface will also allow users to assign upload quotas on a monthly basis. This feature guarantees that user consent is adhered to, while at the same time eliminating unintentional upload caused by connectivity malfunctions that usage has experienced.

Impact on Risk

The independent module verifies known technical failures (i.e. filtering failure and false positives), while user controlled uploads provides coverage for known user consent and privacy risks. Collectively, the two mitigations will thwart high consequence fault tree paths and significantly reduce the chance of uploading irrelevant or privacy invasive footage.


Dashcam uploads irrelevant/private footage
Filtering not applied (SPEC failure)
Model outputs too many false positives
Filtering applied but parameters wrong (ASM)
Connectivity bug sends full buffer
User consent not respected (ASM)
Filtering module
Misconfiguration of filter logic"
Mitigation: Independent Pre-Upload Verification
Low confidence threshold
Model bias in real-world conditions
Wrong parameters from authorities
Corrupted data transfer
Network driver bug
Unexpected disconnect restarts upload
Mitigation: User-controlled Approval & Quotas
Opt-in setting ignored
Consent database corrupted