



U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND ARMAMENTS CENTER

AI-ASSISTED TRANSFORMATION OF OPERATIONAL RECORDINGS INTO
REVIEWABLE EVIDENCE

23 SEPTEMBER 2026

Distribution Statement:

A

POC:

Robert M. DeMarco

PRESENTED BY: ELIZABETH MEZZACAPPA, TACTICAL BEHAVIOR RESEARCH LABORATORY

COLLECTING DATA IS NOT THE SAME AS EXTRACTING EVIDENCE



OPERATIONAL EXPERIMENTATION NOW COLLECTS MORE HUMAN DATA THAN ANALYSTS CAN EXPLOIT MANUALLY

WHAT WE COLLECT

- System logs
- Performance metrics
- Audio and video
- After-action discussions

WHAT ANALYSTS NEED

- Relevant statements
- Construct labels
- Source context
- Reviewable findings

The problem is not collecting data; it is turning recordings into evidence quickly enough to matter.

OPERATIONAL RECORDINGS CAPTURE HUMAN-SYSTEMS EVIDENCE



Recordings capture what Soldiers and operators say during use, troubleshooting, training, and after-action discussions.

WHAT THE RECORDINGS CONTAIN

- Mission dialogue
- Training interactions
- Technical discussion
- Troubleshooting
- User feedback
- After-action commentary

WHAT THE RECORDINGS CAN TELL US

TRUST

Why did they rely on or reject the system?

WORKLOAD

Where did effort or cognitive burden increase?

USABILITY

What made interaction easy or difficult?

COMMUNICATION

Where did information transfer break down?

FEEDBACK CLARITY

Did operators understand system outputs?

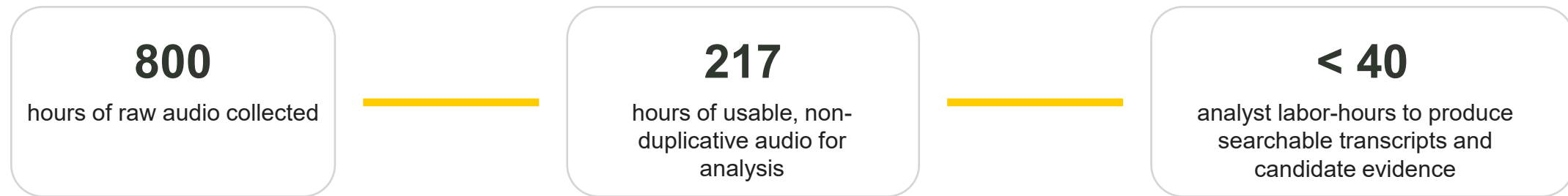
HMT

How did human-machine teaming actually work?

SCALE OF THE FIELD DATA CHALLENGE



One two-week field experiment produced hundreds of hours of overlapping audio from multiple recorders, personnel, and locations.



WHY “RAW HOURS” ARE NOT THE SAME AS “UNIQUE EVIDENCE”

- Multiple recorders may capture the same dialogue from different positions
- Overlapping coverage must be organized so repeated capture is not mistaken for independent evidence
- The analysis starts by preserving source, date, recorder, and mission-relevant context

The scale creates a search problem before it becomes an interpretation problem.

CLARITY CHANGES THE ANALYST'S JOB



TRADITIONAL WORKFLOW

- 1 Listen
- 2 Search
- 3 Transcribe / clean
- 4 Locate evidence
- 5 Organize material

Massive search problem

CLARITY WORKFLOW

- 1 Start with candidate evidence
- 2 Check context
- 3 Verify source
- 4 Revise or reject
- 5 Synthesize findings

Targeted validation problem

**The AI is not being asked to produce the answer;
it transforms a massive search problem into a smaller validation problem.**

FIVE-STAGE CLARITY WORKFLOW

**1****ORGANIZE**

Event, date, recorder,
and activity context

2**TRANSCRIBE**

Whisper-generated
SRT transcripts with
timestamps

3**EXTRACT**

AI-generated candidate
evidence tables

4**REVIEW**

SME review, revision,
rejection, or
reassignment

5**SYNTHESIZE**

Evidence tables,
summaries, briefings,
and products

DESIGN PRINCIPLE

- AI output is treated as draft evidence, not final findings
- Every analytical product should preserve a path back to transcript, timestamp, and original source audio
- Human analysts retain responsibility for interpretation, validation, and final reporting

CANDIDATE EVIDENCE, NOT FINAL FINDINGS



CLARITY generates structured tables that analysts can inspect, challenge, revise, and validate.

TIME	SOURCE	EXCERPT	CONSTRUCT	STATUS	ANALYST ACTION
00:14:22	Rec-03	"I could not tell why it recommended that."	Feedback clarity	Candidate	Check source audio
00:31:09	Rec-07	"I trusted it after it was consistent."	Trust	Revised	Link to prior event
01:05:48	Rec-02	"This took my attention away from the task."	Workload	Accepted	Use in evidence matrix
01:42:16	Rec-05	"We needed more training before using it."	Training burden	Merged	Duplicate coverage

TRACEABLE

Every entry has source and time context

REVIEWABLE

Analysts accept, revise, reject, merge, or reassign

NOT FINAL

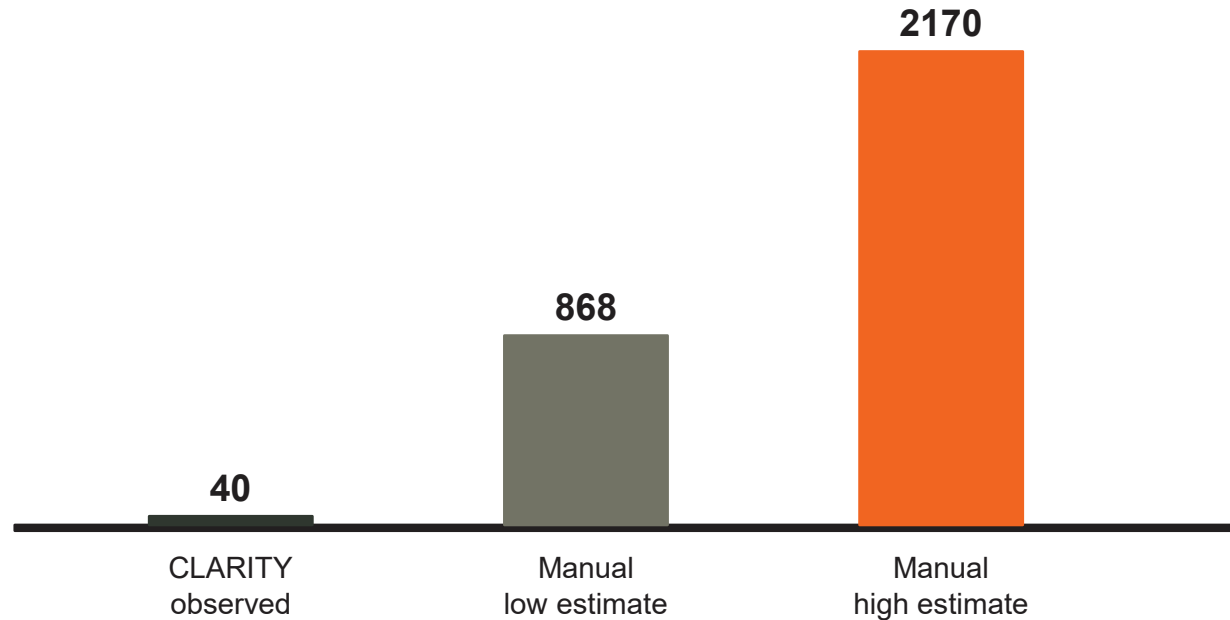
The model identifies material for analyst attention

ANALYST BURDEN WAS REDUCED



Applied workflow estimate: results demonstrate potential magnitude, not a controlled benchmark.

ESTIMATED ANALYST HOURS



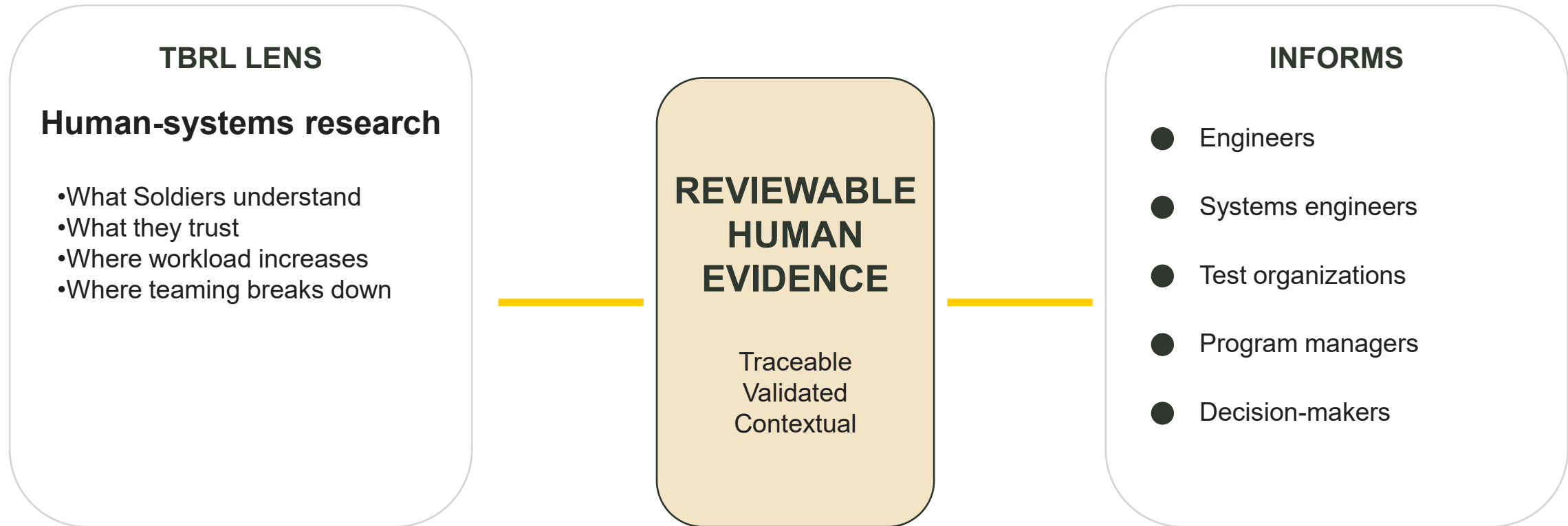
95–98%

estimated reduction in early-stage analysis burden

Compared with a four-to-ten labor-hour manual baseline per hour of audio

The big result is the role shift. Analysts move from broad discovery toward targeted validation and synthesis.

HUMAN EVIDENCE CAN INFORM ENGINEERING AND EVALUATION



Contribution: underused human interaction data becomes a practical analytical input



LESSONS LEARNED AND WHERE CLARITY IS GOING

1

TRACEABILITY

Evidence must link back to transcript, timestamp, and source recording

2

STRUCTURE

Evidence tables are more useful than unconstrained summaries

3

HUMAN REVIEW

AI output remains candidate evidence until reviewed

4

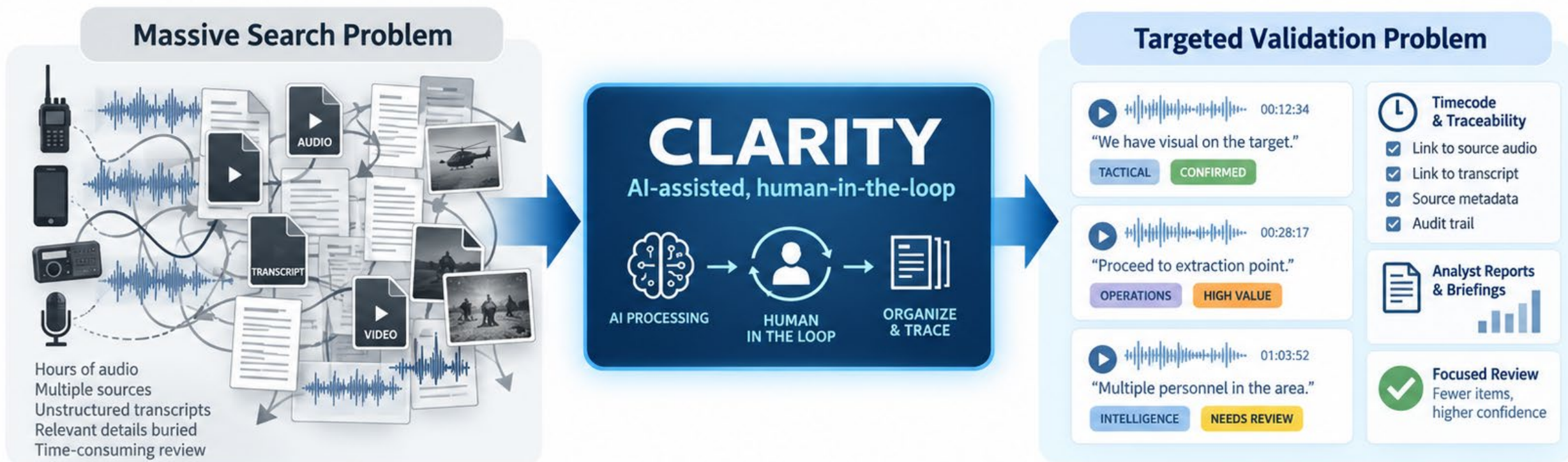
VALIDATION

Future work must compare AI-assisted review against manual coding

CURRENT EVOLUTION

- Refining an analyst-driven workflow into a reusable CLARITY capability
- Improving multiple-recording handling, source traceability, structured prompts, and reviewable outputs
- Reducing researcher time spent on machine-assistable tasks so SME judgment focuses on interpretation

From Operational Recordings to Reviewable Evidence



“The AI is not being asked to produce the answer. It is being used to transform a massive search problem into a much smaller validation problem.”

Human analysts still decide what the evidence means.

THANK YOU.

