

THE VENDOR INTERROGATION SCRIPT

10 Questions Before You Sign Off on an Automated Grievance Mechanism

Before you deploy any system that will process community grievances, compensation claims, or resettlement decisions, you need answers to these questions. In writing. In the contract. The vendor who can answer them clearly is worth your time. The vendor who deflects is telling you something important about where liability will land when the system fails.

SECTION A: THE LIABILITY ARCHITECTURE

1. When this system makes a classification error that harms a community member, who is legally responsible?

You are looking for specificity. 'Shared responsibility' means nobody. Get names, roles, contractual clauses.

■ Red flag: 'The human reviewer approved it, so...'

2. What is the maximum number of decisions a human reviewer will be expected to validate per hour?

Do the math yourself. If the number makes meaningful review physically impossible, you are being asked to be a liability sponge.

■ Red flag: Hesitation, vague answers, or 'it depends on workflow configuration.'

3. Can the human operator pause ALL decisions in a category without vendor approval?

This is bulk Stop Work Authority. If you can only override cases one by one, you have a veto that cannot scale to the speed of harm.

■ Red flag: 'Bulk pause requires a change request ticket.'

SECTION B: THE OVERRIDE ECONOMICS

4. Is there a cap on how often operators can override the system's recommendations?

Some systems flag 'operator bias' if you override too frequently. This penalises due diligence. You need to know the threshold and what happens when you exceed it.

■ Red flag: An override cap exists but wasn't mentioned until you asked.

5. What happens to throughput metrics when an operator exercises caution?

If careful review tanks your KPIs, the system is economically designed to discourage scrutiny.

■ Red flag: 'Efficiency gains' prominently featured; quality metrics absent.

SECTION C: THE TRANSPARENCY TEST

6. For any decision I am asked to validate, can I see the raw inputs the model used?

Not a summary. Not a confidence score. The actual data that drove the classification. If you cannot see what the model saw, you cannot meaningfully review it.

■ Red flag: 'Proprietary methodology' or 'aggregated for privacy.'

7. Can you show me which categories of grievance this model performs worst on?

Every model has blind spots. A vendor who knows where theirs are has tested honestly. A vendor who claims uniform performance has not.

■ Red flag: '94% accuracy across all categories.'

8. Has this model been validated on the specific dialect, language variant, and cultural context of this deployment?

'El agua está enferma' should not be downgraded because it doesn't match the keyword training set for 'contamination.' Local expression matters.

■ Red flag: 'The model is language-agnostic.'

SECTION D: THE FAILURE PROTOCOL

9. What is the system's default state when a threshold is breached or a rule is triggered?

The answer should be 'Hold.' Not 'Flag and proceed.' Not 'Route to queue.' Hold. The system should require active energy to harm, not active energy to save.

■ Red flag: 'Decisions continue processing while flagged items await review.'

10. Where are the known failure modes documented, and do they attach to relevant outputs in the audit trail?

Pre-registered blind spots are a sign of engineering maturity. 'We couldn't have predicted this' is not an acceptable post-incident explanation when the limitation was knowable in advance.

■ Red flag: Failure documentation exists but is not linked to live outputs.

HOW TO USE THIS DOCUMENT

These questions are designed for procurement discussions, vendor evaluations, and contract negotiations. They translate the principles of the Calvin Convention (pre-deployment rule sovereignty, human-defined uncertainty, default to hold, evidence access, bulk control, and pre-registered failure modes) into operational due diligence. A vendor who can answer all ten clearly, in writing, with contractual commitments, is building systems designed for genuine human oversight. A vendor who cannot is building something else.

—

From the Sociable Systems newsletter

Exploring how complex systems behave under real-world pressure

Subscribe: [linkedin.com/newsletters/sociable-systems](https://www.linkedin.com/newsletters/sociable-systems)