TRIDENT: A Redundant Architecture for Caribbean-Accented Emergency Speech Triage

Galbraith, E., Sutherland, C., and Morgan, D.

SMG Labs Research Group

December 10, 2025

Abstract

Emergency speech recognition systems exhibit systematic performance degradation on non-standard English varieties, creating a critical gap in services for Caribbean populations. We present TRIDENT (**T**ranscription and **R**outing **I**ntelligence for **D**ispatcher-**E**mpowered **N**ational **T**riage), a three-layer dispatcher-support architecture designed to structure emergency call inputs for human application of established triage protocols (the Emergency Severity Index for routine operations and START for mass casualty events), even when automatic speech recognition fails.

The system combines Caribbean-accent-tuned ASR, local entity extraction via large language models, and bio-acoustic distress detection to provide dispatchers with three complementary signals: transcription confidence, structured clinical entities, and vocal stress indicators. **Our key insight is that low ASR confidence, rather than representing system failure, serves as a valuable queue prioritization signal—particularly when combined with elevated vocal distress markers indicating a caller in crisis whose speech may have shifted toward basilectal registers.** A complementary insight drives the entity extraction layer: trained responders and composed bystanders may report life-threatening emergencies without elevated vocal stress, requiring semantic analysis to capture clinical indicators that paralinguistic features miss.

We describe the architectural design, theoretical grounding in psycholinguistic research on stress-induced code-switching, and deployment considerations for offline operation during disaster scenarios. This work establishes a framework for accent-resilient emergency AI that ensures Caribbean voices receive equitable access to established national triage protocols. Empirical validation on Caribbean emergency calls remains future work.

Keywords: automatic speech recognition, Caribbean English, emergency dispatch, vocal stress detection, creole continuum, edge computing, Emergency Severity Index, dispatcher support, position paper

1 Introduction

When a caller dials emergency services during a crisis, the interaction between human distress and automated systems creates a critical dependency on speech recognition accuracy. Modern automatic speech recognition (ASR) systems, however, exhibit well-documented performance disparities across demographic groups [16]. For Caribbean English speakers—a population of over 40 million across the Anglophone Caribbean and diaspora—these disparities compound with a linguistic phenomenon: under acute stress, speakers tend to shift toward basilectal (more creole-heavy) speech registers, precisely the varieties on which ASR systems perform worst.

1.1 Clinical Context: Established Triage Protocols

Caribbean health ministries have adopted internationally-validated triage protocols to standardize emergency response: the Emergency Severity Index (ESI) for routine operations and START (Simple Triage and Rapid Treatment) for mass casualty events. These protocols assume dispatchers can accurately capture caller information—an assumption that fails systematically when ASR systems cannot reliably transcribe Caribbean speech.

1.2 TRIDENT: Dispatcher-Empowered Architecture

This paper presents TRIDENT (Transcription and Routing Intelligence for Dispatcher-Empowered National Triage), an architectural framework designed not to replace established triage protocols, but to ensure Caribbean-accented callers receive equitable access to them. Rather than attempting to eliminate ASR errors on Caribbean speech—an unrealistic goal given current technology—we build a **dispatcher-support system** that remains functional when transcription fails. **We frame this work as a position paper and system proposal**, establishing theoretical foundations and design rationale while acknowledging that end-to-end empirical validation on Caribbean emergency calls remains future work.

Our central contribution is a three-layer dispatcher-support framework that provides human dispatchers with structured inputs for protocol application. The system generates three complementary signals:

1. **Transcription confidence:** Flags unreliable transcripts so dispatchers know to listen directly to call audio rather than relying on text alone.
2. **Structured entity extraction:** Extracts clinical indicators needed for ESI/START application—location, mechanism of injury, breathing status, number of persons affected, presence of vulnerable populations—even from partial or degraded transcriptions.
3. **Bio-acoustic distress detection:** Provides a novel signal not currently captured by standard triage protocols—physiological stress markers derived from vocal acoustics that indicate caller crisis state independent of transcript content.

1.3 Key Insights

Two complementary insights motivate this design:

1. **Content beyond voice:** Trained first responders, medical professionals, and composed bystanders may report life-threatening emergencies without elevated vocal stress. Semantic extraction of clinical entities captures information that paralinguistic features alone would miss—ensuring that "children trapped in burning building," spoken calmly, provides dispatchers with the structured data needed for appropriate ESI or START classification.
2. **Uncertainty as prioritization signal:** Low ASR confidence, rather than representing system failure, serves as a valuable indicator for queue prioritization—particularly when combined with elevated vocal distress markers indicating a caller in crisis whose speech may have shifted toward basilectal registers. This reframes accent-induced transcription errors from bugs into features that correlate with genuine caller distress, ensuring these callers receive priority human attention for proper triage assessment.

1.4 Addressing Gaps in Emergency AI

The architecture addresses four gaps in existing emergency AI systems:

1. **Cloud dependency** with accent-agnostic ASR

2. **Text-only analysis** ignoring paralinguistic stress signals
3. **Dialect blindness** to stress-induced register shifting
4. **Infrastructure fragility** during disaster scenarios

Critically, TRIDENT addresses these gaps while **respecting the clinical authority of established protocols**. The system structures inputs and prioritizes dispatcher queues, but triage decisions remain with trained human professionals applying Ministry of Health-mandated frameworks.

1.5 Scope and Generalizability

Note on stress-induced register shift: While we focus on Caribbean creole continua, the phenomenon of dialect reversion under cognitive load is not unique to this population. The inhibitory control model of bilingual processing [14] and research on the Lombard effect (speech modifications in noisy environments) suggest our framework may generalize to other bidialectal populations worldwide. Caribbean emergency services serve as our motivating case study, but the architectural principles apply broadly to any context where:

- Speakers use dialectal varieties underrepresented in ASR training data
- Stress may induce shifts toward non-standard registers
- Established triage protocols require accurate capture of caller information
- Infrastructure resilience is critical for emergency response

2 Related Work

The proposed dispatcher-support architecture draws on and extends research across five domains: ASR for accented and low-resource speech varieties, artificial intelligence (AI) in emergency dispatch (including the clinical triage protocols AI systems must support), vocal stress detection, dialect reversion under cognitive load, and edge computing for disaster resilience. We review each in turn, identifying the gaps that motivate TRIDENT's three-layer design.

2.1 The Accent Gap in Automatic Speech Recognition

Modern ASR systems exhibit systematic performance degradation on non-standard English varieties—a disparity with serious implications for equitable access to voice-enabled services. Koenecke et al. [16] conducted the seminal quantitative study, evaluating five major commercial ASR systems across racial demographics. Their findings were stark: word error rates (WER) averaged 0.35 for Black speakers compared to 0.19 for White speakers, with 23% of Black speaker audio producing WER exceeding 0.50—functionally unusable transcription—compared to just 1.6% for White speakers. Critically, the researchers traced these disparities to acoustic models rather than language models, as the performance gap persisted even on identical phrases.

The Edinburgh International Accents of English Corpus (EdAcc) benchmark extends this analysis to global accent variation [22]. Testing revealed that OpenAI's Whisper-large model achieved 19.7% WER on EdAcc compared to just 2.7% on LibriSpeech test-clean—a seven-fold performance degradation on accented speech. The study specifically identified Jamaican English among the accents with highest error rates, directly validating concerns about Caribbean speech recognition.

Caribbean English remains especially underserved despite representing millions of speakers. Madden et al. [19] developed the first substantial Jamaican Patois speech corpus (42.58

hours) and derived scaling laws for Whisper performance on this variety. Their results are instructive: pre-trained Whisper Large achieved 89% WER on Patois—functionally useless—while fine-tuned Whisper Medium reduced this to 30% WER. Notably, fine-tuned Whisper Tiny outperformed non-fine-tuned Whisper Large, demonstrating that domain-specific data matters more than model size for underrepresented varieties. Their scaling law ($\text{WER} = 158.06 \times \text{M}^{-0.255} \times \text{D}^{-0.269}$) reveals that dataset increases yield greater gains than model scaling for this population, informing our choice of Whisper Medium with Caribbean-specific fine-tuning.

2.2 AI-Assisted Emergency Dispatch

Emergency services worldwide are exploring AI-powered speech recognition and natural language processing to improve call handling efficiency and support triage accuracy. Importantly, these systems are designed to *augment* human dispatchers applying established clinical protocols, not to replace professional judgment. Understanding the clinical frameworks that AI systems must support is essential context for evaluating their design.

2.2.1 Clinical Triage Protocols: The Gold Standards

Emergency Severity Index (ESI). Developed by the Agency for Healthcare Research and Quality, ESI is a five-level acuity scale widely used in the United States and internationally [10]. The protocol stratifies patients from Level 1 (immediate lifesaving intervention required) to Level 5 (no resources needed), based on acuity assessment and anticipated resource utilization. Jamaica's Ministry of Health implemented ESI across all 19 public hospital emergency departments in 2016 [11].

However, a 2020 evaluation found poor interrater reliability between Jamaican practitioners and ESI experts, with triage note quality, completeness of vital sign assessment, and high staff attrition identified as key challenges [11]. These implementation difficulties motivate TRIDENT's focus on structured entity extraction to support dispatcher protocol application.

START Protocol. For mass casualty events such as hurricanes that regularly affect the Caribbean, the START (Simple Triage and Rapid Treatment) protocol provides rapid four-category sorting: BLACK (deceased/expectant), RED (immediate), YELLOW (delayed), and GREEN (walking wounded). The ESI handbook explicitly notes that ESI should *not* be used during mass casualty incidents, where START or similar rapid triage systems are appropriate [10].

Any AI system for emergency dispatch must therefore be evaluated not on whether it makes "better" triage decisions than these validated protocols, but on whether it improves the *inputs* available to human professionals applying them. Our evaluation of existing systems and the design of TRIDENT is guided by this framing: AI as protocol enabler rather than protocol replacement.

2.2.2 Current AI Systems: Capabilities and Limitations

The Emergency Calls Assistant (ECA) framework represents current academic state-of-the-art, achieving 92.7% accuracy in emergency classification using SVM with linear kernel on textual features [2]. The system operates in two phases—speech-to-text conversion followed by NLP classification—and compares favorably against commercial platforms including RapidSOS, Corti, and AlertGO. Like TRIDENT, ECA is designed to support dispatcher decision-making by providing structured classification of caller reports.

However, critical examination reveals systematic gaps. ECA relies on Google Cloud Speech-to-Text API with no offline capability or accent adaptation, assumptions that fail for Caribbean deployment contexts. The system processes only transcribed text, ignoring paralinguistic stress markers that may indicate caller distress even when words are unclear. Furthermore, due to

privacy restrictions on real emergency recordings, ECA was trained on synthetic datasets, raising questions about generalization to actual crisis communications where callers exhibit genuine distress.

Clinical validation studies demonstrate AI's potential for dispatcher support while highlighting implementation challenges. Blomberg et al. [3, 4] evaluated the Corti AI system for cardiac arrest detection, finding that the ML system achieved 84.1% sensitivity compared to dispatchers' 72.5%, with faster time-to-recognition (44 seconds versus 54 seconds median). Critically, Corti operates as a *decision-support tool*: it alerts dispatchers to potential cardiac arrests, but the dispatcher makes the final determination and applies appropriate protocols. However, a subsequent randomized clinical trial found no significant improvement in dispatcher recognition when supported by ML alerts, suggesting that human-AI teaming requires careful interface design beyond raw model performance.

A scoping review of 106 AI studies in prehospital emergency care identified underutilization of multimodal inputs as a key gap [6]. No reviewed system integrated audio-based stress detection with text classification. The review also noted the absence of systems designed for infrastructure-independent operation, a critical limitation for disaster response scenarios where communication networks are degraded precisely when emergency services are most needed.

2.2.3 Gaps Relevant to Caribbean Deployment

Three specific gaps emerge from this literature that motivate TRIDENT's design:

1. **Accent and dialect adaptation:** No existing system addresses the systematic ASR performance degradation on Caribbean English varieties, nor the stress-induced register shifting that characterizes bidialectal speech communities under crisis conditions.

2. **Multimodal distress detection:** While vocal stress research is extensive (Section **??**), no deployed emergency AI system incorporates bio-acoustic analysis as a parallel signal to text-based classification—missing critical information when ASR fails.

3. **Infrastructure resilience:** Cloud-dependent architectures assume reliable internet connectivity, an assumption that fails catastrophically during the hurricanes, earthquakes, and floods that drive emergency call surges in Caribbean contexts.

TRIDENT addresses these gaps while maintaining the fundamental principle that AI systems should *empower* dispatchers to apply established protocols (ESI, START) more effectively, not replace clinical judgment with algorithmic decision-making.

2.3 Vocal Stress Detection

The bio-acoustic analysis layer of our system builds on extensive research establishing acoustic correlates of psychological stress. A systematic review analyzing 38 peer-reviewed studies found that fundamental frequency (F0) is the most consistent stress marker, with 15 of 19 studies reporting significant mean F0 increases under stress conditions [24]. Intensity and amplitude increases showed similarly consistent patterns, while speech rate, jitter, and shimmer produced heterogeneous results across studies.

It is important to note methodological heterogeneity in this literature. While F0 elevation is the most replicated finding, some studies report null or contradictory results depending on stress type (acute vs. chronic), measurement methodology, and population characteristics. Hansen and Patil [15] found that certain stress conditions produce F0 *decreases* in some speakers, particularly under conditions of extreme fatigue or hopelessness. Our system design accounts for this by using F0 as one component of a multi-feature distress score rather than a sole indicator.

Research specifically examining emergency communications provides direct validation for our approach. Van Puyvelde et al. [27] analyzed real-life emergency recordings including cockpit voice recorders and 911 calls, documenting F0 increases from 123.9 Hz to 200.1 Hz during life-threatening emergencies—a 62% increase. F0 range expanded dramatically from 124.2 Hz to 297.3 Hz. Interestingly, jitter *decreased* during emergency stress, contrary to intuition, providing an additional discriminative feature. These findings directly inform our distress detection thresholds.

Studies of actual emergency call centers demonstrate both the promise and limitations of acoustic stress detection. Lefter et al. [18] achieved 4.2% Equal Error Rate for automatic stress detection in emergency telephone calls by fusing prosodic and spectral detectors—compared to 19% EER for individual detectors, highlighting the importance of multi-feature approaches. Demenko and Jastrzębska [7] found over-one-octave pitch shifts in highly stressful Polish police emergency calls, achieving 80-84% classification accuracy.

However, a critical reality check comes from Deschamps-Berger et al. [8], who found that while benchmark IEMOCAP data yielded 63% Unweighted Accuracy for emotion recognition, real emergency calls achieved only 45.6%—a substantial domain shift that deployment systems must account for. This finding reinforces our design decision to use bio-acoustic analysis as a triage signal rather than a sole decision-maker, routing high-distress calls to human dispatchers rather than attempting fully automated classification.

Recent work on multimodal fusion in emergency contexts supports our architecture. Feng and Devillers [9], analyzing the French CEMO emergency call center corpus, found that audio components often encode more emotive information than text in crisis contexts, with multimodal fusion yielding 4-9% absolute accuracy gains over unimodal models. This validates our approach of maintaining parallel ASR and bio-acoustic pathways that can compensate for each other's failures.

2.4 Dialect Reversion Under Cognitive Load

A theoretical foundation for Caribbean-specific ASR in emergency contexts comes from psycholinguistic research on bilingual processing under stress. The inhibitory control model establishes that non-target languages remain continuously active and must be suppressed through cognitive effort [14]. For Caribbean speakers navigating the creole continuum—from basilect (most creole features) through mesolect to acrolect (Standard English)—maintaining acrolectal speech requires sustained executive function.

The creole continuum is not simply a stylistic choice but a dynamic system of linguistic control, modulated by cognitive load. Research on cognitive load effects demonstrates that this inhibition fails under stress. Gollan and Ferreira [12] found that under high cognitive load, bilingual speakers use significantly less intraclausal code-switching, instead reverting to monolingual chunks of their dominant language. Importantly, cognitive load also affects lexical access timing—Kroll et al. [17] demonstrated that retrieval of L2 (non-dominant language) vocabulary slows significantly under dual-task conditions, providing a mechanism for stress-induced register shift.

Patrick's [20] foundational sociolinguistic analysis of the Jamaican Creole continuum establishes that stress levels influence speakers' positioning on this spectrum, with most speakers being mesolectal in normal conditions but capable of shifting toward either pole.

The implications for emergency services are significant: a professional who speaks Standard English at work may revert toward basilectal Patois when their house is flooding. Standard ASR systems, trained predominantly on acrolectal varieties, will exhibit precisely the performance degradation documented in the accent gap literature at the moment when accurate recognition is most critical. Our system addresses this by fine-tuning on Caribbean broadcast data that includes mesolectal speech, and by providing bio-acoustic fallback when ASR confidence drops—which may itself serve as a proxy indicator for basilectal reversion.

2.5 Edge Computing for Disaster Resilience

The case for offline-capable emergency AI is made starkly by infrastructure failure during recent disasters. Hurricane Maria's impact on Puerto Rico saw 95% of cell towers fail, with the entire island losing power and over 66% of the population lacking potable water [23]. Communication infrastructure failure caused delays in mortality reporting and created substantial information vacuums, contributing to a disputed death toll ultimately estimated at approximately 3,000. Recovery required over 200 days for full power restoration.

Recent advances in model compression make edge deployment increasingly feasible. Quantization studies demonstrate that 4-bit (INT4) quantization reduces Whisper model size by 45-87% with minimal WER degradation, and may actually reduce hallucinations by acting as a regularizer. Gondi and Pratap [13] demonstrated that transformer-based ASR achieves real-time inference on Raspberry Pi hardware with PyTorch mobile optimization. For the NLP component, 4-bit quantized Llama 3 8B runs at 2-5 tokens per second on Raspberry Pi 5—too slow for real-time conversation but adequate for background entity extraction tasks.

A survey of edge technologies for disaster management identifies prediction, detection, response, and recovery phases where edge computing enables real-time processing without cloud dependency [1]. The survey specifically identifies a gap in offline-capable speech and language processing at the edge—precisely the capability our system provides. Pre-positioned edge computing resources at hospitals, shelters, and emergency coordination centers, loaded with Caribbean-tuned models, could maintain triage capability even during complete grid and network failure.

2.6 Summary: Positioning Our Contribution

The literature reveals a clear opportunity space for Caribbean emergency services. As introduced in Section 1.4, existing dispatch AI systems exhibit four critical limitations for Caribbean deployment: cloud dependency with accent-agnostic ASR, text-only analysis, dialect blindness, and infrastructure fragility.

How the literature establishes each gap:

- **Cloud dependency:** ECA and Corti rely on commercial cloud APIs (Google Speech-to-Text) with documented performance degradation on non-standard English varieties (Section 2.1, Madden et al. scaling law findings). No existing system adapts for Caribbean accents or creole continua.
- **Text-only analysis:** Current approaches process only transcribed text, ignoring paralinguistic stress signals documented in Section 2.3 (Van Puyvelde et al., Schmalz et al.). This misses critical information when words are unclear or mistranscribed.
- **Dialect blindness:** No existing system accounts for stress-induced register shifting demonstrated in Section 2.4 (inhibitory control model, creole continuum research)—the phenomenon whereby speakers under acute stress revert toward basilectal varieties, precisely when accurate transcription matters most.
- **Infrastructure fragility:** Hurricane Maria case (Section 2.5) demonstrates how cloud-dependent architectures fail during the disasters that generate emergency call surges.

TRIDENT's contribution is a dispatcher-support architecture that addresses each gap while respecting the clinical authority of established triage protocols:

- **Caribbean-adapted ASR:** Fine-tuned Whisper models provide the transcription accuracy that makes downstream entity extraction viable for Caribbean speech varieties.

- **Structured entity extraction:** Local Llama 3-based NLP extracts clinical indicators needed for ESI/START application—location, mechanism of injury, breathing status, vulnerable populations—operating without internet connectivity.
- **Bio-acoustic distress detection:** A parallel signal pathway that functions even when ASR fails, transforming low transcription confidence from a system limitation into a queue prioritization feature that routes distressed callers to immediate human attention.
- **Offline operation:** Complete system deployment on edge hardware (Raspberry Pi 5) enables function during infrastructure failures when emergency services are most critical.

The result is the first dispatcher-support system designed specifically for Caribbean emergency services—not to make triage decisions, but to ensure that Caribbean-accented callers receive equitable access to the ESI and START protocols that their health ministries have adopted. TRIDENT empowers dispatchers with better information and intelligent queue prioritization; clinical judgment remains where it belongs—with trained human professionals.

3 System Architecture

TRIDENT implements a three-layer dispatcher-support architecture where each component provides independent value while contributing to intelligent queue prioritization. The system does not make clinical triage decisions—those remain with trained dispatchers applying ESI or START protocols—but ensures dispatchers receive the highest-priority calls first along with structured information to support protocol application. Figure 1 illustrates the system flow.

3.1 Design Philosophy: Enabling Protocol Application

TRIDENT's architecture reflects a core principle: **AI should empower dispatchers to apply established protocols more effectively, not replace clinical judgment**. Caribbean health ministries have adopted validated triage frameworks, ESI for emergency departments, START for mass casualty incidents, that represent decades of clinical refinement. TRIDENT's role is to solve the *input problem*: ensuring these protocols can be applied equitably to Caribbean-accented callers whose speech current ASR systems fail to transcribe accurately.

Each architectural layer addresses a specific input challenge:

- **Layer 1 (ASR):** Produces transcripts and confidence scores, enabling dispatchers to know when to trust text versus listen directly to audio.
- **Layer 2 (NLP):** Extracts structured clinical entities—location, mechanism of injury, breathing status, vulnerable populations—that map directly to ESI/START decision points.
- **Layer 3 (Bio-Acoustic):** Detects physiological distress markers that indicate caller crisis state, providing a signal not currently captured by standard protocols but valuable for queue prioritization.

The following subsections detail each layer's implementation.

3.2 Layer 1: Caribbean-Tuned ASR

The ASR layer employs OpenAI's Whisper Large model (769M parameters) fine-tuned with Low-Rank Adaptation (LoRA) on Caribbean broadcast speech. Competition experience suggests that Whisper Large is more accurate than Whisper Medium for Caribbean speech, even though the latter is the default model used by OpenAI. This notwithstanding, for TRIDENT, we selected

Whisper Medium over Large based on Madden et al.'s [19] scaling law, which demonstrates diminishing returns from model size compared to domain-specific data for Caribbean varieties. Furthermore, Whisper Medium is more efficient to run on a Raspberry Pi 5, which is the edge device we are using for TRIDENT.

Fine-tuning Configuration:

- Base model: openai/whisper-medium
- Adaptation: LoRA (rank=16, alpha=32)
- Training data: BBC Caribbean broadcast corpus (∼28,000 clips)
- Trainable parameters: ∼0.5% of total model

Confidence Scoring: The system computes **utterance-level** confidence as the mean log-probability across all decoded tokens, normalized to a 0-1 scale. Specifically:

$$\text{confidence} = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log P(t_i|t_1 \ldots t_{i-1}, \text{audio})\right) \tag{1}$$

We use utterance-level rather than token-level confidence because emergency triage requires a holistic assessment of transcription reliability. Token-level confidence would require additional aggregation logic and may miss systematic degradation patterns (e.g., consistently low confidence across an entire basilectal utterance).

Confidence Threshold: We set the "low confidence" threshold at 0.7 based on initial calibration experiments, though sensitivity analysis is needed to optimize this value (see Limitations).

3.3 Layer 2: Local NLP Entity Extraction

When ASR produces usable transcription (confidence ≥ 0.7), the NLP layer extracts structured emergency information using Llama 3 8B running locally via Ollama. The extraction schema targets entity types that map directly to ESI and START triage protocol decision points, enabling dispatchers to apply these frameworks more efficiently.

3.3.1 Entity Extraction Schema

The extraction schema targets four entity categories critical for protocol application:

- **LOCATION:** Street addresses, landmarks, geographic references—essential for dispatch routing
- **MECHANISM/HAZARD:** Emergency type (fire, flood, medical, violence, traffic)—maps to ESI resource prediction
- **CLINICAL INDICATORS:** Breathing status, consciousness, bleeding, mobility—maps to ESI Decision Points A and B
- **SCALE:** Number of people involved, injuries mentioned, vulnerable populations—maps to START mass casualty sorting and ESI resource needs

3.3.2 Mapping to ESI Decision Points

The ESI algorithm proceeds through four decision points [10]. TRIDENT's entity extraction targets information relevant to each:

ESI Decision Point	Clinical Question	TRIDENT Extraction Target
A: Immediate lifesaving intervention?	Airway, breathing, circulation compromise	"not breathing," "choking," "heavy bleeding," "unresponsive"
B: High-risk situation?	Could patient deteriorate?	Mechanism of injury, chest pain, altered mental status
C: Resource needs?	How many resources required?	Hazard type, complexity indicators, number affected
D: Vital signs?	Abnormal vitals requiring uptriage?	Any reported vitals, distress indicators

Table 1: Mapping between ESI decision points and TRIDENT entity extraction targets. Extracted entities support but do not replace dispatcher clinical judgment.

3.3.3 Mapping to START Categories

For mass casualty events, dispatchers apply START rather than ESI. TRIDENT extracts indicators relevant to START's rapid sorting:

START Category	Sorting Criteria	TRIDENT Extraction Target
GREEN (Minor)	Can walk	"walking," "minor injuries," "okay"
YELLOW (Delayed)	Breathing, follows commands	"injured but stable," "conscious"
RED (Immediate)	Not walking, breathing issues, or not following commands	"trapped," "not breathing," "unresponsive," "heavy bleeding"
BLACK (Expectant)	Not breathing after airway intervention	"dead," "not breathing," no pulse mentioned

Table 2: Mapping between START triage categories and TRIDENT entity extraction targets for mass casualty scenarios.

3.3.4 Handling Garbled Input

A critical design question is how the NLP layer behaves when ASR produces low-quality transcriptions. We address this through confidence-aware prompting:

```
SYSTEM: You are extracting emergency information from a
speech transcript. The transcription confidence is
{confidence_score}.

If confidence is below 0.7, the transcript may contain errors.
Extract what you can, but:
1. Mark uncertain extractions with [UNCERTAIN]
2. Do not hallucinate or guess missing information
3. Prioritize extracting any recognizable location names
4. Note phonetically similar alternatives for garbled terms

TRANSCRIPT: {asr_output}
```

Example of garbled transcript handling:

ASR Output (confidence=0.52)	NLP Extraction
"mi house a bun down pan [unintelligible] road near di gas station"	LOCATION: "[UNCERTAIN] road, near gas station"; HAZARD: "fire (house burning)"; CLINICAL: "unknown"; SCALE: "unknown"

Table 3: Example of NLP extraction from low-confidence ASR output. Uncertain fields are marked rather than hallucinated.

When confidence is very low (<0.4), the NLP layer produces minimal structured output and the call is flagged for immediate human review, relying on the bio-acoustic layer to inform queue prioritization.

3.3.5 Content Indicator Scoring

Beyond entity extraction, the NLP layer computes a **Content Indicator Score** ($S_c \in [0, 100]$) that informs queue prioritization. This score quantifies the urgency implied by the *semantic content* of the transcript, independent of how the caller sounds.

Important clarification: The Content Indicator Score determines *queue position*, not clinical triage level. A high score means the call should reach a dispatcher quickly; the dispatcher then applies ESI or START to determine the actual clinical priority.

This addresses a critical gap: a trained first responder or composed bystander may report a mass casualty event in a calm voice, producing low bio-acoustic distress despite extremely urgent content. Without content analysis, such calls would be deprioritized in the queue.

Classification-Based Approach. Rather than brittle keyword matching, we leverage the LLM's semantic understanding to classify transcript content along four dimensions. This approach offers critical advantages for Caribbean speech: the model can recognize that "mi granmodda drop dung an she nah move" conveys the same urgency as "my grandmother collapsed and she's not moving" without requiring an exhaustive enumeration of creole variants. The LLM also handles negation ("no one is trapped"), indirect references ("she nine months pregnant" → vulnerable), and context-dependent interpretation that keyword matching cannot capture.

The LLM outputs structured classifications according to the following schema:

```
{
  "hazard_category": "violent_crime" | "medical" | "fire" |
                     "flood" | "traffic" | "infrastructure" | "other",
  "life_threat_level": "imminent" | "potential" | "none",
  "vulnerable_population": true | false,
  "situation_status": "escalating" | "stable" | "resolved",
  "persons_affected": <integer>
}
```

A deterministic scoring function then maps these classifications to the Content Indicator Score, separating the flexibility of neural language understanding from the interpretability of rule-based scoring:

$$S_c = \min\left(100,\ S_{\text{hazard}} + S_{\text{threat}} + S_{\text{vuln}} + S_{\text{scale}}\right) \tag{2}$$

Component 1: Hazard Category (S_{hazard}). Different emergency types carry inherent urgency levels for queue prioritization:

Component 2: Life-Threat Level (S_{threat}). The LLM classifies the immediacy of danger to life based on semantic understanding of the full transcript context:

Hazard Category	Score
`violent_crime`	30
`medical`	25
`fire`	25
`flood`	20
`traffic`	15
`infrastructure`	10
`other`	5

Table 4: Hazard category weights for queue prioritization

- `imminent`: Active, immediate threat to life (trapped, not breathing, active violence, drowning) → +30
- `potential`: Situation could become life-threatening (injuries, spreading fire, chest pain) → +15
- `none`: No apparent threat to life → +0

Component 3: Vulnerable Population (S_{vuln}). Boolean classification indicating presence of children, elderly, pregnant individuals, or persons with disabilities. If `true` → +15. This reflects both ethical prioritization and reduced self-rescue capacity.

Component 4: Scale and Escalation (S_{scale}). Combines two factors:

- `persons_affected`: +5 per person, capped at +20
- `situation_status = "escalating"`: +10 (fire spreading, water rising, more vehicles involved)

Example calculations:

Transcript	LLM Classification	S_c
"Pothole on Nelson Street"	`infrastructure, none, false, stable, 0`	10
"Car accident, one person injured"	`traffic, potential, false, stable, 1`	35
"House fire, spreading to neighbor's yard"	`fire, potential, false, escalating, 0`	50
"Mi granmodda drop dung, she nah breathe"	`medical, imminent, true, stable, 1`	75
"Pickney dem trap inna di fire"	`fire, imminent, true, stable, 2+`	80

Table 5: Content indicator scoring via LLM classification. The model's semantic understanding captures urgency from both standard English and Caribbean creole variants. High scores elevate queue priority; clinical triage remains with dispatchers.

The Content Indicator Score feeds into the queue prioritization engine (Section 3.6), ensuring that semantically urgent calls reach dispatchers promptly even when vocal distress markers are absent.

Note on weight calibration: The weights presented here represent initial values based on general emergency response principles. In deployment, these weights are tunable parameters that should be calibrated in consultation with local emergency services leadership to reflect institutional priorities, regional hazard profiles, and operational experience. The architecture separates LLM classification (which requires ML expertise to modify) from the scoring function (which emergency managers can adjust without technical intervention).

3.4 Layer 3: Bio-Acoustic Distress Detection

The bio-acoustic layer operates on raw audio, independent of ASR success, extracting features correlated with psychological distress. Based on the vocal stress literature [24, 27, 28], we focus on features that capture physiological arousal through vocal production changes.

3.4.1 Feature Extraction

Using librosa, we extract the following acoustic features:

1. **Fundamental Frequency (F0):** Mean pitch extracted via autocorrelation method
   - Typical baseline: 85–180 Hz (male), 165–255 Hz (female) [25]
   - Stress indicator: Elevation above speaker baseline
2. **F0 Coefficient of Variation (CV):** Pitch instability measure
   - Computed as $CV = \sigma_{F0}/\mu_{F0}$
   - Normalizes for baseline differences across speakers
   - Stress indicator: $CV > 0.3$ suggests vocal instability
3. **Energy (RMS amplitude):** Mean intensity across utterance
   - Normalized to 0–1 scale relative to recording gain
   - Stress indicator: Elevated intensity during distress vocalizations
4. **Jitter:** Cycle-to-cycle variation in F0 period
   - Relatively independent of prosodic patterns [27]
   - Pathology threshold: >1.04% [5]

3.4.2 Distress Score Calculation

The distress score combines multiple acoustic indicators into a composite metric. We weight features according to their documented reliability and sex-independence:

$$D = w_{\text{pitch}} \cdot P + w_{\text{var}} \cdot V + w_{\text{energy}} \cdot E + w_{\text{jitter}} \cdot J \tag{3}$$

where:
The pitch elevation component now uses sex-adaptive parameters:

$$P = \min\left(1.0, \max\left(0, \frac{\bar{F}_0 - B}{R}\right)\right) \quad \text{(pitch elevation)} \tag{4}$$

where (B, R) adapts based on estimated speaker sex:

$$(B, R) = \begin{cases} (120, 80) & \text{if } \bar{F}_0^{(\text{init})} < 165 \text{ Hz (estimated male)} \\ (200, 100) & \text{otherwise (estimated female)} \end{cases} \tag{5}$$

The baseline B and range R parameters adapt based on a heuristic sex estimation from the initial 3 seconds of speech. A male speaker at 170 Hz (stressed) now contributes $P = (170 - 120)/80 = 0.625$ rather than the previous formulation's 0.0, addressing the male pitch penalty.

The remaining components are:

$$V = \min\left(1.0, \frac{CV_{F0}}{0.5}\right) \quad \text{(pitch instability)} \tag{6}$$

$$E = \min\left(1.0, \frac{\bar{E}}{0.1}\right) \quad \text{(energy)} \tag{7}$$

$$J = \min\left(1.0, \frac{\text{jitter}}{0.02}\right) \quad \text{(perturbation)} \tag{8}$$

The weights reflect relative reliability from the literature:

- $w_{\text{pitch}} = 0.30$ — F0 elevation is the most consistent stress marker but is sex-dependent
- $w_{\text{var}} = 0.35$ — F0 coefficient of variation is sex-normalized and robust
- $w_{\text{energy}} = 0.20$ — intensity elevation accompanies distress
- $w_{\text{jitter}} = 0.15$ — perturbation measures are prosody-independent

3.4.3 Threshold Classification

- **High Distress:** $D > 0.5$
- **Low Distress:** $D \leq 0.5$

These thresholds are calibrated against Van Puyvelde et al.'s [27] findings on vocal markers in emergency versus baseline speech.

Note on sex differences: The distress score prioritizes sex-normalized features (CV, jitter) over absolute F0 elevation to mitigate the substantial baseline differences between male (85–175 Hz) and female (165–270 Hz) speakers. See Section 6.2 for detailed discussion of remaining bias risks.

3.5 The Complementarity Principle

The theoretical foundation for our multi-layer design rests on what we term the **Complementarity Principle**: the three signal dimensions capture distinct failure modes and urgency indicators that compensate for each other's blind spots, ensuring dispatchers receive the most critical calls first regardless of which individual signal might fail.

Dimension 1: Transcription Confidence. The conditions that degrade ASR performance (high stress, code-switching to basilect, environmental noise) are precisely the conditions that often accompany genuine emergencies. Low confidence is not merely a technical limitation to be hidden—it correlates with caller distress and should elevate queue priority while flagging the call for direct audio review.

Dimension 2: Content Indicators. Semantic analysis of transcript content captures urgency that vocal characteristics may miss. Trained professionals, repeat callers, and composed bystanders often report critical emergencies without elevated vocal stress—their calm delivery masks the urgency that only content analysis reveals. When transcription confidence is high, extracted entities map directly to ESI/START decision points.

Dimension 3: Bio-Acoustic Distress. Vocal stress markers (elevated pitch, intensity, instability) provide a parallel assessment channel that operates on raw audio, independent of transcription success. A caller whose speech is entirely unintelligible to ASR will still produce detectable distress signals. This dimension captures information not currently used by ESI or START protocols, representing TRIDENT's novel contribution to dispatcher awareness.

This creates a robust prioritization space with complementary coverage:

Dimensional ordering. The three dimensions are evaluated in deliberate sequence: *Confidence*, *Content*, *Concern*. This ordering reflects operational logic: (1) *Can we understand the caller?*—ASR confidence determines whether transcription is reliable enough for downstream analysis; (2) *What is being reported?*—semantic content establishes the substance of the emergency; (3) *How distressed does the caller sound?*—bio-acoustic indicators validate and can elevate priority, but do not override content. This sequence ensures that a composed professional reporting a mass casualty event receives appropriate priority based on content, while a highly distressed caller reporting a minor issue is not over-prioritized based on vocal expression alone.

- **High Confidence + Low Content + Low Concern:** Routine call; dispatcher applies ESI using extracted entities at normal pace
- **High Confidence + High Content + Low Concern:** The composed reporter—urgent content from a calm caller requires elevated queue position; dispatcher reviews entities and applies ESI, likely assigning ESI-2 or ESI-3
- **High Confidence + Low Content + High Concern:** Anxious caller, possibly minor issue—dispatcher assesses whether distress reflects emergency or anxiety
- **High Confidence + High Content + High Concern:** All signals aligned; immediate queue position for rapid ESI/START application
- **Low Confidence + Low Content + Low Concern:** Likely technical issue; dispatcher reviews audio quality before processing
- **Low Confidence + High Content + Low Concern:** Garbled but fragments suggest urgency—elevated priority; dispatcher listens directly
- **Low Confidence + Low Content + High Concern:** Distressed caller with unintelligible speech—immediate priority; dispatcher listens and applies protocol based on direct assessment
- **Low Confidence + High Content + High Concern:** Maximum queue priority—all indicators suggest crisis; immediate dispatcher attention

Two cells represent our key insights. The **High Confidence + High Content + Low Concern** cell captures callers whose semantic content demands urgent attention despite calm delivery: the trained first responder, medical professional, or composed bystander whose measured voice belies the severity of their report. The **Low Confidence + Low Content + High Concern** cases capture the complementary pattern—callers in crisis whose speech has shifted toward basilectal registers, where ASR failure combined with vocal stress becomes valuable prioritization information rather than system failure.

Together, these insights ensure that neither semantic nor paralinguistic signals alone determine queue position—and that clinical triage decisions remain with trained dispatchers who can assess the full context of each call.

3.6 Queue Prioritization Engine

The Queue Prioritization Engine integrates three independent signals to determine the order in which calls receive dispatcher attention. **Critically, this system determines queue position, not clinical triage category.** Clinical triage—assigning ESI levels 1–5 or START colors (RED/YELLOW/GREEN/BLACK)—remains the responsibility of trained dispatchers applying Ministry of Health protocols.

The prioritization logic ensures that:

1. Callers most likely to need immediate intervention reach dispatchers first
2. Dispatchers receive structured information to support rapid protocol application
3. Calls with unreliable transcriptions are flagged for direct audio review

3.6.1 Three-Dimensional Prioritization Space

Each call is mapped to a point in prioritization space defined by:

- **Transcription Confidence** (C): High (≥ 0.7) or Low (< 0.7)
- **Content Indicators** (S_c): High (≥ 50) or Low (< 50)
- **Bio-Acoustic Distress** (D): High (> 0.5) or Low (≤ 0.5)

The $2 \times 2 \times 2$ combination yields eight queue priority cells, shown in Table 6.

Confidence	Content	Concern	Queue	Dispatcher Action
High	Low	Low	**Q5-ROUTINE**	Apply ESI using extracted entities
High	High	Low	**Q2-ELEVATED**	Priority review; calm reporter, urgent content*
High	Low	High	**Q3-MONITOR**	Review for anxiety vs. emergency
High	High	High	**Q1-IMMEDIATE**	Immediate attention; apply ESI/START
Low	Low	Low	**Q5-REVIEW**	Check audio quality; possible technical issue
Low	High	Low	**Q2-ELEVATED**	Listen to audio; fragments suggest urgency
Low	Low	High	**Q1-IMMEDIATE**	Priority audio review; possible dialect shift†
Low	High	High	**Q1-IMMEDIATE**	Highest priority; all indicators elevated

Table 6: Three-dimensional queue prioritization matrix. *Addresses trained responder/composed bystander scenario. †Preserves core insight: low ASR confidence + high vocal concern may indicate stress-induced basilectal shift requiring human ears.

3.6.2 Queue Priority Levels

Q1-IMMEDIATE: Top of queue. Dispatcher reviews within seconds. System flags call for potential crisis requiring direct audio assessment.

Q2-ELEVATED: High priority queue. Dispatcher attention within 1–2 minutes. Extracted entities displayed prominently to support rapid ESI/START application.

Q3-MONITOR: Moderate priority. May indicate anxious caller with non-urgent situation. Dispatcher assesses and de-escalates if appropriate.

Q5-ROUTINE: Standard queue. Extracted entities available; dispatcher applies ESI at normal pace.

Q5-REVIEW: Standard queue but flagged for audio quality check. May indicate technical issues rather than emergency content.

Note on Q4: The current matrix does not produce a Q4 outcome. Future refinement with real operational data may identify scenarios warranting an intermediate priority level. A theoretical case: High Confidence + Low Content + Moderate Concern (anxious caller, minor issue).

3.6.3 Relationship to Clinical Triage Protocols

Table 7 illustrates how TRIDENT's queue prioritization relates to—but does not replace—clinical triage protocols.

TRIDENT Output	Dispatcher Action	Protocol Application
Q1-IMMEDIATE	Immediate audio review; assess caller state	Dispatcher determines ESI-1/2 or START-RED based on clinical assessment
Q2-ELEVATED	Review extracted entities; listen if uncertain	Dispatcher applies ESI using structured data; may be ESI-2 through ESI-4
Q3-MONITOR	Assess distress source; de-escalate if needed	Often ESI-4/5 after dispatcher determines no emergency
Q5-ROUTINE/REVIEW	Process normally using extracted metadata	Full ESI protocol application; typically ESI-3 through ESI-5

Table 7: TRIDENT queue priority does not determine clinical triage level. Dispatchers apply ESI or START protocols after reviewing TRIDENT's structured outputs and/or call audio.

3.6.4 Dispatcher Interface

Figure 2 and Figure 3 illustrate the dispatcher interface for contrasting scenarios. The interface presents:

- Queue priority level with visual urgency coding
- Transcription confidence (with recommendation to review audio if low)
- Extracted clinical entities mapped to ESI/START decision points
- Bio-acoustic distress indicators
- One-click access to call audio for direct assessment

4 Theoretical Foundations

4.1 Why Accent-Tuned ASR Is Necessary But Insufficient

Fine-tuning Whisper on Caribbean speech will improve transcription accuracy, but it cannot eliminate the accent gap entirely. Madden et al. [19] achieved 30% WER on Jamaican Patois with fine-tuning—a dramatic improvement from 89% baseline, but still far above the <5% WER typical for standard English. In emergency contexts, even 30% WER means nearly one-third of words may be incorrect, potentially including critical location or hazard information.

Moreover, fine-tuning on broadcast speech cannot fully capture emergency speech characteristics: elevated noise (sirens, screaming, wind), emotional vocal qualities, and the stress-induced basilectal reversion discussed above. A system relying solely on ASR, no matter how well-tuned, will fail precisely when it is needed most.

4.2 Why Bio-Acoustic Analysis Is Necessary But Insufficient

Conversely, bio-acoustic distress detection alone cannot provide the semantic information needed for emergency dispatch. A caller may exhibit extreme vocal stress while saying "my house is on fire" or "I lost my keys"—the distress signal is identical, but the appropriate response differs dramatically.

Furthermore, as Deschamps-Berger et al. [8] demonstrated, laboratory accuracy of emotion recognition systems (63%) drops substantially in real emergency calls (45.6%). Bio-acoustic features provide reliable *gradient* information about caller state but cannot substitute for semantic content.

4.3 The Integration Thesis

Our architecture integrates these complementary information sources based on the following thesis: **In emergency contexts, the correlation between ASR failure and genuine distress creates an opportunity to use recognition uncertainty as a routing signal rather than an error to be minimized.**

This thesis rests on the psycholinguistic literature establishing that:

1. Stress triggers cognitive load effects that impair executive function [12]
2. Impaired executive function leads to reduced inhibition of dominant language varieties [14]
3. For Caribbean speakers, dominant varieties include basilectal forms underrepresented in ASR training [20, 19]
4. Stress simultaneously elevates bio-acoustic markers (F0, intensity) that can be detected independently of speech content [27]

The logical conclusion: when ASR confidence drops and bio-acoustic distress rises, the system has detected a caller in genuine crisis whose speech has shifted beyond standard recognition capabilities. This combination should trigger immediate human review—not because the system has failed, but because it has successfully identified a caller who needs human attention most.

5 Deployment Considerations

5.1 Operational Context: Supporting Protocol Application

TRIDENT is designed to integrate with existing emergency dispatch workflows, not replace them. Caribbean health ministries have adopted standardized triage protocols—ESI for routine emergency department operations, START for mass casualty incidents—and TRIDENT's deployment model respects this clinical framework.

Day-to-day operations (ESI context): TRIDENT processes incoming calls to extract structured entities (location, mechanism, clinical indicators) and assigns queue priority. Dispatchers receive calls in priority order along with extracted data that maps to ESI decision points. The dispatcher applies ESI to determine clinical acuity level (1–5) and appropriate response.

Mass casualty events (START context): During hurricanes, earthquakes, or other disasters generating call surges, TRIDENT's queue prioritization becomes critical for managing volume that exceeds dispatcher capacity. Extracted entities map to START categories (RED/YELLOW/GREEN/BLACK), enabling dispatchers to rapidly sort callers even when transcription quality degrades due to infrastructure stress.

Key principle: TRIDENT determines *which calls dispatchers see first* and *what structured information they receive*. Clinical triage decisions—ESI level assignment, START color coding, resource dispatch—remain with trained human professionals applying Ministry of Health protocols.

5.2 Operational Deployment Models

A critical question for any emergency AI system is: *when in the call workflow does it actually operate?* TRIDENT's processing latency (45–60 seconds on edge hardware) precludes real-time transcription during live dispatcher-caller conversation. This subsection explicitly addresses the operational contexts where TRIDENT provides value, ordered from highest to lowest impact.

5.2.1 Primary Value: Surge Queue Prioritization

TRIDENT's greatest value emerges during **disaster surge conditions**—hurricanes, earthquakes, floods—when call volume exceeds dispatcher capacity and callers must wait in queue.

Operational flow:

1. Caller dials emergency services; all dispatchers are engaged
2. Caller enters queue and hears automated message: "Please hold. Briefly describe your emergency and location."
3. Caller provides initial statement (15–30 seconds)
4. TRIDENT processes audio while caller waits (45–60 seconds)
5. Queue is reordered by priority level (Q1-IMMEDIATE through Q5-ROUTINE)
6. When dispatcher becomes available, highest-priority call routes first
7. Dispatcher receives: transcription, confidence flag, extracted entities, distress indicators
8. Dispatcher applies ESI or START protocol with TRIDENT's structured data and/or direct audio review

Why this context maximizes value:

- Calls are waiting regardless—TRIDENT uses wait time productively
- Queue prioritization ensures most critical callers reach dispatchers first
- Extracted entities enable faster protocol application when dispatcher connects
- Low ASR confidence flags alert dispatchers to potential dialect shift or audio quality issues before they engage

Policy and operational requirements:

- Automated initial message prompting callers to describe their emergency
- Integration with existing queue management infrastructure
- Dispatcher training on interpreting TRIDENT outputs and priority levels
- Clear protocols for when human override of queue priority is appropriate

This deployment model represents TRIDENT's primary design target. Caribbean emergency services face predictable annual surge events (hurricane season, June–November) where this capability would directly impact response effectiveness.

5.2.2 Secondary Value: Parallel Processing and Documentation Support

During normal operations when dispatchers answer immediately, TRIDENT can operate as a **parallel processing layer**—a "second listener" that supplements dispatcher perception.

Operational flow:

1. Dispatcher answers immediately and begins conversation
2. Audio streams to TRIDENT in background
3. Approximately 60 seconds into call, TRIDENT results appear on dispatcher screen
4. Dispatcher uses extracted entities to confirm notes and supplement information
5. Structured data auto-populates CAD system fields, reducing manual entry

Why this context provides value:

- Structured entity extraction improves documentation quality and consistency
- Automatic capture reduces cognitive load during high-stress calls
- Creates audit trail of information captured from caller
- Low-confidence flags alert dispatcher to segments requiring re-confirmation

Limitations in this context:

- Results arrive mid-call, not at start—less useful for initial triage decisions
- Dispatcher has already heard the caller—TRIDENT confirms rather than informs
- Value depends on CAD system integration and dispatcher willingness to reference parallel output

This deployment model provides incremental value during normal operations but does not transform the dispatch workflow. Its primary benefit is documentation quality and consistency rather than triage support.

5.2.3 Tertiary Value: Voicemail and Callback Triage

During extreme surge conditions when call volume overwhelms even queue capacity, calls may overflow to voicemail systems.

Operational flow:

1. All lines occupied, queue full; call routes to voicemail
2. Automated message: "All dispatchers are currently assisting other callers. Please leave your name, location, and describe your emergency."
3. TRIDENT processes voicemail messages in batch
4. Dispatchers receive prioritized callback list with extracted entities
5. Callbacks proceed in priority order

Why this context provides value:

- Batch processing suits TRIDENT's latency profile
- Prioritization ensures callbacks reach critical situations first

- Extracted entities enable dispatchers to prepare before callback

Limitations:

- Callback introduces inherent delay—inappropriate for immediately life-threatening emergencies
- Voicemail messages may be less coherent than live caller statements
- Represents failure mode (overwhelmed system) rather than normal operation

This model is valid for extreme disaster scenarios but should not be considered a primary deployment target.

5.2.4 Future Potential: Automated Initial Capture

The highest-impact deployment model would involve **automated initial capture** before dispatcher connection:

1. Caller dials emergency services
2. Automated system: "110 Emergency. Please state your location and describe your emergency."
3. System records 20–30 seconds of caller statement
4. TRIDENT processes while caller hears: "Please hold, connecting you to a dispatcher"
5. Call routes to dispatcher who already has: transcription, entities, distress score, confidence flags
6. Dispatcher engages caller with full context from first word

This model would maximize TRIDENT's value by ensuring dispatchers receive structured information *before* engaging callers. However, it requires significant operational changes:

- Modification of initial call-answering protocols
- Caller acceptance of brief automated interaction before human contact
- Robust fallback for callers who cannot provide verbal statements
- Policy approval from health ministry and emergency services leadership

We present this model as future potential rather than current recommendation, acknowledging that workflow changes of this magnitude require extensive stakeholder consultation and pilot testing.

5.2.5 Early Exit Architecture

To address the 45–60 second latency inherent in full three-layer processing, TRIDENT implements an early exit mechanism that bypasses NLP when ASR and bio-acoustic signals alone warrant immediate prioritization.

Early Exit Conditions:

1. **High Distress + Low Confidence:** If $D > 0.8$ and $C < 0.4$, route immediately to Q1-IMMEDIATE. This captures callers exhibiting extreme vocal stress whose speech has likely shifted to basilectal registers.

2. **Extreme Distress:** If $D > 0.9$ regardless of confidence, route to Q1-IMMEDIATE. Severe physiological distress warrants immediate human attention.

Under early exit, ASR and bio-acoustics complete in approximately 12 seconds (10s transcription + 2s preprocessing, with bio-acoustic extraction parallel). NLP may continue in the background, updating the dispatcher's screen with extracted entities after the call has already been prioritized.

This reduces Time-to-Q1 from approximately 55 seconds to 12 seconds for clearly distressed callers—a critical improvement for surge queue scenarios.

5.2.6 Summary: Matching Deployment to Context

Deployment Model	Operational Context	Primary Value	Requirements
Surge Queue Prioritization	Disaster conditions, call backlog	Queue ordering, pre-dispatch context	Automated prompt, queue integration
Parallel Processing	Normal operations	Documentation support, consistency	CAD integration, dispatcher training
Voicemail Triage	Extreme overflow	Callback prioritization	Voicemail system, callback protocols
Automated Initial Capture	Any (future)	Full pre-dispatch context	Significant workflow changes

Table 8: TRIDENT deployment models matched to operational context. Surge queue prioritization represents the primary design target; other models provide incremental value with varying implementation requirements.

TRIDENT's architecture supports all four models, but we are explicit that **surge queue prioritization during disaster conditions** represents the scenario where the system provides maximum value with feasible operational integration. Caribbean emergency services face predictable annual surge events where this capability would directly support more equitable application of ESI and START protocols to Caribbean-accented callers.

5.3 Hardware Requirements

The complete system is designed for deployment on Raspberry Pi 5 (8GB RAM) or equivalent edge hardware:

Component	Model	Size	Inference Speed
ASR	Whisper Medium (INT4)	∼400MB	∼10s per 30s audio
NLP	Llama 3 8B (4-bit)	∼4GB	2-5 tokens/sec
Bio-acoustic	librosa + numpy	<50MB	Real-time

Table 9: Hardware requirements for edge deployment

Total system footprint: ∼4.5GB, well within Raspberry Pi 5 8GB capacity.

5.4 Latency Analysis

End-to-end processing time for a 30-second call segment:

- Audio preprocessing: ∼2 seconds

- ASR transcription: ∼10 seconds
- Bio-acoustic extraction: ∼1 second (parallel with ASR)
- NLP entity extraction: ∼30-45 seconds
- Queue priority assignment: <1 second
- **Total: ∼45-60 seconds**

This latency profile is unsuitable for real-time call answering but well-matched to surge queue prioritization, where calls are waiting regardless and TRIDENT uses queue time productively (see Section 5.2 for detailed deployment model analysis).

For real-time operation or parallel processing during live calls, GPU acceleration (e.g., NVIDIA Jetson) would reduce total latency to under 10 seconds, enabling results to appear early in dispatcher-caller conversations.

5.5 Offline Operation

All components operate without internet connectivity:

- Whisper model weights stored locally
- Llama 3 served via local Ollama instance
- Bio-acoustic analysis uses standard signal processing libraries
- Queue prioritization logic implemented in local Python

This enables deployment at emergency coordination centers that may lose internet connectivity during disasters while maintaining local power (generator/battery backup). Critically, offline capability ensures that TRIDENT can support dispatcher application of ESI and START protocols precisely when infrastructure degradation makes accurate call processing most difficult—during the disasters that generate emergency call surges.

5.6 Integration with Existing Dispatch Systems

TRIDENT is designed as a **pre-processing layer** that integrates with, rather than replaces, existing Computer-Aided Dispatch (CAD) systems. The integration model:

1. **Input:** Audio stream or recording from existing telephony infrastructure
2. **Processing:** TRIDENT extracts entities, computes distress indicators, assigns queue priority
3. **Output:** Structured data package passed to CAD system, including:
    - Queue priority level (Q1-IMMEDIATE through Q5-ROUTINE)
    - Transcription with confidence score
    - Extracted entities mapped to ESI/START decision points
    - Bio-acoustic distress indicators
    - Flag for audio review if transcription confidence is low
4. **Dispatcher interface:** CAD system presents calls in priority order; dispatcher applies ESI or START protocol using TRIDENT's structured data and/or direct audio review

This architecture requires no changes to clinical protocols or dispatcher training on triage methodology—only familiarization with TRIDENT's output format and the meaning of queue priority levels.

6 Limitations and Future Work

6.1 Current Limitations

Validation gap (most critical). This paper presents an architectural framework with theoretical grounding but limited empirical validation on real emergency calls. Performance claims for each layer are based on component evaluations and related literature rather than end-to-end system testing. The three-dimensional queue prioritization matrix (ASR confidence × distress × content indicators) is theoretically motivated but has not been validated against expert dispatcher judgments.

Protocol integration. While this paper frames TRIDENT as a dispatcher-support system for ESI and START protocol application, the entity extraction schema and queue prioritization logic were developed independently of clinical stakeholder input. Full integration with Ministry of Health workflows would require:

- Validation that extracted entities map correctly to ESI decision points A–D
- Confirmation that queue priority levels align with operational dispatcher workflows
- Assessment of whether bio-acoustic distress indicators provide actionable information beyond what dispatchers already perceive
- Training material development for dispatcher familiarization with TRIDENT outputs

This clinical integration work represents essential future collaboration with Caribbean emergency services professionals. The current paper establishes technical feasibility; operational validation requires partnership with the health ministries whose protocols TRIDENT aims to support.

Training data constraints. Caribbean emergency speech corpora do not exist. ASR fine-tuning was performed on broadcast speech, which differs significantly from emergency call acoustics in noise profiles, emotional content, and register distribution. The gap between training domain (broadcast) and deployment domain (emergency calls) may introduce systematic errors not captured in current evaluation.

Bio-acoustic threshold calibration. Distress detection thresholds are derived from literature on non-Caribbean, predominantly Western populations. Baseline vocal characteristics may vary across Caribbean demographics, requiring population-specific calibration.

6.2 Sex Differences in F0 Baseline

Fundamental frequency is sexually dimorphic: male voices typically range 85–175 Hz while female voices range 165–270 Hz [25, 26]. Setting a single absolute F0 threshold for distress detection risks differential sensitivity across speaker sex.

Architectural mitigation strategies:

- Prioritizing sex-normalized features: F0 coefficient of variation ($CV = \sigma_{F0}/\mu_{F0}$) captures pitch instability independent of baseline; jitter measures cycle-to-cycle perturbations that are "relatively independent from prosodic patterns" [27]
- Weighting normalized features (CV: 0.35, jitter: 0.15) more heavily than absolute F0 elevation (0.30) in the distress score calculation

Research confirms that stress manifests with "striking parallels in men and women" [21]—both sexes show increased pitch mean, minimum, and variation under acute stress. The challenge is not that stress manifests differently, but that baseline values differ.

Residual bias risks:

- **False positive risk:** A relaxed female speaker near the upper baseline range may contribute to elevated distress scores
- **False negative risk:** A stressed male speaker with naturally low F0 may not contribute sufficiently to the pitch component

Automatic sex identification from voice is itself an imperfect classifier, particularly for voices near the overlap region of male and female F0 distributions. Rather than introduce a potentially error-prone sex classification step, we employ the sex-normalized feature strategy above. A validation study with sex-stratified analysis on Caribbean emergency calls is essential to: (1) calibrate population-appropriate thresholds, (2) confirm that normalized measures maintain sensitivity across speaker demographics, and (3) determine whether Caribbean populations exhibit different baseline distributions requiring adjustment.

Content indicator classification. The Content Indicator Score depends on LLM classification quality. While leveraging Llama 3's semantic understanding avoids brittle keyword matching, it introduces new failure modes:

- Classification errors propagate deterministically to queue priority
- Caribbean creole expressions not well-represented in LLM training data may be misclassified
- The model may fail to recognize culturally-specific threat indicators or landmarks

Empirical evaluation of classification accuracy on Caribbean emergency transcripts is needed, with particular attention to false negatives (urgent content classified as non-urgent) that could delay dispatcher attention to critical calls.

Single-speaker assumption. The current architecture assumes single-speaker input. Multi-party calls, common in emergencies ("put your mother on the phone"), are not handled. Speaker changes mid-call could confuse bio-acoustic analysis and entity extraction continuity.

Threshold sensitivity. Multiple thresholds govern system behavior: ASR confidence (0.7), distress score (0.5), and content indicators (50). These values were selected based on literature and initial calibration but have not been rigorously optimized. Sensitivity analysis examining system performance across threshold combinations is needed to understand precision-recall tradeoffs for each queue priority level.

6.3 Future Work

Clinical stakeholder collaboration. The most important next step is partnership with Caribbean emergency services to validate TRIDENT's utility in real dispatch workflows. This includes:

- Observation studies of current ESI/START application challenges
- Dispatcher feedback on extracted entity usefulness and queue priority alignment
- Iterative refinement of the entity extraction schema based on clinical input
- Development of dispatcher training materials for TRIDENT integration

Caribbean Emergency Speech Corpus. A critical enabler for future progress is a dedicated corpus combining Caribbean-accented speech with emergency domain content and stress annotations. We are exploring partnerships with Caribbean emergency services to develop such a resource, with appropriate privacy protections and community consent.

6.3.1 Data Collection Methodology: VoicefallJA

We are developing a gamified speech elicitation platform, *VoicefallJA*, designed to collect stressed Caribbean speech ethically and at scale.

Game Design. VoicefallJA presents falling-word targets that players must speak aloud before words exit the screen. Difficulty progression (increased speed, shorter response windows) induces naturalistic cognitive load, eliciting stress responses without deception. Phrase prompts span the creole continuum—from acrolectal ("The hospital is on Nelson Street") through mesolectal ("Di hospital deh pon Nelson Street") to basilectal ("Di haspital deh dung a Nelson Street side")—enabling register-annotated collection.

Distribution. The Progressive Web App is designed for WhatsApp-based distribution through institutional partnerships, specifically Methodist Church in the Caribbean and the Americas (MCCA) congregational networks spanning Jamaica and Montserrat. Target: 100–300 speakers, 5,000+ utterances.

Ethical Framework. Our consent model includes: (1) persistent recording indicators, (2) "panic button" mid-session revocation, (3) play-only mode without audio recording, and (4) community benefit mechanisms ensuring Caribbean communities receive value from their linguistic contributions.

Annotation Schema. Each utterance is annotated with: register label (ACR/MES/BAS), game-induced stress level (1–5), prosodic features (F0 mean/SD, jitter, shimmer, speech rate), and optional demographics.

Timeline. Q1 2026: beta launch; Q2–Q3 2026: data collection; Q3 2026: dataset release targeting 20+ hours of register-annotated speech.

Empirical validation. End-to-end evaluation with emergency dispatch professionals assessing whether TRIDENT's queue prioritization aligns with expert judgment. This should include:

- Comparison of three-dimensional prioritization against two-dimensional (confidence × distress) baseline
- Sex-stratified analysis of bio-acoustic distress detection accuracy
- Assessment of entity extraction accuracy on Caribbean creole transcripts
- Measurement of dispatcher efficiency gains (if any) when using TRIDENT outputs

Ablation studies. Rigorous testing to quantify the contribution of each architectural component:

- Does bio-acoustic analysis improve queue prioritization over ASR-only approaches?
- Do content indicators catch urgent calls missed by distress detection alone?
- What is the marginal value of Caribbean-tuned ASR versus off-the-shelf Whisper?

Sex-adaptive distress detection. Implementing and validating approaches to further reduce sex bias:

- Within-call F0 *change* detection rather than absolute thresholds
- Automatic speaker characteristic estimation for threshold adaptation
- Ensemble approaches combining multiple normalization strategies

Dialect density estimation. Augmenting the system with automatic estimation of creole feature density, providing dispatchers with guidance on expected communication challenges and informing decisions about when to rely on extracted text versus direct audio review.

Multilingual extension. Caribbean emergency services handle calls in English, Spanish, French, Dutch, and various creoles. Extending the architecture to multilingual operation would significantly expand impact, though each language introduces its own ASR adaptation and entity extraction challenges.

Edge deployment optimization. While the architecture is designed for offline operation, current latency profiles (45–60 seconds per call) limit real-time applicability. Optimization for edge hardware (Raspberry Pi, embedded GPU) would enable faster queue prioritization at emergency coordination centers operating with degraded connectivity.

7 Conclusion

TRIDENT presents a dispatcher-support architecture that ensures Caribbean-accented emergency callers receive equitable access to ESI and START triage protocols. By combining accent-adapted speech recognition, local NLP entity extraction, and bio-acoustic distress detection, the system empowers dispatchers to apply established protocols even when automated transcription fails.

The architecture operationalizes two complementary insights established in Section 1.3: that ASR uncertainty combined with vocal distress signals priority callers requiring human attention, and that calm delivery of urgent content must not delay dispatcher response. These insights drive the three-dimensional queue prioritization matrix that routes calls based on confidence, content, and concern signals.

Critically, TRIDENT respects the clinical authority of established protocols. The system determines which calls dispatchers see first and provides structured information to support rapid protocol application—but triage decisions remain with trained human professionals. This design philosophy reflects a broader principle for emergency AI: technology should empower human expertise, not attempt to replace it.

We hope this architectural framework contributes to more equitable emergency services—not just for Caribbean populations, but for the billions of speakers worldwide whose accents and dialects remain underserved by current speech technology. When a caller dials for help, the system that answers should understand them. TRIDENT is a step toward that goal.

References

[1] Mohamed Aboualola, Khalid Abualsaud, Tamer M. S. Khattab, Nizar Zorba, and Hossam S. Hassanein. Edge technologies for disaster management: A survey of social media and artificial intelligence integration. *IEEE Access*, 11:73782–73802, 2023.

[2] Afraa Attiah and Manal Kalkatawi. AI-powered smart emergency services support for 9-1-1 call handlers using textual features and svm model for digital health optimization. *Frontiers in Big Data*, 8:1594062, 2025.

[3] Stig Nikolaj Blomberg et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation*, 138:322–329, 2019.

[4] Stig Nikolaj Blomberg et al. Effect of machine learning on dispatcher recognition of out-of-hospital cardiac arrest during calls to emergency medical services: A randomized clinical trial. *JAMA Network Open*, 4(1):e2032320, 2021.

[5] Paul Boersma and David Weenink. *Praat: doing phonetics by computer*, 2013. Version 5.3.51.

[6] Marcel Lucas Chee, Mark Leonard Chee, Haotian Huang, Katelyn Mazzochi, Kieran Taylor, Han Wang, Mengling Feng, Andrew Fu Wah Ho, Fahad Javaid Siddiqui, Marcus Eng Hock Ong, Nan Liu, et al. Artificial intelligence and machine learning in prehospital emergency care: A scoping review. *iScience*, 26(8):107407, 2023.

[7] Grażyna Demenko and Magdalena Jastrzębska. Analysis of voice stress in call center conversations. In *Proceedings of Speech Prosody 2012*, pages 183–186, 2012.

[8] Théo Deschamps-Berger, Lori Lamel, and Laurence Devillers. End-to-end speech emotion recognition: Challenges of real-life emergency call centers data recordings. In *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–8, 2021.

[9] Théo Deschamps-Berger, Lori Lamel, and Laurence Devillers. Exploring attention mechanisms for multimodal emotion recognition in an emergency call center corpus. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023*, pages 1–5, 2023.

[10] Emergency Nurses Association. *Emergency Severity Index (ESI): A Triage Tool for Emergency Departments, Version 5*. Emergency Nurses Association, Schaumburg, IL, 2020. Available at https://www.ena.org/practice-resources/resource-library/esi.

[11] Simone French, Georgiana Gordon-Strachan, Kevon Kerr, Jacquiline Bisasor-McKenzie, Lambert Innis, and Paula Tanabe. Assessment of interrater reliability of the emergency severity index after implementation in emergency departments in jamaica using a learning collaborative approach. *Journal of Emergency Nursing*, 46(6):875–882, 2020.

[12] Tamar H. Gollan and Victor S. Ferreira. Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3):640–665, 2009.

[13] Santosh Gondi and Vineel Pratap. Performance evaluation of offline speech recognition on edge devices. *Electronics*, 10(21):2697, 2021.

[14] David W. Green. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2):67–81, 1998.

[15] John H. L. Hansen and Sanjay Patil. Speech under stress: Analysis, modeling and recognition. *Speaker Classification I*, pages 108–137, 2007. Chapter in Springer Lecture Notes in Computer Science.

[16] Allison Koenecke et al. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14):7684–7689, 2020.

[17] Judith F. Kroll, Susan C. Bobb, and Zofia Wodniecka. Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9(2):119–135, 2006.

[18] Iulia Lefter, Leon J. M. Rothkrantz, David A. van Leeuwen, and Pascal Wiggers. Automatic stress detection in emergency (telephone) calls. *International Journal of Intelligent Defence Support Systems*, 4(2):148–168, 2011.

[19] Jordan Madden, Matthew Stone, Dimitri Johnson, and Daniel Geddez. Towards robust speech recognition for Jamaican Patois music transcription. *arXiv preprint arXiv:2507.16834*, 2025.

[20] Peter L. Patrick. *Urban Jamaican Creole: Variation in the Mesolect*. John Benjamins Publishing, Amsterdam, 1999.

[21] Katarzyna Pisanski, Joanna Nowak, and Piotr Sorokowski. Multimodal stress detection: Testing for covariation in vocal, hormonal and physiological responses to Trier Social Stress Test. *Hormones and Behavior*, 106:52–61, 2018.

[22] Ramon Sanabria, Nikolay Bogoychev, Nina Markl, Andrea Carmantini, Ondrej Klejch, and Peter Bell. The edinburgh international accents of english corpus: Towards the democratization of english asr. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023*, pages 1–5, 2023.

[23] Carlos Santos-Burgoa, John Sandberg, Erick Suárez, Ann Goldman-Hawes, Scott Zeger, Alejandra Garcia-Meza, Cynthia M. Pérez, Kenneth Rivera, Adriana Colón Ramos, Jose Figueroa, et al. Differential and persistent risk of excess mortality from hurricane maria in puerto rico: A time-series analysis. *The Lancet Planetary Health*, 2(11):e478–e488, 2018.

[24] Lilien Schewski, Mathew Magimai Doss, Guido Beldi, and Sandra Keller. Measuring negative emotions and stress through acoustic correlates in speech: A systematic review. *PLOS ONE*, 20(7):e0328833, 2025.

[25] Ingo R. Titze. Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, 85(4):1699–1707, 1989.

[26] Hartmut Traunmüller and Anders Eriksson. The frequency range of the voice fundamental in the speech of male and female adults. *Journal of the Acoustical Society of America*, 97(4):2634–2639, 1995.

[27] Martine Van Puyvelde, Xavier Neyt, Francis McGlone, and Nathalie Pattyn. Voice stress analysis: A new framework for voice and effort in human performance. *Frontiers in Psychology*, 9:1994, 2018.

[28] André Veiga et al. The fundamental frequency of voice as a potential stress biomarker: A systematic review and meta-analysis. *Stress and Health*, 2025.

A Implementation Details

Repository: `https://github.com/smg-labs/project-filter` *(to be made public upon acceptance)*

Dependencies:

- Python 3.11+
- openai-whisper
- transformers, peft (LoRA fine-tuning)
- ollama (Llama 3 serving)
- librosa (audio feature extraction)
- jiwer (WER evaluation)

Hardware requirements:

- Training: NVIDIA GPU with 16GB+ VRAM recommended
- Inference: CPU-only operation supported; 8GB RAM minimum

B Acknowledgments

This work was developed during the Caribbean Voices AI Hackathon organized by the UWI AI Innovation Centre. We thank the organizers for creating the competition and providing the BBC Caribbean speech corpus that motivated this research. We also thank Dr. Sikopo Nymabe-Galbraith for providing invaluable feedback on the research.

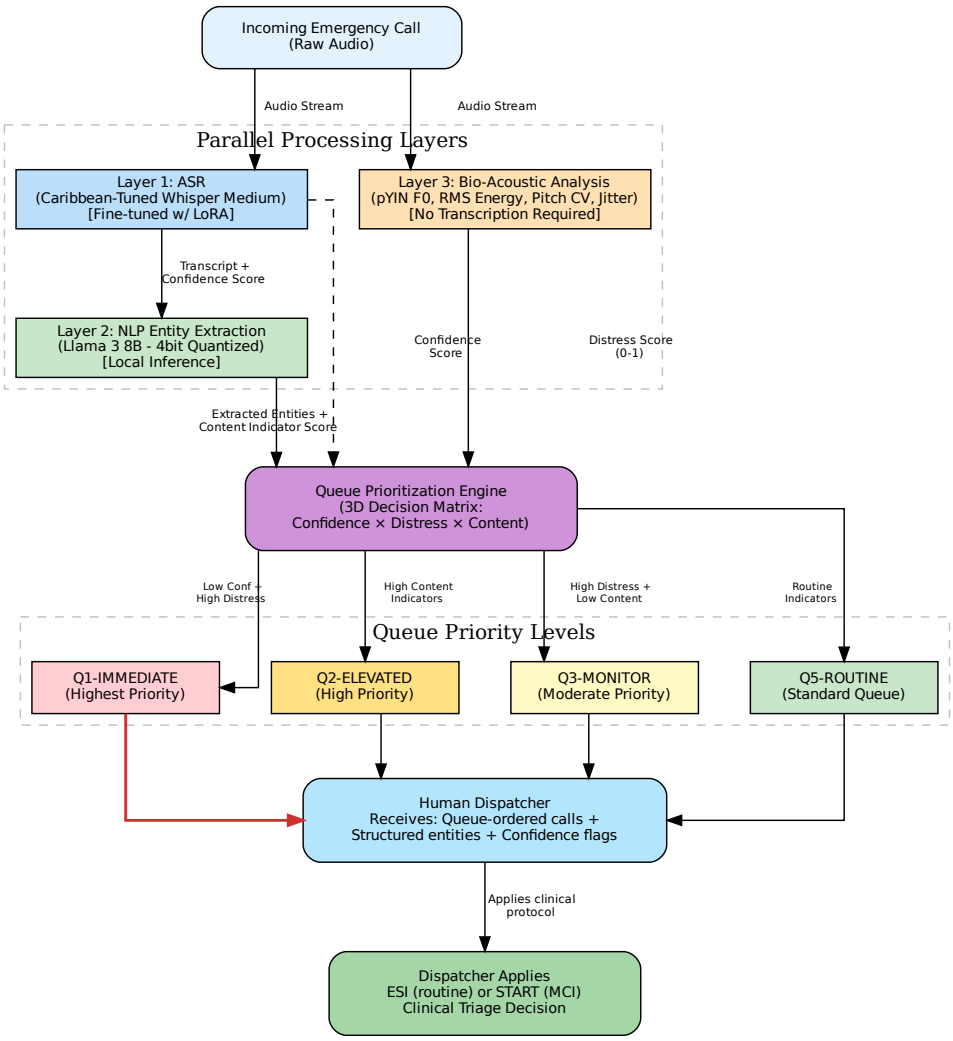

Figure 1: The TRIDENT architecture. The system processes raw audio through two parallel streams: (Left) A Caribbean-adapted ASR and NLP pipeline for entity extraction and content analysis, and (Right) a bio-acoustic analysis layer for detecting physiological distress markers. The **Queue Prioritization Engine** integrates three independent signals—transcription confidence, extracted clinical indicators, and vocal distress—to determine queue position for dispatcher attention. This ensures that (1) calls with low transcription confidence but high vocal distress receive immediate human review, and (2) semantically urgent calls from calm reporters are not delayed due to absent vocal stress markers. The dispatcher then applies established triage protocols (ESI for routine operations, START for mass casualty events) using both TRIDENT's extracted entities and direct audio review.

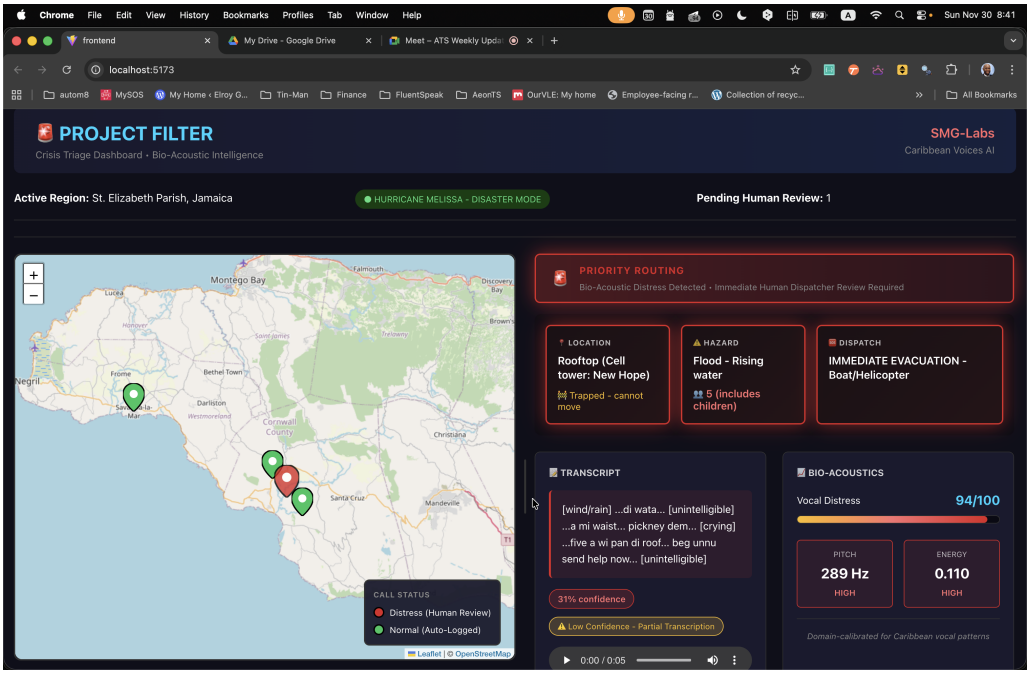

Figure 2: Dispatcher interface for a routine scenario (Q5-ROUTINE). High transcription confidence enables reliable entity extraction. The dispatcher can apply ESI protocol using the structured location, hazard type, and resource need data. Audio review is available but not flagged as necessary.

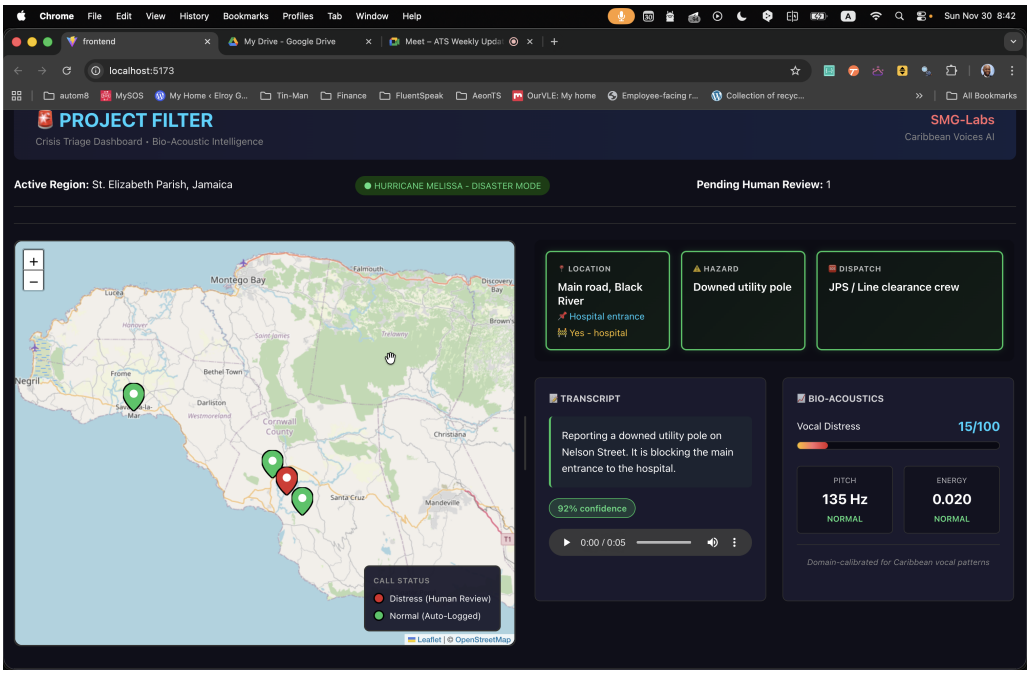

Figure 3: Dispatcher interface for a high-priority scenario (Q1-IMMEDIATE). Elevated distress markers combined with low transcription confidence trigger immediate queue placement. The interface prominently recommends audio review and displays partial entity extraction with uncertainty markers. The dispatcher will listen directly and apply ESI or START protocol based on their clinical assessment.