Your team runs CampusFest.

TEAM HUB | Keep this trusted sheet on your table in every round.

Trusted event policy
Organizers never ask for login information, private codes or combinations, or payment. Check any request at the Info Desk in person.

Facts from the known Info Desk

1. The art club offers free stickers in the gym.

2. Volunteer check-in starts at 3:30 at the library desk.

3. The original music plan was 4:00 in the courtyard.

4. Rain update: music moves to the gym and starts at 4:15. A volunteer directs people from the courtyard.

5. Lost property: collect the water bottle at the library desk before 3:45.

Everyone votes on every message.

Reader: read it. Recorder: write it. Model Operator: apply the frozen rule. Verifier: explain the trusted check. Rotate roles once when retraining starts.

LEGIT means an organizer-authored message in this story. In real life, a filter prediction does not prove who sent a message. Use a channel you already know.

TRAIN: find the shortcut.

Open when instructed. These six examples include their true labels.

T1 | LEGIT

CampusFest: The art club is running a free sticker table in the gym.

T2 | LEGIT

CampusFest: Volunteer check-in is at the library desk at 3:30.

T3 | LEGIT

CampusFest: The original music plan is 4:00 in the courtyard.

T4 | SCAM

URGENT: Your prize expires. Reply with the practice access code to claim it.

T5 | SCAM

URGENT: Send your login password to keep your place at CampusFest.

T6 | SCAM

URGENT: Pay a gift-card fee to unlock your free event ticket.

Compare four clues: **URGENT / PASSWORD / PRACTICE ACCESS CODE / GIFT CARD.**
For each candidate, a match predicts SCAM; no match predicts LEGIT.

TEST: six new messages.

Keep closed until instructed. Vote first, then apply the frozen model.

A | VOTE BEFORE REVEAL

CampusFest: Free stickers are available at the art-club table in the gym.

B | VOTE BEFORE REVEAL

URGENT: Send your login password to reserve a free ticket.

C | VOTE BEFORE REVEAL

URGENT: Rain has moved the music set to the gym. A volunteer will meet you at the courtyard entrance.

D | VOTE BEFORE REVEAL

The event photo team needs your login password to add your picture.

E | VOTE BEFORE REVEAL

URGENT: A water bottle was left at the library desk. Collect it before 3:45.

F | VOTE BEFORE REVEAL

You have a free event ticket. Send a gift card to activate it.

For each message: **your vote → model prediction → trusted check → facilitator reveal.** Record the result on page 7. Keep the model frozen through A-F.

Learn it. Then lock it.

RULE WORKSHEET | Keep the reference rules covered until their reveal.

Which single clue fits the first six labels?

URGENT	Training errors: ____ / 6
PASSWORD	Training errors: ____ / 6
PRACTICE ACCESS CODE	Training errors: ____ / 6
GIFT CARD	Training errors: ____ / 6

Our proposed rule:

COVER BELOW THIS LINE UNTIL THE RULE REVEAL

Initial reference model
Contains URGENT? Predict SCAM. Otherwise, predict LEGIT.

KEEP THIS PART COVERED UNTIL RETRAINING

Improved reference model

1. Contains PASSWORD? Predict SCAM.
2. Otherwise, contains GIFT CARD? Predict SCAM.
3. Otherwise, contains PRACTICE ACCESS CODE? Predict SCAM.
4. Otherwise, predict LEGIT.

Match whole words or phrases. Ignore capitals. Treat hyphens as spaces.

FRESH: check the new rule.

Keep closed until after retraining. These messages were not used to train.

G | VOTE BEFORE REVEAL

URGENT: The performer needs extra setup time. Music now starts at 4:15 in the gym.

H | VOTE BEFORE REVEAL

The booth captain needs your practice access code to reserve a prize.

J | VOTE BEFORE REVEAL

Please send your password so the event team can save your photos.

K | VOTE BEFORE REVEAL

URGENT: Bring a gift card to pay the fee for your free ticket.

What did the model get right?

What could still fool it?

CHALLENGE: be the red team.

Open after the fresh check. Write one fictional scam, then swap clockwise.

Your goal: find a weakness in the **improved** model and explain a trusted check that catches it. Use fewer than 25 words. Use fictional roles and practice information.

Try another way to ask for a forbidden thing.

Examples: passphrase, PIN, verification code, locker combination, or a payment-app fee. The trusted policy covers login information, private codes or combinations, and payments.

From: [fictional CampusFest role]

Hook: [prize, photos, or event]

Ask: [a forbidden thing]

Receiving team
Model prediction: LEGIT / SCAM
Trusted policy: LEGIT / SCAM
Which forbidden request did you find? _____

Swap: team 1 → 2 → 3 → 4 → 5 → 1. Everyone votes.

Keep score. Find the gap.

TEAM RECORD | Each person votes; the Recorder writes the model's answer.

Team number: _____

Card	Our vote	Model says	Story label	Model correct?
A				
B				
C				
D				
E				
F				
G				
H				
J				
K				

Initial model: ____ / 6 correct. Improved model: ____ / 4 correct.

BOSS: What did the model miss?

Exit note: How would you double-check a real request?

Vote with both sides.

BALLOTS | Print this page 3 times. Cut on solid lines, fold on dashed lines.

LEGIT SCAM	LEGIT SCAM
LEGIT SCAM	LEGIT SCAM
LEGIT SCAM	LEGIT SCAM
LEGIT SCAM	LEGIT SCAM