X-Integrity First Meeting Guide

Purpose
Use this guide for the first technical review with X engineering, trust and safety, ranking, and policy stakeholders.

Meeting objective
Leave the room with alignment on three things:
1. the system is a forensic evidence layer, not a truth oracle
2. the integration path is low-risk because it is asynchronous and fail-to-neutral
3. the pilot can begin in advisory mode without changing the upload critical path

Recommended attendees
- media ingest / platform backend
- ranking / recommendations
- trust and safety
- community notes / reviewer ops
- legal or transparency representative
- SRE / infrastructure lead

Suggested 45-minute agenda

1. Opening frame - 5 min
Use this sentence:
> We are proposing an integrity operating layer that measures media consistency before virality, while leaving final platform treatment under X control.

2. Problem statement - 5 min
Cover:
- virality often outpaces verification
- current responses are reactive
- platforms need evidence before scale, not only correction after scale

3. System walkthrough - 10 min
Explain:
- Gatekeeper
- PRI
- Consensus
- Evaluator
- policy-configured treatment hint
- replay sidecar

4. Integration flow - 10 min
Walk through:
1. post-upload async ingest hook
2. parallel module execution
3. forensic state callback
4. feature store injection
5. notes / review / ranking consumption

5. Safety posture - 5 min
Emphasize:
- fail-to-neutral
- degraded mode on GEI/API issues
- sparse-region fairness
- no dependence on account ideology or political content labels

6. Pilot proposal - 5 min
Recommend:
- 14-day pilot
- high-velocity breaking-news vertical
- cohort model from shadow logging to ranking assist
- success metrics tied to lead time, fairness, and replay consistency

7. Q&A and close - 5 min
Ask for:
- one engineering owner
- one ranking owner
- one trust and safety owner
- decision on advisory-only pilot scope

Questions you should be ready to answer

Engineering
- What is the upload-path latency impact?
- How are timeouts handled?
- What fields are persisted for replay?
- How does the cache interact with snapshot replay?

Ranking
- Are these hard actions or soft features?
- Can the scores be used only in breaking-news surfaces first?
- How do we separate forensic evidence from product policy?

Trust and Safety
- How are rural and low-data cases protected?
- What triggers a manual review?
- How do structured reason strings appear to reviewers?

Legal / policy
- Can decisions be replayed deterministically?
- What is the boundary of the claim?
- How do we avoid overclaiming certainty?

Non-negotiable architectural points
- forensic state must remain separate from treatment hint
- replay sidecar is mandatory
- fail-to-neutral remains the default
- lane-aware routing must preserve honest constructed media
- account reputation is not part of the forensic core

Good next deliverables after the meeting
- pilot success metrics memo
- API contract review
- advisory-mode rollout checklist
- first region/vertical selection memo
- reviewer dashboard mock

What success looks like
At the end of the meeting, X should not feel asked to adopt a black-box detector.
They should feel they were handed:
- a forensic middleware layer
- a controllable rollout ladder
- a replayable audit mechanism
- a measurable pilot plan