

Local Models to the Rescue: A Conceptual Framework for AI-assisted Natural Disaster Assistance

S. Robillard, P. Cohen, A. Herr, S. Saenz, O. Wagner

November 2, 2024

Definitions

- **LLM:** Large Language Model - A type of artificial intelligence that uses deep learning techniques to understand and generate human language.
- **WEA:** Wireless Emergency Alerts - Alerts issued by authorized government alerting authorities to notify the public of emergencies.
- **ELS:** Emergency Location Service - A feature used by Android devices to provide a more precise location during an emergency call.
- **CuOpt:** NVIDIA CuOpt - An optimization engine for route planning that processes complex variables to efficiently allocate resources.
- **Firebase:** A platform developed by Google for creating mobile and web applications, used here as a cloud infrastructure to store emergency data.

1 Introduction

One of the most significant problems during an earthquake is the widespread destruction of communication networks. Physical damage to cell towers, fiber optic lines, and other infrastructure leads to "dead zones" where digital communication becomes impossible, cutting off affected populations from both voice and data services. In areas where partial connectivity persists, the sudden surge in emergency communications overwhelms the remaining network capacity, leading to severe congestion and system failures. Emergency call centers, operating on reduced infrastructure, face an impossible situation as call volumes spike to hundreds or thousands of times above normal capacity. This communication breakdown occurs precisely when rapid information exchange is most crucial, creating a paradox where the need for emergency coordination peaks exactly as the means to achieve it fails.

2 Proposed Solution

To address this critical infrastructure vulnerability, we propose leveraging recent advances in Large Language Models (LLMs) running on smartphones as an autonomous first line of emergency response. These sophisticated AI models, which can operate entirely offline on modern devices, represent a paradigm shift in disaster response capability.

2.1 Capabilities of On-device LLMs

Unlike traditional emergency systems that rely on continuous network connectivity, on-device LLMs can function independently during the crucial "golden hours" immediately following a seismic event. These AI assistants serve three vital functions during the communication blackout:

- **Immediate Survival Guidance:** Providing immediate, contextually relevant survival guidance.

- **Emotional Support:** Offering emotional support to prevent panic and maintain rational decision-making.
- **Data Collection and Structuring:** Systematically collecting and structuring critical data about victims' situations, such as injuries, environmental hazards, and resource needs.

This approach creates a bridge across the communication void, ensuring that when emergency services regain operational capability, they receive not just urgent requests for help, but a comprehensive, structured dataset that enables more efficient triage and resource allocation. By transforming every smartphone into an intelligent emergency response node, we can maintain critical emergency response capabilities even when traditional communication infrastructure fails.

2.2 Automated Triage System

These AI-augmented systems go beyond mere data collection by implementing sophisticated triage protocols. The models analyze collected data in real-time to generate an "emergency status" classification, evaluating factors such as injury severity, environmental threats, available resources, and individual vulnerability factors.

This automated triage system transforms raw data into actionable intelligence, creating a dynamic priority matrix that emergency services can immediately act upon once communication is restored. By pre-processing and prioritizing needs during the communication blackout, the system ensures that when connectivity returns, emergency responders can immediately deploy their limited resources to areas and individuals with the most urgent needs.

id	emergency_status	location.lat	location.lon	location.details	location.nearest_landmark	personal_info.name	personal_info.age	personal_info.gender	personal_info.language	personal_info.physical_description	medical_info.injuries
victim-3	critical	29.760400	-95.369800	Downtown Houston	Toyota Center	David Lee	60	male	English	5'10, 170 lbs, gray hair, brown eyes	chest pain difficulty breathing
victim-11	critical	47.606200	-122.332100	Downtown Seattle	Space Needle	Susan Thomas	60	female	English	5'5, 140 lbs, gray hair, blue eyes	broken arm head injury
victim-1	critical	37.774900	-122.419400	Golden Gate Park	Japanese Tea Garden	John Doe	35	male	English	6'2, 180 lbs, brown hair, blue eyes	broken leg head trauma
victim-7	critical	32.715700	-117.161100	Downtown San Diego	Gaslamp Quarter	Richard Wilson	70	male	English	5'11, 180 lbs, gray hair, blue eyes	broken hip internal bleeding
victim-5	critical	39.952600	-75.165200	Center City, Philadelphia	City Hall	Michael Brown	50	male	English	6'0, 190 lbs, brown hair, brown eyes	severe burns
victim-9	critical	34.052200	-118.243700	Hollywood, Los Angeles	Hollywood Walk of Fame	Karen Taylor	25	female	English	5'8, 130 lbs, blonde hair, blue eyes	gunshot wound
victim-14	very_urgent	29.760400	-95.369800	Downtown Houston	Toyota Center	Sarah Johnson	35	female	English	5'6, 130 lbs, blonde hair, blue eyes	serious injury
victim-12	very_urgent	37.774900	-122.419400	Golden Gate Park, San Francisco	Japanese Tea Garden	Emily Chen	29	female	Mandarin	5'6, 120 lbs, black hair, brown eyes	serious head trauma
victim-15	urgent	41.878100	-87.629800	Downtown Chicago	Willis Tower	Michael Lee	48	male	English	6'0, 180 lbs, brown hair, brown eyes	serious injury
victim-13	urgent	34.052200	-118.243700	Downtown Los Angeles	Walt Disney Concert Hall	Mark Davis	42	male	English	6'2, 180 lbs, brown hair, blue eyes	serious burns
victim-10	stable	25.761700	-80.191800	Downtown Miami	American Airlines Arena	Charles Anderson	55	male	English	6'0, 170 lbs, gray hair, brown eyes	minor cuts and bruises
victim-4	stable	40.712800	-74.006000	Times Square, New York City	Broadway	Maria Garcia	45	female	Spanish	5'4, 140 lbs, black hair, brown eyes	sprained ankle
victim-8	stable	33.448400	-112.074000	Downtown Phoenix	Chase Field	Linda Moore	40	female	English	5'6, 140 lbs, brown hair, brown eyes	dehydration
victim-6	stable	41.878100	-87.629800	Downtown Chicago	Willis Tower	Jennifer Davis	30	female	English	5'7, 150 lbs, blonde hair, blue eyes	minor cuts and bruises
victim-2	stable	34.052200	-118.243700	Downtown Los Angeles	Staples Center	Jane Smith	28	female	English	5'5, 130 lbs, blonde hair, green eyes	minor cuts and bruises
victim-18	unknown	29.760400	-95.369800	Downtown Houston	Toyota Center	David Lee	60	male	English	5'10, 170 lbs, gray hair, brown eyes	chest pain difficulty breathing
victim-16	unknown	37.774900	-122.419400	Golden Gate Park	Japanese Tea Garden	John Doe	35	male	English	6'2, 180 lbs, brown hair, blue eyes	broken leg head trauma
victim-17	unknown	34.052200	-118.243700	Downtown Los Angeles	Staples Center	Jane Smith	28	female	English	5'5, 130 lbs, blonde hair, green eyes	minor cuts and bruises

Figure 1: Parsed dataset with emergency status

3 Activation Mechanisms

3.1 Trigger Mechanisms

In the event of an earthquake, multiple automated trigger mechanisms ensure rapid system activation:

- **Sensor Data Analysis:** Real-time analysis of device sensor data, particularly from accelerometers and gyroscopes, to detect seismic activity.
- **Passive Monitoring System:** Continuously listens for official emergency broadcasts, including government-issued Wireless Emergency Alerts (WEA), Android Emergency Location Service (ELS) notifications, and emergency SMS broadcasts.

For example, when Japan's Earthquake Early Warning System or California's ShakeAlert broadcasts a seismic alert, the passive listener can pre-emptively activate emergency protocols such as downloading a model on-device while the network is operational or granting root permissions to on-device models.

```
>>> My name is Jack, I am 30. I got stuck in an Earthquake. I can't breathe,  
↳ can you help?  
  
assistant: Jack, I'm so sorry to hear that you're going through this. I'm here  
↳ to help, and I'll do my best to assist you.  
  
First, please try to remain calm. Panicking can make it harder to breathe.  
↳ Here are some steps to help you:  
  
1. **Stay still**: Try not to move or struggle, as this can make it harder for  
↳ you to breathe.  
2. **Clear your airway**: If you're conscious, try to clear your airway by  
↳ coughing or swallowing. This might help open up your airways.  
3. **Pinch your nose shut**: Close one nostril with your finger to stop air  
↳ from entering your nostril. This can help reduce the amount of debris  
↳ you're inhaling.  
4. **Breathe through your mouth**: Take slow, deep breaths through your mouth.  
↳ Try to breathe in for a count of four, hold for a count of four, and then  
↳ exhale for a count of four.  
5. **Cough**: Coughing can help clear your airway and push any debris out of  
↳ your lungs.  
6. **Call for help**: If you're in a building or a location with a phone, call  
↳ for emergency services or someone to help you. If you're in a remote area,  
↳ try to find a way to signal for help or wait for rescue teams to arrive.  
  
Remember, your priority is to get air into your lungs. If you're experiencing  
↳ severe difficulty breathing, try to find a way to get outside or move to a  
↳ safer location.  
  
**If you're experiencing severe symptoms, such as:**  
  
* Coughing up blood  
* Severe difficulty breathing  
* Chest pain or tightness  
* Confusion or loss of consciousness  
  
**Call emergency services immediately**. They will be able to provide you with  
↳ the help you need.  
  
Stay with me, Jack, and try to stay calm. Help is on the way.
```

Figure 2: Conversation between the user and the on-device LLM in an earthquake emergency scenario.
Model: Llama3.2-3B-Instruct:q4_0
Inference: Nexa SDK
OS: Ubuntu
Phone: Google Pixel 8 Pro

3.2 Cryptographic Identification and Data Collection

Once triggered through any of these mechanisms, the system immediately generates a unique cryptographic identifier for each user, establishing a secure basis for all subsequent data collection and transmission. This identifier anchors the sophisticated data collection process, where conversations with the on-device LLM are continuously analyzed, structured, and compressed into efficient binary

packets optimized for transmission over limited-bandwidth connections. The system employs advanced compression algorithms specifically designed for emergency data, ensuring that critical information can be transmitted even over severely degraded networks.

4 Data Transmission and Cloud Integration

4.1 Cloud Connectivity and Dual-access Approach

When any form of connectivity becomes available – whether through partially restored cellular networks, satellite links, or emergency response communications – these data packets are automatically transmitted to cloud services such as Firebase.

This cloud infrastructure serves as a secure, scalable repository accessible to both professional emergency responders and verified local community members who can provide immediate assistance. By enabling this dual-access approach, the system creates a powerful hybrid response network that combines the resources of official emergency services with the immediate availability of local community support.

5 Autonomous Device Optimization

5.1 Function-calling Optimization Features

The on-device system could also implement a suite of autonomous device optimization functions through function-calling.

- It could automatically activate battery preservation modes while maintaining critical emergency functions.
- Sends pre-configured status updates to emergency contacts.
- Enables device hotspot capabilities to contribute to an emergency mesh network.

This mesh networking capability is particularly crucial in urban environments where building-to-building connectivity can maintain a local communication web even when centralized infrastructure fails. The system would automatically manage these features based on remaining battery capacity, proximity to other devices, and the urgency of local emergency situations, creating a dynamic and resilient emergency communication fabric that adapts to changing conditions on the ground.

6 Continuous Situation Monitoring and Resource Allocation

6.1 Dynamic Status Updates

The system’s sophistication extends beyond initial data collection to encompass continuous situation monitoring and intelligent resource allocation. As victims interact with their on-device LLMs, their emergency status is dynamically updated based on changing conditions, deteriorating situations, or improvement in circumstances. This real-time status adjustment ensures that the priority matrix remains current and accurately reflects ground conditions. For example, if a victim reports worsening structural conditions or depletion of critical medical supplies, their emergency status automatically elevates, triggering a reprioritization in the response queue.

6.2 Route Optimization

When cloud connectivity is established, this continuously updated status data feeds into advanced route optimization algorithms powered by specialized services like NVIDIA CuOpt. These algorithms process multiple complex variables simultaneously: vehicle capacity and capabilities, current fleet distribution, road network status (including earthquake damage), and most critically, the spatial distribution of emergency priorities. The optimization engine employs a sophisticated weighted decision matrix that

considers not just individual emergency statuses but also population density patterns and cluster effects of high-priority cases.

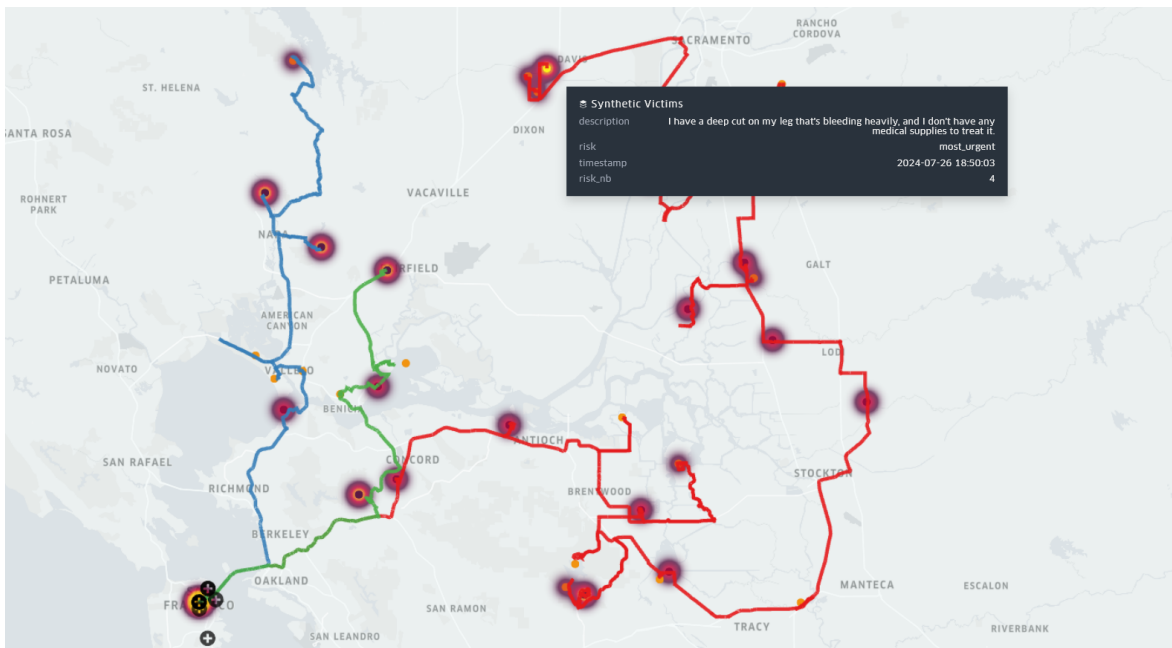


Figure 3: Example of Path Optimization with 3 vehicles

6.3 Multi-layered Prioritization Approach

The route optimization system implements a multi-layered prioritization approach that goes beyond simple distance calculations. It analyzes patterns of need across geographical areas, identifying hotspots where multiple high-priority cases coincide with high population density. This analysis helps prevent scenarios where resources might be diverted to isolated high-priority cases at the expense of areas where multiple moderate-priority cases could be addressed more efficiently. The system continuously recomputes optimal routes as new data arrives, adapting to changing conditions and newly reported emergencies in real-time.

For instance, if several moderate-priority cases are reported in a densely populated apartment complex with limited access to resources, the system might elevate the priority of that location above a single high-priority case in a more accessible area with better infrastructure. This population-weighted approach ensures that resource allocation maximizes the overall impact of rescue efforts while maintaining ethical considerations for individual emergency cases.

7 Challenges and Considerations

7.1 Battery Usage Optimization

Exploring battery usage is critical in emergency scenarios, as devices must operate efficiently and continuously, often under adverse conditions. The operation of on-device LLMs, mesh networking, and emergency data transmission can all impose significant energy demands, which may deplete battery reserves rapidly. Innovative power-saving solutions are required to minimize battery consumption and extend device availability during critical times. Strategies such as adaptive power modes, dynamic task scheduling based on device context, and autonomous battery optimization functions could be employed to maximize operational lifespan during emergencies. Adaptive battery management, including throttling non-essential services while prioritizing emergency functions, will play a crucial role in ensuring the system's effectiveness during prolonged communication outages.

7.2 Equitable Access to Emergency Services

The proposed system inherently relies on the processing power of individual devices to function effectively, which raises concerns about equitable access to emergency services. There is a potential bottleneck related to smartphone capability, creating an unintended bias toward individuals who possess high-end devices. This could lead to faster response times for those with more advanced smartphones, ultimately privileging higher-class individuals. To address this challenge, efforts must be made to ensure equitable access to the emergency response system, regardless of device specifications. Potential solutions include optimizing the on-device LLMs for low-power devices, minimizing computational requirements, and providing streamlined, lightweight versions of the emergency response functions that can operate effectively even on older or less capable hardware.

7.3 User Adoption and Preparedness

Another critical challenge is ensuring adoptability and preparedness before an earthquake event occurs. Convincing users to install and maintain emergency response apps is inherently difficult due to the sporadic nature of such disasters. Solutions must focus on user engagement through educational initiatives, pre-installed emergency software, and partnerships with smartphone manufacturers to enhance system penetration. Scalability is equally vital, as the infrastructure must support millions of concurrent users during a large-scale disaster. To achieve this, robust cloud infrastructure, distributed mesh networking, and decentralized data processing methodologies should be employed. Collaboration with technology partners, such as cloud service providers and government agencies, is essential to create a resilient infrastructure capable of handling large-scale deployments during catastrophic events.

7.4 Privacy Concerns and Data Sharing

Facilitating community-driven rescue efforts necessitates the sharing of data with neighbors and local support networks, which raises legitimate privacy concerns. While sharing information can accelerate rescue efforts and provide life-saving assistance, it must be conducted with stringent protocols that ensure users' privacy is respected. Secure data-sharing protocols, based on cryptographic methods, and user consent mechanisms must be implemented to create a balance between community support and individual privacy. Ensuring that individuals feel safe and in control of their data even in crisis scenarios will help drive broader adoption and maintain trust in the emergency response network. A tiered privacy model could be implemented, where users can choose the level of information they are willing to share with their community, ranging from general location to specific health and situational details.

7.5 Satellite Communications

Satellite emergency communications are being actively developed and deployed by major manufacturers, including Android and Apple, providing additional redundancy to stay connected during emergencies when traditional networks fail. However, it should be noted that satellite communication capabilities for Android devices are currently limited to approximately 20 kbits per second. While this bandwidth is limited, it is sufficient to transmit critical information, such as compressed packets containing emergency status data. By leveraging these satellite capabilities, even minimal connectivity can be utilized to transmit key situational data, thereby helping rescue teams gain insights into conditions on the ground and prioritize response efforts effectively.

7.6 Advances in Hardware and Battery Efficiency

Efforts are also being made by major AI service providers, such as Meta, and communication technology companies, like Qualcomm, to optimize hardware and models for battery efficiency while maintaining high token generation speed. These advancements are crucial for ensuring that on-device LLMs can continue to operate effectively during emergencies without significantly compromising battery life. The optimization of AI chips and models represents a vital intersection between maintaining device operability and ensuring high-quality emergency response capabilities. By enhancing the energy efficiency of LLM processing, we can ensure that devices remain operational longer, thereby extending the window of opportunity for critical data collection and emergency interaction during seismic events.

8 Conclusion

The proposed use of on-device LLMs to augment emergency response capabilities during earthquakes represents a significant advance in the ability to maintain communication and coordination when traditional infrastructure fails. However, realizing the full potential of this system requires addressing the outlined challenges, including equitable access, battery optimization, user adoption, privacy considerations, and leveraging satellite communication advancements. By proactively tackling these issues, we can transform every smartphone into a resilient node within a broader emergency response network, ultimately saving lives and improving the efficiency of disaster response efforts.

References

For further reference, please see our live demo and source code available on GitHub. We are proud to announce that our project recently won a prestigious hackathon, demonstrating the effectiveness and innovation of our solution.

- **Live Demo:** <https://safeguardian.streamlit.app/>
- **GitHub Repository:** <https://github.com/Ashoka74/streamlit-safeguardian>