




IPSC IT FEST 2026

# Rulebook

*The Black Box*

The Doon School

DEHRADUN

## Terminology

| Term | Meaning |
|---|---|
| Autopsy | The initial manual online audit of a deliberately flawed dataset prior to model training. |
| Baseline Model | The initial machine learning model trained directly on the original flawed dataset without modifications. |
| Failure Prediction | A written hypothesis detailing anticipated model errors based on observed dataset flaws. |
| Confusion Matrix | A grid evaluating model performance by displaying actual versus predicted class outcomes. |
| Override | The controlled addition of up to 10 new targeted photos to fix identified dataset biases. |
| The Black Box AI Deployed Model | The designated web-based platform used to train and test image classification models. |

# Rulebook

## 1. Overview

1.1 To audit, debug, and fix a severely flawed image dataset so a computer vision model can accurately classify unseen test images.

1.2 Teams open a sealed dataset filled with missing samples, bad lighting, and heavy background bias. In Phase 1, they inspect the photos online and write down where they think the AI will fail. In Phase 2, they upload the broken photos to The Black Box AI Deployed Model and build a baseline model. In Phase 3, judges give them 30 unseen test images to test the model, map the mistakes on a grid, and spot the top two causes of failure. In Phase 4, teams shoot up to 10 new targeted photos to fix those specific flaws, retrain the model, and re-test it. Finally, in Phase 5, teams present their findings, explain their data fixes, and answer questions from the judges.

## 2. Teams and eligibility

2.1 2 Participants per school. 1 Participant is also permitted, at their own risk.

2.2 Open to school students in any grade. No prior competitive history is required.

2.3 No sort of substitutions will be permitted at any time. In special circumstances, substitutions may be during registration at the main desk.

## 3. Format and progression

3.1 The event consists of a single continuous competitive session divided into 5 distinct phases:

Phase 1 (Autopsy): Online-based dataset audit and submission of failure predictions (45 minutes).

Phase 2 (Baseline): Dataset upload, model training, and baseline metric logging (30 minutes).

Phase 3 (Hunt): Model evaluation on 30 hidden test images and confusion matrix analysis (45 minutes).

Phase 4 (Override): Capture of up to 10 patch images, model retraining, and re-testing (45 minutes).

Phase 5 (Presentation): Oral presentation of diagnostic findings and judge Q&A (10 minutes per team).

3.2 All teams progress sequentially through Phases 1 through 4. All teams present in Phase 5. Stage gates are strictly temporal; teams must submit deliverables for each phase at the specified deadlines to remain eligible for scoring in subsequent phases.

3.3 Phase 5 (Presentation) serves as the final stage. Each team receives 6 minutes to present their complete audit, predictions, confusion matrix findings, override strategy, and before/after test metrics. This is followed by a 4-minute Question & Answer session with the panel of judges.

3.4 In the event of a tie in total score, the winning team will be determined by applying the following criteria sequence:

• Highest classification accuracy achieved on the 30 hidden test images following Phase 4 (Override).

• Highest score awarded in Phase 1 (Autopsy & Failure Prediction).

• Highest score awarded in Phase 3 (Confusion Matrix & Error Tagging Accuracy).

• Lowest total time taken to submit deliverables across Phases 1 through 4.

## 4. Run structure

4.1 A 600-second (10-minute) setup window is provided prior to Phase 1. During this window, teams may power on provided hardware, verify internet connectivity, open designated web browsers, and check digital camera or webcam functionality. No event files or envelopes may be opened during setup.

4.2 The lead judge will announce "10 seconds to start" followed by an audible tone or whistle signaling "Go." Teams may only unseal the dataset envelope or open the dataset folder upon hearing the start signal.

4.3 The following are permitted during a run:

• Inspecting images manually.

• Training models exclusively within the designated platform The Black Box AI Deployed Model.

• Using team-assigned devices to photograph up to 10 patch images during Phase 4.

• Documenting findings on provided worksheets and template logs.

4.4 The following are strictly prohibited during a run:

• Writing custom code or script-based machine learning pipelines.

• Using image processing software (e.g., Photoshop, GIMP) to edit or crop dataset images.

• Deleting or altering the original provided training images in the baseline set.

• Taking more than 10 total patch photos during Phase 4.

• Communicating with individuals outside the assigned two-person team.

4.5 A run will end in the following circumstances:

Completion: A run concludes when Phase 5 presentation and judge Q&A are finished.

Time Expiry: Work must cease immediately upon the judge's signal at the end of each phase deadline.

Retirement: A team may voluntarily retire by submitting a written notice to the Event Head; retired teams receive zero points for remaining phases.

4.6 Model retrainings during Phase 2 and Phase 4 are unrestricted within the allocated phase time. However, official metric logging occurs strictly once per phase upon formal submission to the judges. No re-testing of the hidden 30-image test set is permitted outside designated review passes.

4.7 No external human assistance from advisors, teachers, or non-team members is permitted. Any unauthorized intervention results in immediate disqualification under §6.3.

## 5. Scoring

5.1 Points are awarded across a 100-point total scale based on 5 core components:

• Phase 1 Audit & Prediction Quality: 20 points

• Phase 2 Baseline Execution: 10 points

• Phase 3 Failure Analysis & Confusion Matrix: 25 points

• Phase 4 Override Strategy & Post-Patch Accuracy: 30 points

• Phase 5 Presentation & Defense: 15 points

5.2 Phase duration is tracked via an official master countdown timer displayed in the room. Time of submission for Phase 1, Phase 3, and Phase 4 deliverables is recorded to the nearest second by event heads for tie-breaking purposes.

## 6. Penalties and disqualification

6.1 Penalties, each with the points deducted or the sanction applied.

• Late Deliverable Submission: 2 points deducted per minute late (up to 5 minutes max per phase).

• Exceeding Phase 4 Image Limit: 5 points deducted per photo exceeding the 10-photo limit.

• Editing Original Training Set: 10 points deducted per occurrence.

6.2 Failure to submit deliverables for two consecutive phases results in disqualification from the active run.

6.3 Below are the conditions that cause disqualification from the event.

• Unauthorized communication with external parties or advisors during active phases.

• Use of prohibited automated scripts or external software tools.

• Tampering with hidden test images prior to official distribution in Phase 3.

• Unsportsmanlike conduct directed at judges, organizers, or other competitors.

## 7. Appeals

7.1 Only the officially registered participant from a school may lodge an appeal. The appeal must be submitted in writing within a strict window of 15 minutes after the official ruling or score is announced.

7.2 First instance: the Judge. Their ruling is final within the event.

7.3 If a team believes a systemic rule was broken, they may escalate the issue to the Event Coordinators. Only issues regarding gross misapplication of rules or scoring math errors may be escalated; subjective judging decisions cannot be contested.