

INCIDENT RESPONSE & AI INVESTIGATIONS FRAMEWORK

AI-Assisted. Human-Led. Trusted Investigations. Practical framework for governing the responsible use of AI in Incident Response

**A LIVING
FRAMEWORK**

**AI-ASSISTED.
HUMAN-LED.**

**INCIDENT
RESPONSE**

Created by: **Heather Barnhart**

DFIR Curriculum Lead and Head of Faculty, SANS Institute Safe AI Integration | SWGDE Best Practices for Digital & Multimedia Evidence

The human examiner is always the authority | Disclaimer: This is a living document. Guidance reflects current understanding and is subject to change as AI technology, legal precedent, and industry practices evolve. Always confirm content is current before use.

CORE PRINCIPLES

Why This Framework Exists

Why This Framework Exists

AI adoption in DFIR is already happening, with or without formal policies. Most teams lack governance models, validation standards, and defensible practices for integrating AI into critical investigative workflows.

Validation with current tools is often insufficient. As AI assertion confidence increases, the problem compounds without clear guardrails.

Augmentation, Not Replacement

AI is a force multiplier. It accelerates work; it does not own conclusions. The human examiner remains the final authority at every phase.

The analyst must independently verify AI findings, make the final determination, and accept full responsibility for all assertions.

Force Multiplier

AI is a force multiplier, not a replacement for trained analysts or examiners. It accelerates work. It does not own conclusions.

Human Accountability

Human accountability must remain central to all investigative decisions. AI systems should not assume legal or professional responsibility.

Validation Required

AI outputs must be validated before operational or legal use. Every AI-assisted finding used in an investigation requires human review.

Designed to Evolve

This framework is designed to evolve alongside AI technology, legal precedent, and industry best practices.

HUMAN OVERSIGHT MODELS

Risk Categories

1 Human IN the Loop Maximum Control

Humans approve AI actions before execution.

Required for high-risk, legally sensitive, or irreversible tasks.

Applies to:

Containment approval | Eradication sign-off

2 Human ON the Loop Monitor and Intervene

Humans supervise AI-driven processes in real time.

AI acts, but humans monitor and can intervene at any point.

Applies to:

Detection | Response analysis

2 Human OVER the Loop Governance

Humans maintain governance and policy oversight. AI operates within defined, approved boundaries.

Applies to:

Preparation | Policy

RISK CATEGORY

Reversibility Risk

Activities that may alter systems, evidence, or operational environments require increased human oversight. Actions may not be easily undone.

e.g. Evidence imaging, system changes, collection

RISK CATEGORY

Accountability Risk

Attribution, validation, testimony, legal conclusions, and final decision-making require direct human ownership. Responsibility cannot be delegated to AI.

e.g. Attribution, expert testimony, final conclusions

Critical Principle: Regardless of oversight model, digital forensic conclusions require direct human validation before they are acted upon or entered into any legal or evidentiary record.

HUMAN OVERSIGHT MODELS

Risk Categories

	AI Recommended Low Risk	AI Optional Elevated Risk	AI Never Human Only	Human Driven
IR1	Preparation Human Over Loop	• Policy gap analysis and playbook drafting assistance • AI-assisted training simulations and tabletop exercises • Automated asset inventory, classification, and risk scoring		AI RECOMMENDED FOR508 FOR572 FOR577
IR2	Detection Human On Loop	• AI can add efficiency to the Detection phase. This may require Human in the Loop and Human on the Loop. • Automated IoC triage, enrichment, and threat scoring • AI anomaly detection and log correlation		AI RECOMMENDED FOR508 FOR572 FOR589
IR3	Response and Containment Human In Loop	• AI-suggested containment actions; Human In the Loop required before execution • Automated evidence preservation tasks • AI-assisted timeline construction and report drafting		AI OPTIONAL FOR508 FOR572 FOR608
IR4	Eradication and Recovery Human In Loop	• Human confirms threat fully removed before recovery begins • Recovery prioritization requires human judgment • Backup integrity validation and run-book draft generation		AI OPTIONAL FOR508 FOR528 FOR577
IR5	Lessons Learned Human In Loop	• A human must validate AI findings at all times • Human examiner must independently verify all conclusions • Legal and evidentiary integrity requires human accountability		HUMAN DRIVEN FOR508 FOR572 FOR589
AI MAY BE A COLLABORATOR A HUMAN MUST MAKE THE FINAL DECISION				

HIGH-RISK AREAS

Operational Use Cases

SANS DFIR

HIGH-RISK AND PROHIBITED

Some actions must never be delegated to AI, regardless of capability, confidence score, or efficiency gains.

No Independent Attribution

AI should not independently determine attribution. Attribution requires human expertise and legal accountability.

No Legal or Evidentiary Conclusions

AI should not make legal or evidentiary conclusions. These carry professional weight that cannot be delegated to automated systems.

Testimony and Reporting Ownership

Human experts must retain ownership of testimony and final reporting. Expert witness credibility cannot be transferred to an AI system.

Full Documentation Required

All AI-assisted actions must be documented. Transparent records of where and how AI was used are essential for defensible investigations.

OPERATIONAL USE CASES

Where AI can responsibly accelerate workflows, with human oversight at every step.

Timeline Reconstruction

AI-assisted timeline reconstruction and evidence triage accelerates identification of key events across large forensic datasets.

Triage Assistance

AI can help prioritize, score and flag evidence at scale requiring deeper investigation.

Potential Evidence Identification

AI can scan for additional sources of evidence and flag this for the examiner.

Report Drafting

AI-assisted drafting reduces analyst burden while preserving human authority over all final findings and language. Human must author and sign off.

Critical Boundary: AI CANNOT put a human behind an artifact to prove attribution.

ACCELERATE DFIR. RESPONSIBLY.

Start With Low-Risk
Use Cases

Human Examiner Is
Final Authority

Transparent Governance
and Validation

AI can accelerate DFIR operations responsibly. The human examiner remains the final authority. Organizations should begin with low-risk use cases and build governance before expanding capability.

A LIVING FRAMEWORK FOR SAFE AI INTEGRATION IN DFIR