FROM SPEECH RECOGNITION TO ALGORITHMIC TRIAGE: HOW POST-9/11 INTELLIGENCE AUTOMATION RECONFIGURED POWER, BIAS, AND ACCOUNTABILITY

Utkarsh Srivastava
Department of Economics
Birla Institute of Technology and Science (BITS) Pilani, Hyderabad Campus
Hyderabad, Telangana, India
f20240633@hyderabad.bits-pilani.ac.in

Rahul D. Ray
Department of Electrical and Electronics Engineering
Birla Institute of Technology and Science (BITS) Pilani, Hyderabad Campus
Hyderabad, Telangana, India
f20242213@hyderabad.bits-pilani.ac.in

Rishi Mohapatra
Department of Chemical Engineering
Birla Institute of Technology and Science (BITS) Pilani, Hyderabad Campus
Hyderabad, Telangana, India
f20240796@hyderabad.bits-pilani.ac.in

ABSTRACT

Post-9/11 security reforms normalized large-scale automated surveillance by reframing intelligence failure as a problem of data integration. In parallel, advances in artificial intelligence, particularly speech recognition and voice biometrics rendered spoken communication computable at population scale. This paper argues that voice-based AI operates as algorithmic triage: upstream systems that probabilistically filter, rank, and render speech intelligible prior to human judgment. We formalize algorithmic triage as an epistemic infrastructure with identifiable stages and failure modes, and show why voice is a uniquely powerful and dangerous modality, entangling identity, behavior, and cultural difference in a single signal. We further propose a voice-specific sociotechnical audit framework as a work-in-progress research agenda. We argue that algorithmic triage erodes conditions of positive peace by normalizing perpetual suspicion and shifting accountability away from contestable human institutions. Addressing these dynamics requires not only ethical critique, but methodological tools for interrogating how voice AI is embedded within security infrastructures.

1 INTRODUCTION

The attacks of September 11, 2001 catalyzed a profound institutional and epistemological shift in security governance. Beyond expanding surveillance, the post-9/11 reforms fundamentally reconstituted the intelligence problem from one of human analysis and political judgment to one of computational data processing (Lyon, 2003; Webb, 2007). This created an infrastructural demand for artificial intelligence, not merely as a tool for scaling, but as a new epistemic foundation capable of generating actionable intelligence from previously unintelligible data flows. Among these modalities, voice occupies a distinctive position: intimate, difficult to withhold, socially marked, and increasingly treated as both biometric identifier and behavioral signal (Jansen et al., 2021; Leix Palumbo & Prey, 2024).

This paper argues that post-9/11 intelligence automation reconfigured power not primarily through scale, but through algorithmic triage: AI systems that pre-emptively determine which signals are rendered intelligible, prioritized, or discarded. While existing work critiques surveillance expansion broadly, fewer studies examine how specific AI modalities enact this triage in practice. Unlike other biometric modalities, voice is continuously produced through social interaction, rendering it especially susceptible to repurposing within preemptive security infrastructures and making voice-based AI a critical site for examining how algorithmic governance intersects with peace, trust, and democratic participation.

2 METHODOLOGICAL APPROACH

Our methodology employs a tripartite, infrastructure-focused analysis to dissect voice-based algorithmic triage. We first construct a theoretical model of algorithmic triage by integrating infrastructure studies with intelligence studies, positioning it as a new epistemic layer within security governance (Lyon, 2003; Monahan & Regan, 2012). Second, we perform a technical genealogy, mapping the co-evolution of speech processing architectures (from HMM-GMM to end-to-end deep learning) and their re-framing as security instruments Nassif (2019); Fendji (2022). This reveals how technical milestones in speaker diarization and emotion recognition enabled new inferential logics for security. Third, we operationalize critique by developing a sociotechnical audit framework, translating structural concerns into discrete, investigable dimensions that can be applied under conditions of institutional opacity.

This approach aligns with prior calls to examine not only algorithmic outputs, but the institutional contexts and governance arrangements within which such systems operate (Katzenbach & Ulbricht, 2019; Margulies, 2016). The audit framework presented here is intended as work-in-progress research, offering a structured method that can be applied to specific voice-based security systems in future empirical studies.

Importantly, this methodological orientation treats critique itself as a form of research practice. By specifying audit dimensions, data sources, and evaluative questions, the framework enables systematic inquiry even in contexts where direct access to models or proprietary datasets is limited. This is particularly relevant for studying security and intelligence systems, where opacity and secrecy are themselves central features of governance.

3 THE POST-9/11 SURVEILLANCE TURN

Post-9/11 security reforms were driven by the belief that intelligence failures stemmed from insufficient information sharing rather than flawed assumptions or institutional incentives. The creation of the Department of Homeland Security and the proliferation of fusion centers exemplify this shift toward centralized, data-driven governance (Monahan & Regan, 2012). Fusion centers integrate data from federal agencies, local law enforcement, and private-sector actors. Monahan and Regan describe them as "zones of opacity," where heterogeneous datasets are algorithmically correlated and circulated without clear accountability (Monahan & Regan, 2012). In this environment, correlation increasingly substitutes for interpretation.

This institutional shift created demand for automated prioritization. As data volumes exceeded human analytic capacity, algorithmic systems became infrastructural necessities rather than optional tools (Margulies, 2016). These systems enabled a temporal shift from reactive intelligence to preemptive security, oriented toward probabilistic future risk rather than past action (Lyon, 2003).

4 ALGORITHMIC TRIAGE AS EPISTEMIC INFRASTRUCTURE

Algorithmic triage functions as an upstream epistemic gatekeeper: a suite of AI processes that probabilistically filter, rank, and structure raw sensory data into intelligible, actionable categories before any human analyst engages. Its power derives from being population-scale, continuous, and architecturally opaque. For voice, we specify a four-stage technical pipeline: (1) Signal Capture/Enhancement, where noise suppression and voice activity detection already discard data deemed non-informative; (2) Feature Disentanglement, where neural encoders separate speaker identity from linguistic content

and para-linguistic features; (3) Multi-Task Inference, where concurrent models classify identity, language, accent, and prosody, outputting confidence-weighted labels; and (4) Priority Scoring, where a meta-model synthesizes these labels against a risk heuristic to assign a ranking for analyst review. Bias is systematically embedded and amplified at each stage for instance, through acoustic models tuned to dominant dialects making errors non-transparent and non-contestable (Zweig et al., 2018).

This differs fundamentally from analyst-driven prioritization. Whereas human triage is deliberative and contestable, algorithmic triage embeds assumptions into technical infrastructure. As Margulies notes, intelligence agencies increasingly rely on machine- based surveillance to narrow the field of analysis in advance (Margulies, 2016). Understanding algorithmic triage as infrastructure reveals how power is exercised through design choices rather than explicit decisions. By embedding prioritization logics upstream, such systems constrain interpretation before analysts engage with data, helping explain why accountability mechanisms focused solely on human decision-makers fail to address systemic harms

5 From Speech Recognition to Voice-Based Inference

Advances in automatic speech recognition made voice computable at scale. The shift from statistical models to deep neural networks and transformer-based architectures improved robustness in noisy and multilingual environments (Nassif, 2019; Fendji, 2022; Sharrab, 2025).

These advances enabled a transition from transcription to inference. Voice biometrics systems increasingly incorporate "soft biometrics" such as accent, language, or perceived emotion (Jansen et al., 2021). As a result, voice functions simultaneously as content, identity, and behavioral signal.

Because voice encodes cultural and social difference, its use in triage systems amplifies existing inequalities. Leix Palumbo and Prey show how voice analytics blur state and corporate surveillance, complicating accountability and consent (Leix Palumbo & Prey, 2024).

These dynamics are especially salient in multilingual and transnational contexts, where accent and language variation are often conflated with uncertainty or risk. As voice-based systems scale across jurisdictions, the technical handling of linguistic difference becomes inseparable from questions of belonging, legitimacy, and surveillance exposure.

6 Militarization Through Infrastructural Integration

AI militarization operates as an infrastructural absorption, seamlessly integrating civilian research and commercial platforms into the intelligence lifecycle (Hoadley & Lucas, 2018; Sayler, 2020). For voice AI, this "diffuse militarization" occurs through three mechanisms: procuring commercial speaker recognition APIs, funding academic research on problems like low-resource language detection, and developing standards to ingest consumer voice data into security fusion centers. This pathway bypasses oversight, as capabilities are quietly operationalized within existing data pipelines rather than as new weapon systems (Garcia, 2019), blurring the boundary between the consumer tech sector and the security state.

Operating deep within data pipelines, these systems render their surveillance role invisible. Linguistic difference is transformed into a risk signal, disproportionately affecting marginalized populations (Jansen et al., 2021; Katzenbach & Ulbricht, 2019). Thus, militarization functions as a gradual **alignment process**, where everyday tools like voice AI are absorbed into security infrastructures without clear political authorization, evading democratic oversight.

7 Toward a Sociotechnical Audit of Voice-Based Security Systems

To move beyond critique, we propose a voice-specific sociotechnical audit framework as work-in-progress research. Unlike generic fairness audits focused narrowly on model performance, this framework interrogates how voice-based systems are institutionally embedded and how their outputs shape downstream governance practices (Katzenbach & Ulbricht, 2019; Caroleo, 2025). The framework examines four interrelated dimensions:

(1) Data Provenance and Representation. This dimension examines whose voices are captured, under what conditions, and with what forms of consent. Prior research shows that voice datasets often underrepresent linguistic diversity while treating accent and language variation as noise or risk signals (Jansen et al., 2021; Leix Palumbo & Prey, 2024). Auditing data provenance involves analyzing demographic coverage in training corpora and reviewing data collection and retention policies.

(2) Model Architecture and Inferential Scope. This dimension investigates the technical instantiation of inference. It requires auditing which specific model architectures (e.g., x-vector extractors, wav2vec 2.0, emotion classification CNNs) are deployed, what their intended and latent inferential capabilities are (e.g., can accent classification be derived from a speaker verification model?), and how performance disparities manifest not just as error rate variances but as fundamentally different confusion matrices across linguistic and demographic groups Jansen et al. (2021); Nassif (2019); Fendji (2022).

(3) Workflow Integration and Human-in-the-Loop Dynamics. Moving beyond the simplistic "upstream" metaphor, this audit traces the precise points of human-algorithm interaction. It examines how priority scores and confidence intervals are visually presented to analysts, whether analysts can access model metadata or training characteristics, and how institutional protocols (like "escalate all high-priority flags") effectively delegate discretionary power to the algorithmic system, making human oversight a procedural formality (Zweig et al., 2018; Margulies, 2016).

(4) Outcomes and Distributional Impact. This dimension examines how voice-based triage redistributes suspicion and visibility across populations, with prior work indicating disproportionate impacts on marginalized communities, particularly migrants and speakers of non-dominant dialects (Jansen et al., 2021; Garcia, 2019). Outcome analysis focuses on patterns of false positives, escalation, and downstream consequences.

7.1 Illustrative Application and Research Feasibility

To demonstrate the framework's application, we sketch an audit of a documented system: the Speaker Identification Integrated Project (SiiP) (Jansen et al., 2021). A full audit would involve: (1) Provenance: Using procurement documents and API specifications to reconstruct the training data ecology, searching for clauses on demographic representativeness. (2) Inference: Analyzing published accuracy metrics for subgroup performance (e.g., Equal Error Rate disparities for speakers of Maghrebi French vs. Metropolitan French) and testing for unintended attribute leakage. (3) Integration: Examining analyst training materials for SiiP to see if confidence score interpretation is taught, and if workflows allow for overriding algorithmic priority rankings. (4) Impact: Correlating deployment zones with demographic data to identify potential disproportionate targeting patterns. This structured approach moves from technical documentation to actionable governance questions.

Documentary Micro-Audit (Proof of Concept). To demonstrate feasibility, we conducted a limited documentary audit of publicly described design characteristics of SiiP, focusing on audit dimension (1) Data Provenance and Representation (Jansen et al., 2021). Although SiiP emphasizes interoperability and large-scale voice matching across national contexts, publicly available documentation does not indicate any mandated evaluation of system performance across dialectal, accentual, or sociolinguistic subgroups. Linguistic variation is instead treated as a technical challenge to be normalized rather than a potential source of discriminatory impact (Jansen et al., 2021).

From an audit perspective, this absence is analytically significant: it demonstrates how bias can arise not only from model behavior, but from institutional design choices that render certain forms of difference invisible. Although limited in scope, this documentary analysis shows how meaningful audit insights can be derived from publicly accessible materials alone, underscoring the value of sociotechnical methods for interrogating otherwise inaccessible security systems and establishing a baseline for more extensive empirical audits as access permits.

8 Conclusion

Algorithmic triage undermines positive peace by normalizing perpetual suspicion and reframing communication as risk (Lyon, 2003). Peace-oriented responses therefore need not focus on technical optimization: non-deployment, researcher refusal, and sociotechnical audits establish boundaries

on legitimate use and create empirical pathways to challenge infrastructural harm (Garcia, 2019). This paper reframes voice-based surveillance as algorithmic triage an epistemic infrastructure that redistributes power, visibility, and accountability upstream of human judgment and proposes a concrete audit methodology to move beyond ethical critique. Such work is essential if peace is to be understood as democratic accountability, social trust, and freedom from surveillance-induced harm (Lyon, 2003; Katzenbach & Ulbricht, 2019).

REFERENCES

Laura Caroleo. Do we need a voice methodology? *Societies*, 15(9):241, 2025.

Jean Louis K E et al. Fendji. Automatic speech recognition using limited vocabulary. *Applied Artificial Intelligence*, 36(1), 2022.

Eugenio Garcia. The militarization of artificial intelligence. 2019.

Daniel S Hoadley and Nathan J Lucas. Artificial intelligence and national security, 2018.

Fieke Jansen, Javier Sánchez-Monedero, and Lina Dencik. Biometric identity systems in law enforcement and the politics of (voice) recognition. *Big Data & Society*, 8(2), 2021.

Christian Katzenbach and Lena Ulbricht. Algorithmic governance. *Internet Policy Review*, 8(4), 2019.

Daniel Leix Palumbo and Robert Prey. Sounding out voice biometrics. *Big Data & Society*, 11(4), 2024.

David Lyon. *Surveillance after September 11*. Polity, 2003.

Peter Margulies. Surveillance by algorithm. *Florida Law Review*, 68:1045, 2016.

Torin Monahan and Priscilla M Regan. Zones of opacity: Data fusion in post-9/11 security organizations. *Canadian Journal of Law and Society*, 27(3):301–317, 2012.

Ali Bou et al. Nassif. Speech recognition using deep neural networks. *IEEE Access*, 7:19143–19165, 2019.

Kelley M Sayler. Artificial intelligence and national security. Technical report, 2020.

Yousef O et al. Sharrab. Advancements in speech recognition. *IEEE Access*, 2025.

Maureen Webb. *Illusions of Security: Global Surveillance and Democracy in the Post-9/11 World*. City Lights Books, 2007.

Katharina A Zweig, Georg Wenzelburger, and Tobias D Krafft. On chances and risks of security related algorithmic decision making systems. *European Journal for Security Research*, 3(2): 181–203, 2018.