AI-ASSISTED CRIMINAL INTERROGATIONS: A HUMAN-IN-THE-LOOP FRAMEWORK FOR TRANSCRIPTION, COMPLIANCE MONITORING AND POST-INTERVIEW ANALYSIS

*Gauri U. Ansurkar & **Gauri S. Mhatre

* Department of Information Technology and Computer Science, KSD's Model College (Empowered Autonomous), Mumbai, India.

** Department of Information Technology and Computer Science, KSD's Model College (Empowered Autonomous), Mumbai, India .

Abstract:

The growing use of artificial intelligence in criminal justice has intensified debate around reliability, transparency, and procedural fairness, particularly in the highly sensitive context of criminal interrogations. Although recent AI advances enable analysis of speech, behavior, and interaction patterns, prior research consistently shows that inferential applications—such as deception detection, emotion recognition, or predictive judgments of credibility—remain scientifically contested and risk introducing bias, automation dependence, and undue influence into investigative decision-making. In contrast, established interrogation research emphasizes accurate documentation, procedural compliance, and non-coercive interviewing as foundational to investigative integrity and evidentiary reliability.

This paper proposes a human-in-the-loop framework for AI-assisted criminal interrogation rooms that intentionally excludes predictive or judgment-oriented functions. Instead, the framework prioritizes non-inferential, supportive AI capabilities designed to enhance transparency and structured review while preserving human authority. These include automated transcription and structured summarization of interviews, contextual visualization of interactional trends without evaluative labeling, consistency and timeline analysis across multiple interviews, and automated indicators for monitoring rights notification and procedural compliance.

The framework introduces a modular system architecture and dashboard-oriented review layer that separates interview recording from post-interview analysis. This design mitigates automation bias, reduces the risk of coercive influence during questioning, and supports accountable review by investigators, courts, and defense stakeholders. Overall, the proposed approach offers a practical and ethically grounded model for responsible AI integration in criminal interrogations.

Keywords: *AI-assisted criminal interrogation; human-in-the-loop artificial intelligence; investigative interview transcription; procedural compliance monitoring; post-interview analytics; explainable decision-support systems.*

Introduction:

The integration of AI into criminal justice, including interrogation contexts, has generated substantial debate about reliability, transparency, and procedural fairness. Research across predictive policing, risk assessment, and AI enabled interrogation tools highlights persistent concerns about validity, biases, and how inferential judgments (e.g., deception

detection, emotion recognition) can unduly influence investigations [1].

Foundational interrogation scholarship emphasizes the primacy of accurate documentation, non-coercive interviewing, and strict procedural safeguards to ensure evidentiary reliability and integrity [3] There is a gap in practical design work that reconciles AI usefulness with ethical risk in interrogation settings: specifically, how to harness AI for documentation and oversight without enabling inferential or predictive judgments that may bias outcomes or compromise rights[4][5][6]. The explicit mechanistic separation of real-time capture from post-hoc analysis, and the deliberate exclusion of inference-oriented functions, requires rigorous articulation, empirical evaluation, and legal-ethical grounding across diverse investigatory contexts (e.g., variations in rights advisement, recording practices, and defense access) [3][4]

A human-in-the-loop, non-inferential AI framework for interrogation rooms that defers judgment to humans and uses AI only for defensible, documentary roles: automated transcription, structured summarization, contextual visualization without diagnostic labels, cross-interview consistency analyses, and automated rights and procedural compliance monitoring [1][2][6][8]

A modular system architecture that separates real-time recording from post-interview analysis to mitigate automation bias and coercive influence, enabling clear pathways for oversight by investigators, courts, and defense counsel [3][4].

Concrete dashboard concepts and design principles that operationalize transparency, accountability, and ethical safeguards, providing a blueprint for responsible AI integration and a basis for empirical and legal validation [5][6][8][9]

Literature Review:

The integration of artificial intelligence into criminal justice has expanded rapidly, with applications in predictive policing, risk assessment, and forensic analysis. However, its use in criminal interrogations remains limited and highly contested due to ethical, legal, and scientific concerns. Existing scholarship consistently distinguishes between supportive AI tools that enhance documentation and transparency and inferential systems that attempt to predict guilt, deception, or emotional states, the latter being widely criticized.

Berk et al. [1] examine the rise of predictive and analytical AI in law enforcement and caution that algorithmic systems often reinforce existing biases while obscuring accountability. They emphasize that AI outputs can strongly influence human judgment, particularly in high-stakes contexts. This concern is directly relevant to interrogations, where investigative direction can be shaped by subtle cues and perceived "objective" system feedback.

From a legal and human rights perspective, Završnik et al. [2] argue that AI systems in criminal justice must not undermine core protections such as the presumption of innocence, the right to silence, and the right to a fair trial. They stress that meaningful human oversight is essential wherever AI is deployed in processes affecting individual liberty. Their work supports the necessity of human-in-the-loop designs in interrogation environments.

Interrogation methodology research further reinforces the need for procedural rigor. Meissner et al. [3], through a systematic review of interview and interrogation methods, demonstrate that non-coercive, information-gathering approaches lead to more reliable outcomes and significantly reduce the risk of false confessions. Their findings highlight the importance of accurate documentation, structured questioning, and procedural safeguards—areas where AI can provide support without replacing human judgment.

Recent studies have explored specific AI applications

Original Research Article

in investigative interviewing. Stoykova et al. [4] analyze the use of automatic speech recognition for transcribing investigative interviews, noting its potential to improve accuracy, transparency, and auditability. Their work positions transcription as one of the most defensible and legally acceptable uses of AI in interrogation contexts.

In contrast, Wu et al. [5] critically examine AI-based lie detection systems and conclude that such technologies lack scientific reliability and pose significant risks of misclassification. Wright et al. [9] similarly critique emotion recognition technologies, highlighting issues of algorithmic opacity, cultural bias, and the absence of robust empirical validation. Khoei et al. [7] further demonstrate that emotional AI systems are particularly problematic in crime related applications due to contextual ambiguity and cross-cultural variability.

Minhas et al. [6][8] provide a more constructive perspective by examining AI-supported cognitive interview tools for victims and witnesses. Their studies show that AI can assist with memory recall and statement organization when designed as a supportive, non-judgmental aid, without introducing pressure or evaluative bias. This supports the use of AI for structural and organizational assistance rather than inference or assessment.

Overall, the literature indicates that AI applications focused on documentation, organization, and procedural support are more acceptable, reliable, and ethically defensible than systems that attempt to infer truthfulness, intent, or emotional state [1][2][3][5][7][9]. Despite this, there remains a clear lack of integrated frameworks that combine transcription, compliance monitoring, and post-interview analysis within a human-in-the-loop architecture. This paper addresses this gap by proposing a framework that positions AI as a tool for transparency and review, rather than as an authority in investigative decision-making.

Current Use of AI in Interrogation-Related Practices and Research Gaps:

Artificial intelligence is increasingly being used in various law enforcement processes that are related to, but do not fully support, structured criminal interrogations. Across different jurisdictions, agencies are experimenting with AI to improve documentation, increase efficiency, and manage large volumes of evidence. However, these applications remain fragmented, limited in scope, and largely detached from formal interrogation processes.

1. AI in Transcription of Investigative Interviews (Documentation)

Example : Norway / EU Context

Research on the use of automatic speech recognition (ASR) for transcribing investigative interviews highlights real world experimentation with AI to reduce burdensome manual transcription by police. Studies show that law enforcement agencies across Europe are exploring AI-based audio-to-text systems to improve efficiency and transparency. These systems aim to handle large volumes of interview recordings, extract text, and make interview content searchable and reviewable for investigative teams.[4]

2. Real-World Retrial AI Use Cases by Police (Supplemental Context)

Example : Body-Worn Camera Transcripts in Practice

Beginning in early 2024, several California police departments—including Fresno, San Mateo, East Palo Alto, and Campbell—launched pilot programs using *Draft One*, an AI-enabled tool developed by Axon to transcribe body worn camera audio and generate initial incident report drafts. The Fresno Police Department limited early use to misdemeanor calls to reduce report-writing time. By June 2025, the San Francisco Police Department

OPEN ACCESS

expanded testing to approximately 54 officers for low-severity cases, with mandatory human review of all AI-generated content. While these initiatives demonstrate growing use of AI for documentation support, civil liberties groups have raised concerns regarding accuracy, transparency, and procedural oversight [10].

3. Research on ASR Performance Limitations (Transcription Reliability)

Example: England & Wales Police Interviews

Academic work has evaluated how well commercial ASR services perform on police interview recordings, including challenges with background noise, accents, and transcription errors. Results indicate notable issues in automatic transcript accuracy, especially with regional accents or noisy audio — underscoring that current AI performance may not meet forensic standards without human review.[12]

4. Practical Tools for Investigative Transcription

Example: Commercial AI Transcription Services

AI-powered transcription platforms such as Sonix are marketed to law enforcement for converting audio/video interrogation data into searchable text quickly, enabling collaboration across agencies and tagging evidence for case management.[11]

Research Gap in above examples:

Transcription Reliability and Legal Validity

Although ASR tools automate transcription, their accuracy and reliability under real investigative conditions remain inconsistent. Clear standards for handling errors and preserving evidentiary value are still lacking.

Lack of Procedural Compliance Integration

Current AI uses in law enforcement are fragmented and do not actively support procedural safeguards during or after interrogations. There is little research on linking AI outputs with systematic compliance verification.

Limited Scope Beyond Transcription Quality

Existing studies focus primarily on transcript accuracy and overlook post-interview analytical needs. Functions such as consistency checks, timeline reconstruction, and comparative review remain underexplored.

Insufficient Human-in-the-Loop Oversight Structures

Most tools operate without a unified interface for transparent human review. Research rarely examines AI systems embedded within end-to-end interrogation workflows that prioritize accountability and oversight.

Comparison: Sonix, Verbit, and PolygrAI in Law Enforcement Contexts

Sonix AI : Automated transcription and processing of law enforcement audio/video into searchable, annotated text. Verbit AI : AI-assisted transcription with optional human review, integrated with workflow tools and case systems. PolygrAI : AI-driven interrogation and screening analytics including dynamic question support and risk evaluation.

Interpretation and Research Gap Relative to the Four Framework Pillars:

Documentation:

Sonix and Verbit both provide strong capabilities for automated transcription, producing searchable, timestamped text that reliably supports evidence review and reporting. They demonstrate that AI can assist documentation in law enforcement.

Gap: Both tools primarily address transcription and rudimentary summarization. They do not inherently link automated transcription to procedural compliance monitoring, nor do they offer integrated analytical review across multiple interview sessions as proposed in the framework.

Compliance Monitoring :

None of the tools currently offer automated compliance

checking—for example, monitoring whether rights advice occurred or whether required legal procedures were followed.

Gap: There is a lack of structured, AI-supported procedural adherence analysis within current solutions. **Post-Interview**

Analysis :

PolygrAI extends into behavioural analytics and adaptive interviewing, but these functions go beyond mere transcription into inference and risk assessments—raising potential ethical issues and lacking rigorous validation. Gap: There is no tool that consolidates transcript data into consistency checks, timeline reconstructions, or trend analytics under human review without venturing into contested inferential domains.

Human Oversight & Dashboard :

Sonix and Verbit both enable human editing and collaboration on transcripts, but do not provide a unified dashboard that integrates documentation, compliance indicators, and analytical outputs in a single interface. Gap: A centralized human-in-the-loop dashboard that consolidates all outputs for structured review and decision support is currently absent from these commercial offerings.

Research Objectives and Questions:

The primary objective of this study is to develop a conceptually grounded framework for AI-assisted criminal interrogations that positions artificial intelligence as a supportive tool for documentation, procedural oversight, and post-interview analysis rather than as an authority in investigative decision-making. Specifically, the study aims to:

1. Define defensible roles for AI in interrogation environments, limited to transcription, structured summarization, compliance monitoring, and analytical organization.

2. Examine how human-in-the-loop design can preserve investigator autonomy and mitigate

automation bias. 3. Establish structural boundaries between supportive AI functions and high-risk inferential applications such as emotion recognition and deception detection.

4. Provide a normative and architectural foundation to guide future research, system design, and policy development in ethically sensitive investigative contexts.

Research Questions

RQ1 – Documentation Integrity

RQ1: How can AI-assisted transcription and structured summarization be integrated into interrogation environments to enhance documentation accuracy without altering investigative dynamics?

RQ2 – Procedural Safeguards

RQ2: In what ways can AI-based compliance monitoring support adherence to legal and procedural safeguards during interrogations while preserving human authority?

RQ3 – Post-Interview Analytical Support

RQ3: How can post-interview AI analytics, such as consistency and timeline analysis, assist investigative review without introducing interpretive or inferential bias?

RQ4 – Human Oversight & Bias Control

RQ4: How does a human-in-the-loop design structure mitigate automation bias and maintain investigator decision autonomy in AI-assisted interrogation systems?

Research Methodology:

This study adopts a qualitative, literature-driven methodology to develop a conceptual framework for AI-assisted criminal interrogations. The aim is not system implementation, but to define how AI can be responsibly integrated while preserving human judgment, procedural fairness, and legal safeguards. This approach aligns with normative research in criminal justice AI that prioritizes ethics,

OPEN ACCESS   **Original Research Article**

accountability, and human rights [1][2].

A thematic synthesis was conducted across four domains: interrogation science, AI in law enforcement, legal and human rights scholarship, and AI-supported interviewing technologies. Foundational work on investigative interviewing informed best practices in non-coercion and documentation integrity [3]. Legal and policy literature guided principles of transparency and oversight [1][2]. Research on automatic speech recognition in investigative contexts supported the legitimacy of transcription use [4], while studies on AI-assisted cognitive interviews highlighted supportive, non-influential roles for AI [6][8]. Critical literature on emotion recognition and deception detection was reviewed to identify scientific and ethical limitations [5][7][9].

Framework development followed an interpretive, design-oriented process, organizing key themes— documentation accuracy, compliance monitoring, post-interview review, and human oversight—into functional layers. AI is confined to supportive roles, with all interpretive authority retained by human investigators to mitigate automation bias and algorithmic opacity [1][2][9].

Model Selection and Algorithmic Considerations:

Model Selection

The framework intentionally avoids reliance on a single unified model. Instead, it adopts a modular, task-specific approach in which each algorithm is selected based on its functional role, interpretability, and legal suitability. Models are chosen only for documentary, structural, and procedural support tasks, such as transcription, summarization, compliance verification, semantic comparison, and entity extraction. Inferential and predictive models, including emotion recognition, deception detection, and guilt prediction, are explicitly excluded due to their contested scientific validity and ethical risks. This selective and bounded approach ensures that artificial intelligence remains an assistive tool rather than a decision-making authority, with all interpretive judgment retained by human investigators.

Proposed Human-in-the-Loop Framework:

This framework positions artificial intelligence as a supportive infrastructure within criminal interrogations, limited to documentation, procedural oversight, and post-interview organization of information. AI outputs are designed to be reviewable and non-authoritative, while all interpretive and decision-making responsibility remains with human investigators.

The framework is organized into four functional layers, each addressing a specific research gap while avoiding inferential or judgment-oriented uses of AI.

1. **Documentation Layer (Transcription and Structuring)**

   AI supports the accurate conversion of recorded interviews into structured textual records and organizes content for later review; it does not interpret, prioritize, or alter statements, nor does it generate evaluative summaries.

2. **Procedural Safeguard Layer (Compliance Indicators)**

   AI flags the presence of procedural markers, such as rights notification or interview boundaries, to support oversight; it does not assess legal adequacy, determine compliance outcomes, or intervene during questioning.

3. **Post-Interview Analytical Review Layer (Consistency, Timeline, Trends)**

   AI assists in organizing interview data across sessions by highlighting semantic variation, reconstructing timelines, and visualizing observable interaction patterns; it does not infer deception, credibility, emotional state, or intent.

4. **Human Oversight Layer (Human-in-the-Loop Control)**

   All AI outputs are presented through a dashboard-oriented interface for explicit human review,

validation, or rejection; no output is treated as definitive or operational without human confirmation.

Across all layers, AI functions solely as an aid for recording, organization, and review, ensuring transparency and accountability without influencing investigative judgment.

Conceptual Flow of proposed framework

The flow is deliberately linear and non-intrusive:

Interview → Documentation → Compliance → Post-Interview Analytics → Human Oversight
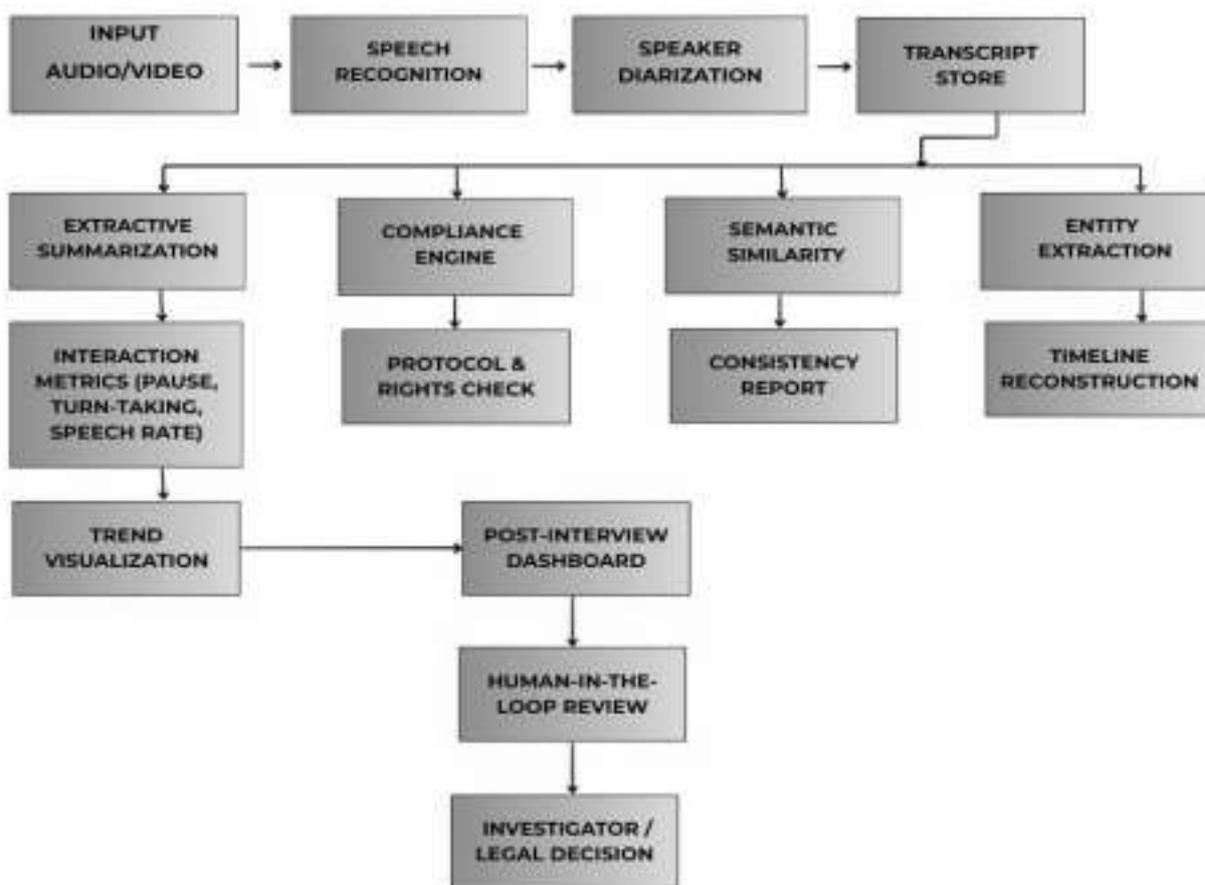

1. Proposed framework

Figure 1 Block diagram illustrating the algorithmic flow of the proposed human-in-the-loop framework for AI assisted criminal interrogations.

Note: No Emotion Detection, No Lie Detection, No Guilt Prediction

Real-time interviewing is isolated from analytical processing to prevent automation bias or coercive influence. Post interview outputs—including transcripts, compliance flags, consistency indicators, and timelines—are consolidated into a single dashboard to support structured, accountable review by investigators, courts, and defense stakeholders. [4][6][8].

Conclusion and Future Work:

Conclusion:

This study presents a human-in-the-loop framework for AI-assisted criminal interrogations that deliberately restricts artificial intelligence to **supportive, non-inferential roles**. By prioritizing transcription, procedural indicators, and post-interview analytical organization, the framework directly responds to documented concerns surrounding inferential AI applications—such as emotion recognition, deception detection, and predictive judgments—which remain scientifically contested and ethically problematic in criminal justice contexts [1][5][7][9].

Grounded in interrogation science, legal scholarship, and critical AI research, the framework emphasizes documentation integrity, procedural transparency, and preservation of human investigative authority [2][3]. The separation of interview capture from post-interview analysis is intended to mitigate automation bias and reduce the risk of undue influence during questioning, consistent with prior findings on investigative interviewing and AI-assisted documentation [1][3][4]. Mandatory human review across all analytical stages further reinforces accountability and due process in legally sensitive environments [2][9].

The primary contribution of this work lies in **boundary definition rather than technical innovation**. By clearly distinguishing defensible assistive uses of AI from high-risk inferential applications, the paper provides a normative and structural foundation for responsible AI integration in criminal interrogation settings [1][2][5].

Future Work:

Future research should focus on validating, contextualizing, and operationalizing the proposed framework through four key directions:

1. **Empirical Validation**

   Controlled simulations or mock interrogation studies should evaluate the framework's impact on transcription accuracy, procedural oversight, and post-interview review efficiency, particularly in comparison to manual documentation practices [4][6].

2. **Human Factors and Interface Design**

   Usability studies are needed to assess how investigators and legal professionals interpret dashboard-based AI outputs, including effects on trust calibration, cognitive load, and decision-making workflows [6][8].

3. **Jurisdictional Adaptation**

   Procedural indicators and review mechanisms should be adapted to jurisdiction-specific legal requirements in collaboration with legal scholars and practitioners to ensure doctrinal alignment and admissibility [2][4].

4. **Privacy and Evidence Integrity**

   Future work should examine secure system architectures, audit logging, and access control mechanisms to protect sensitive interview data and preserve evidentiary integrity across the investigative lifecycle [4].

Together, these directions support the transition from conceptual framing to empirically grounded, ethically responsible AI-assisted interrogation systems that enhance transparency and oversight while preserving human judgment and legal safeguards [1][2][9].

References:

1. R. Berk et al. *"Artificial Intelligence, Predictive Policing, and Risk Assessment for Law*

Enforcement." Annual Review of Criminology (2021). https://doi.org/10.1146/annurev-criminol-051520-012342.

2. *Aleš Završnik et al. "Criminal justice, artificial intelligence systems, and human rights." ERA Forum, 20 (2020): 567 - 583.*
   https://doi.org/10.1007/s12027-020-00602-0.

3. *C. Meissner et al. "PROTOCOL: Interview and Interrogation Methods and their Effects on Investigative Outcomes." Campbell Systematic Reviews, 6 (2010): 1-17. https://doi.org/10.1002/cl2.68.*

4. *Radina Stoykova et al. "The AI Act in a law enforcement context: The case of automatic speech recognition for transcribing investigative interviews." Forensic Science International: Synergy, 9 (2024).*
   https://doi.org/10.1016/j.fsisyn.2024.100563.

5. *Yi-Chang Wu et al. "The use of artificial intelligence in interrogation: lies and truth." IAES International Journal of Robotics and Automation (IJRA) (2023).*
   https://doi.org/10.11591/ijra.v12i4.pp332-340.

6. *Rashid Minhas et al. "Protecting victim and witness statement: examining the effectiveness of a chatbot that uses artificial intelligence and a cognitive interview." AI & SOCIETY, 37 (2021): 265 - 281. https://doi.org/10.1007/s00146-021-01165-5.*

7. *T. T. Khoei et al. "A survey of Emotional Artificial Intelligence and crimes: detection, prediction, challenges and future direction." Journal of Computational Social Science, 7 (2024): 2359 - 2402. https://doi.org/10.1007/s42001-024-00313-3.*

8. *Minhas, R., Elphick, C. & Shaw, J. Protecting victim and witness statement: examining the effectiveness of a chatbot that uses artificial intelligence and a cognitive interview. AI & Soc 37, 265–281 (2022).*
   https://doi.org/10.1007/s00146-021-01165-5

9. *James Wright et al. "Suspect AI: Vibraimage, Emotion Recognition Technology and Algorithmic Opacity." Science, Technology and Society, 28 (2020): 468 - 487.*
   https://doi.org/10.1177/09717218211003411.

10. *GovTech, California Police Use AI to Transcribe Body Cam Videos, December 9, 2024; and San Francisco Testing AI to Write Police Reports, June 11, 2025.*
    https://www.govtech.com/artificial intelligence/california-police-use-ai-to-transcribe-body-cam-videos

11. *Sonix AI for Investigators:*
    https://sonix.ai/en/legal/investigators

12. *Verbit AI Transcription for Law Enforcement: https://verbit.ai/industries/legal/law-enforcement/ 13. PolygrAI Law Enforcement Use Case: https://polygr.ai/law-enforcement-use-case/*