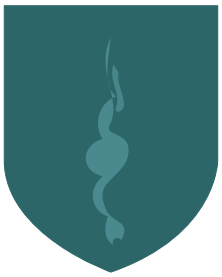


STABLE
SIGNAL

STABLE - SIGNAL Worksheet

Capture the earliest signs that an AI system may have behaved incorrectly, unsafely, or unpredictably.

by **Waydell D. Carvalho**

Published by Cinderpoint

SAFEMACHINE RULES	COMPONENT	EFFECTIVE	VARIANT
2026.05	**STABLE v1**	**2026-05-20**	**Template**

Instruction: Anyone can submit a SIGNAL - employees, customers, or automated alerts.

Use: This triggers the STABLE incident-response flow.

1. What Triggered the SIGNAL?

1. Who noticed the issue?
   - ☐ Customer
   - ☐ Employee
   - ☐ Automated alert
   - ☐ External party
   - ☐ Unknown

2. How was the issue discovered?
   - ☐ Direct complaint
   - ☐ Abnormal output
   - ☐ Support escalation
   - ☐ Log anomaly
   - ☐ Monitoring alert
   - ☐ Other

3. When did the event occur? (Date / time)

2. What Was Observed?

4. Describe the output or behavior that seemed wrong

5. Type of issue (check all)
   - ☐ Incorrect info
   - ☐ Unsafe content
   - ☐ Policy violation
   - ☐ Offensive / Biased
   - ☐ Technical error
   - ☐ Security concern
   - ☐ Unknown

6. Has this happened before?

☐ Yes ☐ No ☐ Not sure

3. Who or What Was Affected?

7. Did the issue reach a user?

☐ Yes ☐ No ☐ Unknown

8. Impacted parties

- ☐ One user
- ☐ Multiple users
- ☐ Internal team
- ☐ Automated system
- ☐ Unknown

9. Did the output propagate (copied / forwarded)?

☐ Yes ☐ No ☐ Unknown

4. Immediate Risk Level

10. Potential or actual harm?

☐ Yes ☐ No ☐ Possibly

11. Describe harm if any

12. Is immediate containment required?

☐ Yes ☐ No ☐ Unknown

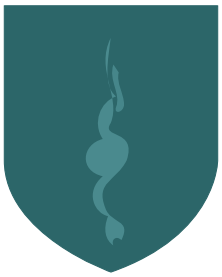
Red Flags

- ☐ Vague / incomplete report
- ☐ Automated alert without explanation
- ☐ Repeated complaints
- ☐ Unknown origin
- ☐ Unknown propagation
- ☐ Possible safety / compliance risk

6. SIGNAL Summary

13. Quick summary

14. Submitted by (name, role, date)



STABLE
TRIAGE

STABLE - TRIAGE Worksheet

Classify the incident, determine urgency, and route it to the correct owner within minutes.

by **Waydell D. Carvalho**

Published by Cinderpoint

SAFEMACHINE RULES	COMPONENT	EFFECTIVE	VARIANT
2026.05	**STABLE v1**	**2026-05-20**	**Template**

Instruction: If unsure, mark UNKNOWN - never guess during triage.

Use: Completed immediately after SIGNAL.

COMPLIANCE	Aligned with EU AI Act Art. 62 and ISO/IEC 42001 §8.4.
SAFETY	Incorrect triage increases risk, harm, and legal exposure.
TIMING	Triage should occur within minutes, not hours.

1. Confirm This Is an AI-Related Incident

1. Does the behavior originate from an AI system?

   ☐ Yes ☐ No ☐ Unknown

2. Which system produced the output? (System ID)

   _____

3. Type of output involved

   ☐ Text
   ☐ Recommendation
   ☐ Classification
   ☐ Action trigger
   ☐ Other

4. Is this part of normal operation?

   ☐ Yes ☐ No ☐ Unclear

2. Immediate Safety Assessment

5. Is the AI still producing risky outputs?

   ☐ Yes ☐ No ☐ Unknown

6. Does this require an urgent STOP or pause?

   ☐ Yes ☐ No ☐ Maybe

7. Risk targets

- ☐ Users
- ☐ Customers
- ☐ Employees
- ☐ Automated systems
- ☐ Data
- ☐ Reputation
- ☐ Compliance

3. Triage Classification Level

8. Severity rating

- ☐ Critical
- ☐ High
- ☐ Medium
- ☐ Low

9. Confidence level

- ☐ High
- ☐ Medium
- ☐ Low
- ☐ Unknown

10. Required escalation

- ☐ Legal
- ☐ Compliance
- ☐ Engineering
- ☐ CX Leadership
- ☐ Executive
- ☐ None

4. Initial Containment Actions

11. Was STOP already executed?

☐ Yes ☐ No ☐ Not needed

12. Disable affected workflow temporarily?

☐ Yes ☐ No

13. Restrict user-facing features?

☐ Yes ☐ No

14. Capture logs now?

☐ Yes ☐ No

5. Information Gathering Checklist

15. Evidence collected

- ☐ Screenshots
- ☐ Logs
- ☐ Transcripts
- ☐ Inputs
- ☐ Outputs
- ☐ System state
- ☐ None

16. SIGNAL report complete?

☐ Yes ☐ No ☐ Needs clarification

17. Additional info required?

☐ Yes ☐ No

Red Flags

- ☐ Unclear system identity
- ☐ Repeated user behavior
- ☐ Harm increasing
- ☐ Legal exposure possible
- ☐ No owner identified
- ☐ No containment applied
- ☐ Missing logs
- ☐ Unclear escalation path

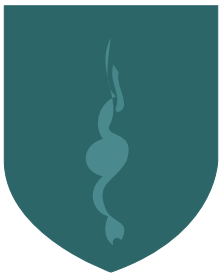
7. TRIAGE Summary

18. Triage decision

19. Assigned owner for ASSESS

20. Response expectation

- ☐ Immediate
- ☐ 30 minutes
- ☐ 2 hours
- ☐ End of day



S T A B L E

A S S E S S

STABLE - ASSESS Worksheet

Identify the root cause of the incident so BUFFER can be executed correctly.

by **Waydell D. Carvalho**

Published by Cinderpoint

SAFEMACHINE RULES	COMPONENT	EFFECTIVE	VARIANT
2026.05	**STABLE v1**	**2026-05-20**	**Template**

Instruction: If unsure, mark UNKNOWN - never guess during incident analysis.

COMPLIANCE	EU AI Act Art. 62, ISO/IEC 42001 §8.4, NIST AI RMF Manage.
LEGAL	Final legal determination must be confirmed by Legal / Compliance.

1. Scope of Impact

1. Who or what was affected?

- ☐ Single user
- ☐ Multiple users
- ☐ Internal staff
- ☐ External customers
- ☐ Automated systems
- ☐ Unknown

2. Estimated number of affected interactions

- ☐ 1-10
- ☐ 11-100
- ☐ 101-1,000
- ☐ > 1,000
- ☐ Unknown

3. Did the output propagate?

☐ Yes ☐ No ☐ Unknown

4. Systems / components involved

2. Harm & Risk Assessment

5. Type of harm (check all)

- ☐ Incorrect info
- ☐ Unsafe suggestion
- ☐ Policy violation
- ☐ Offensive content
- ☐ Security / privacy risk
- ☐ Financial / operational disruption
- ☐ Unknown

6. Could harm occur if unaddressed?

☐ Yes ☐ No ☐ Potentially

7. Did this violate internal policy?

☐ Yes ☐ No ☐ Unclear

8. Possible regulatory impact?

☐ Yes ☐ No ☐ Unclear

3. Reproducibility Check

9. Can the issue be reproduced?

☐ Yes - consistently
☐ Yes - sometimes
☐ No
☐ Unknown

10. Issue tied to specific model versions?

☐ Yes ☐ No ☐ Unknown ☐ Not applicable

11. Has this occurred before?

☐ Yes ☐ No ☐ Uncertain

4. Root-Cause Domain Check

12. Model drift?

☐ Yes ☐ No ☐ Unknown

13. Data drift?

☐ Yes ☐ No ☐ Unknown

14. ALIGN / RULE boundary violated?

☐ Yes ☐ No ☐ Unknown

15. Human misuse contributed?

☐ Yes ☐ No ☐ Unknown

16. Vendor model / config change involved?

Vendor change documentation available?

☐ Yes ☐ No ☐ Unknown

17. Configuration / workflow contributed?

☐ Yes ☐ No ☐ Unknown

18. Multi-system interaction factor?

☐ Yes ☐ No ☐ Unknown

19. Missing or outdated documentation?

☐ Yes ☐ No ☐ Unknown

5. Evidence Completeness Check

20. Evidence collected

- ☐ Logs
- ☐ Screenshots
- ☐ Inputs
- ☐ Outputs
- ☐ Version details
- ☐ System state
- ☐ None

21. Evidence completeness

- ☐ Complete
- ☐ Partial
- ☐ Missing

6. ASSESS Summary

22. Severity classification

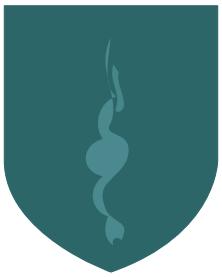
- ☐ Low
- ☐ Medium
- ☐ High
- ☐ Critical

23. Protected class disproportionate impact (if applicable)?

☐ Yes ☐ No ☐ Unknown

24. Assigned owner for BUFFER

25. Required notes before BUFFER



STABLE

BUFFER

STABLE - BUFFER Worksheet

Prevent the incident from spreading or worsening while fixes are prepared.

by **Waydell D. Carvalho**

Published by Cinderpoint

SAFEMACHINE RULES	COMPONENT	EFFECTIVE	VARIANT
2026.05	**STABLE v1**	**2026-05-20**	**Template**

Instruction: BUFFER is containment only. Do NOT modify model weights, prompts, configs, or internal logic.

> **LEGAL** Legal / Compliance must confirm temporary restrictions comply with access obligations.

1. Containment Decision

1. Containment level

- ☐ Level 0 - Monitoring only
- ☐ Level 1 - Limited restriction
- ☐ Level 2 - Partial workflow pause
- ☐ Level 3 - Full STOP (kill switch)
- ☐ Unknown

2. Reason for containment level

2. Temporary Containment Actions (non-destructive)

3. Restrict affected AI feature(s)?

☐ Yes ☐ No ☐ Unknown

4. Restrict user-facing exposure?

☐ Yes ☐ No ☐ Unknown

5. Pause affected workflow (non-destructive)?

☐ Yes ☐ No ☐ Unknown

6. Block high-risk inputs at workflow / UI (NOT inside model)?

☐ Yes ☐ No ☐ Unknown

7. Enable fallback mode?

☐ Yes ☐ No ☐ Unknown

8. Isolate downstream workflows / systems?

☐ Yes ☐ No ☐ Unknown

3. Monitoring & Stability Check

9. Issue reproducible after containment?

☐ Yes ☐ No ☐ Unknown

10. Dashboards / alerts active?

☐ Yes ☐ No

11. Logs confirm containment?

☐ Yes ☐ No ☐ Partial

4. Dependency & Exposure Check

12. Downstream teams notified?

☐ Yes ☐ No

13. Are customers affected?

☐ Yes ☐ No

14. Other workflows impacted?

☐ Yes ☐ No ☐ Unknown

15. Cross-team coordination required?

☐ Yes ☐ No

5. Safety Boundary Verification

16. All prohibited actions blocked?

☐ Yes ☐ No ☐ Unknown

17. Fallback processes functioning safely?

☐ Yes ☐ No ☐ Unknown

18. Any required services blocked accidentally?

☐ Yes ☐ No ☐ Unknown

Red Flags

- ☐ Containment unclear
- ☐ Users still exposed
- ☐ Downstream workflows active
- ☐ Logs not updating
- ☐ Monitoring incomplete
- ☐ Fallback not configured
- ☐ Temporary measures breaking required services
- ☐ Propagation not confirmed blocked

7. BUFFER Summary

19. Final containment state

20. Stable enough for LEARN?

☐ Yes ☐ No ☐ Re-assess

21. Assigned owner for LEARN

22. Legal / Compliance confirmation received?

☐ Yes ☐ No

STABLE

LEARN

STABLE - LEARN Worksheet

Document the complete incident narrative, causal chain, evidence, and organizational learning required before EXECUTE can begin.

by **Waydell D. Carvalho**

Published by Cinderpoint

SAFEMACHINE RULES	COMPONENT	EFFECTIVE	VARIANT
2026.05	**STABLE v1**	**2026-05-20**	**Template**

Instruction: LEARN occurs after SIGNAL ’ TRIAGE ’ ASSESS ’ BUFFER. Do NOT propose fixes here - fixes belong in EXECUTE.

HUMAN FACTOR	All human-factor observations are non-punitive and used only for safety analysis.
LEGAL	Systemic insights gathered here are internal learning only, not regulatory submissions.

1. Incident Narrative Reconstruction

1. Describe the timeline from incident start ’ detection ’ containment

2. Who first detected the issue?

- ☐ Customer
- ☐ Employee
- ☐ Monitoring
- ☐ Unknown

3. Inputs involved

2. Causal Chain Reconstruction

4. Primary root cause from ASSESS

5. Secondary contributing factors (max 3)

6. Organizational or procedural gaps?

If yes, describe.

☐ Yes ☐ No

3. Human Factors & Operational Context

7. Operator misuse involved?

☐ Yes ☐ No ☐ Unclear

8. Missing guidance or unclear documentation?

☐ Yes ☐ No

9. Communication / workload / incentives relevant?

☐ Yes ☐ No

4. Systemic Learning Points (internal use only)

10. Governance gaps (rules, policies, oversight)

11. Monitoring / alerting gaps

12. Missing safeguards that allowed exposure

13. Did fallback / manual modes work in BUFFER?

☐ Yes ☐ No

5. Evidence Review & Documentation Quality

14. Evidence packet completeness

☐ Complete
☐ Partial
☐ Missing

15. Documentation accuracy (configs, rules, constraints)

☐ Yes ☐ No ☐ Partial

Red Flags

- ☐ No clear causal chain
- ☐ Human factors ignored
- ☐ Missing evidence
- ☐ Organizational gaps undocumented
- ☐ No systemic insights
- ☐ Timeline incomplete
- ☐ Buffer validation missing
- ☐ Blame language detected

7. LEARN Summary

16. Key learnings

17. Conditions required before EXECUTE can begin

18. Assigned owner for EXECUTE

STABLE

EXECUTE

STABLE - EXECUTE Worksheet

Implement the corrective actions approved during LEARN, verify they work, and confirm no new risks are introduced.

by **Waydell D. Carvalho**

Published by Cinderpoint

SAFEMACHINE RULES	COMPONENT	EFFECTIVE	VARIANT
2026.05	**STABLE v1**	**2026-05-20**	**Template**

Instruction: EXECUTE begins only after SIGNAL ’ TRIAGE ’ ASSESS ’ BUFFER ’ LEARN.

BOUNDARY	EXECUTE may only implement actions approved in LEARN - no new actions may be added.
APPROVAL	All changes must have documented approval before implementation.
DUAL-CONTROL	All reproduction tests must be verified by a second reviewer.
REGULATORY	Fixes affecting user rights, safety, or access require Legal / Compliance review.

1. Approved Action List

1. Corrective actions approved

2. Legal / Compliance approval?

☐ Yes ☐ No

3. Engineering / IT approval?

☐ Yes ☐ No

4. Confirm no new actions added

☐ Yes ☐ No

2. Implementation Steps

5. Model / prompt / config updates?

Details.

☐ Yes ☐ No

6. Workflow / routing / logic updates?

Details.

☐ Yes ☐ No

7. Documentation updates?

Details.

☐ Yes ☐ No

8. Monitoring / alert threshold updates?

Details.

☐ Yes ☐ No

3. Verification Checklist

9. Reproduction test (second reviewer required)

Second reviewer initials.

- ☐ Resolved
- ☐ Partial
- ☐ Unresolved

10. Check unrelated workflows (regression test)

- ☐ Passed
- ☐ Partial
- ☐ Failed

11. Fallback / manual process after fix

- ☐ OK
- ☐ Needs update
- ☐ N/A

12. Logging / monitoring validation

- ☐ Updated
- ☐ Alerts calibrated
- ☐ Incomplete

4. Risk Re-Evaluation

13. New risks introduced?

☐ Yes ☐ No

14. New safeguard / rule created?

☐ Yes ☐ No

15. Does this fix trigger a SAFEARC update?

Which section.

☐ Yes ☐ No

5. Documentation & Handoff

16. Updated documentation stored correctly?

☐ Yes ☐ No

17. Responsible team briefed?

☐ Yes ☐ No

18. Training / retraining required?

☐ Yes ☐ No

19. Legal / Compliance verification needed?

If yes, documentation.

☐ Yes ☐ No

Red Flags

- ☐ Unapproved fix
- ☐ New risk created
- ☐ Verification incomplete
- ☐ Regression failed
- ☐ Monitoring incomplete
- ☐ Documentation missing
- ☐ SAFEARC trigger not recorded
- ☐ Missing second reviewer

7. EXECUTE Summary

20. Final state

21. Follow-up tasks / owners

22. Ready for STABLE closure?

☐ Yes ☐ No ☐ Needs review

STABLE

CLOSURE

STABLE - CLOSURE Sign-Off (Optional)

Confirm all six STABLE steps (SIGNAL ’ TRIAGE ’ ASSESS ’ BUFFER ’ LEARN ’ EXECUTE) were completed correctly.

by **Waydell D. Carvalho**

Published by Cinderpoint

SAFEMACHINE RULES	COMPONENT	EFFECTIVE	VARIANT
2026.05	**STABLE v1**	**2026-05-20**	**Template**

Closure Sign-Off

1. Incident ID

2. Date of final verification

3. All six STABLE steps completed

   - ☐ SIGNAL
   - ☐ TRIAGE
   - ☐ ASSESS
   - ☐ BUFFER
   - ☐ LEARN
   - ☐ EXECUTE

4. Evidence packet stored in correct repository

   ☐ Yes ☐ No

5. Owner confirms system is stable and safe (name + date)

6. Legal / Compliance final verification

   - ☐ Not required
   - ☐ Required and completed

7. Closure summary

8. Final sign-off (executive name, signature, date)