When Every Second Counts, AI Must Work Offline.

FirstAidQA: A Synthetic Dataset for First Aid and Emergency Response in Low-Connectivity Settings

Mushfiqur Rahman Mushfique Saiyma Sittul Muna Rezwan Islam Salvi Ajwad Abrar

Islamic University of Technology (IUT), Dhaka, Bangladesh

NeurIPS 2025: 5th Muslims in ML Workshop — Paper ID 42

The AI Lifeline Is Broken in Low-Connectivity Zones.

In emergencies—disaster zones, rural clinics, remote regions—access to reliable information can save lives.

Modern Large Language Models (LLMs) have immense potential but are dependent on powerful hardware and stable internet, making them impractical when they are needed most.

First responders and civilians in these settings are often equipped with low-tier devices and no reliable connection, creating a critical information gap.

Existing Medical Datasets Are Built for Clinics, Not Crises

The primary barrier to developing lightweight, offline models is the lack of a suitable dataset.

Current QA Benchmarks

(e.g., BioASQ, MedQA, PubMedQA)

Clinical diagnostics, academic biomedical literature.

Mirrors academic exams.

Medical professionals with formal training.

The Unmet Need for First Aid

(The FirstAidQA Gap)

Practical, instructional, and situational knowledge.

Step-by-step, actionable guidance.

Laypersons, bystanders, and first responders with little to no formal training.

We Created FirstAidQA to Fill This Critical Gap

We introduce the first synthetic Question-Answer dataset specifically tailored to first aid and emergency response, designed to enable the next generation of offline AI assistants.

5,500

High-quality question-answer pairs

Source

Generated from the certified *Vital First Aid Book (2019)*.

Goal

To support instruction-tuning and fine-tuning of Small Language Models (SLMs) for on-device deployment

The Dataset is Grounded in a Certified, Comprehensive First Aid Manual

VITAL FIRST AID

AMERICAN RED CROSS

ILCOR
INTERNATIONAL LIAISON COMMITTEE ON RESUSCITATION

- The source text is the *Vital First Aid Book (2019)*, chosen for its structured and comprehensive coverage of emergency care.
- The manual's content is consistent with international standards from organisations like the American Red Cross and ILCOR.
- Content was carefully segmented to preserve context for generating realistic QA pairs on topics like casualty movement, head injuries, and CPR.

The Dataset's Taxonomy Ensures Comprehensive Coverage of Emergency Scenarios.

General Emergency Procedures

Preparedness, priorities, DRSABCD protocol.

CPR

Methods for adults, children, infants; special cases like drowning.

Road Traffic Accidents

Scene safety, casualty assessment.

Moving Casualties

Relocation methods, spinal precautions.

First Aid Equipment

Use of kits, dressings, splints.

Patient Examination

Vital signs, injury severity.

Specific Medical Conditions

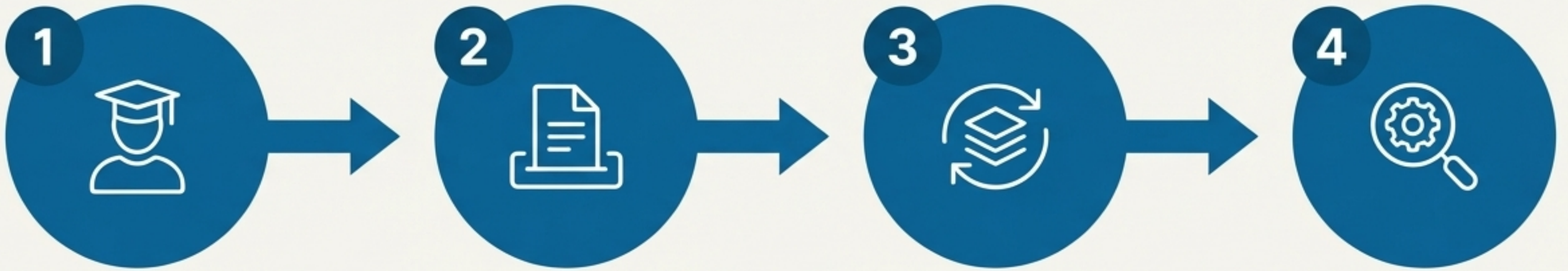
Bleeding, burns, fractures, bites, poisoning.

Neck & Spinal Injuries

Assessment and stabilisation.

A Structured Prompting Strategy Was Used to Generate High-Quality, Relevant Data.

We used ChatGPT-4o-mini to generate the QA pairs through a carefully designed, context-bound process. The output was required in JSON format for seamless integration into machine learning pipelines.



Role Framing

The LLM was instructed to act as a "renowned expert in synthetic dataset creation for first aid."

Context Injection

Each prompt included a specific text chunk from the source manual (e.g., guidance on treating burns).

Iterative Generation

The base prompt generated 20 diverse QA pairs. A follow-up instruction ("Generate 20 more...") was used iteratively.

Refinement

Prompts were adjusted when necessary to improve diversity (e.g., explicitly requesting paediatric or elderly cases).

The Core Prompt Was Engineered for Medical Accuracy and Situational Diversity.

Encourages medically precise, high-quality answers.

Constrains the model to the source text, reducing hallucination.

```
Imagine you are a renowned expert in synthetic dataset creation
for first aid and medical emergencies. Your task is to generate 20
diverse  question-answer pairs from the given corpus of a
certified first aid manual. Answers must be detailed, medically
accurate, and reflect multiple perspectives and scenarios.
Provide the output in JSON format.
```

Ensures questions reflect diverse roles (bystander, responder, lone rescuer).

Enables seamless pipeline integration.

A Multi-Stage Quality Assurance Process Ensures Accuracy, Safety, and Relevance.

1. Strategic Filtering

- Text chunks from the book were manually evaluated.
- Only content that could generate actionable, real-world QA pairs was selected.
- Less applicable theory and background information were excluded.

2. Hallucination and Bias Mitigation

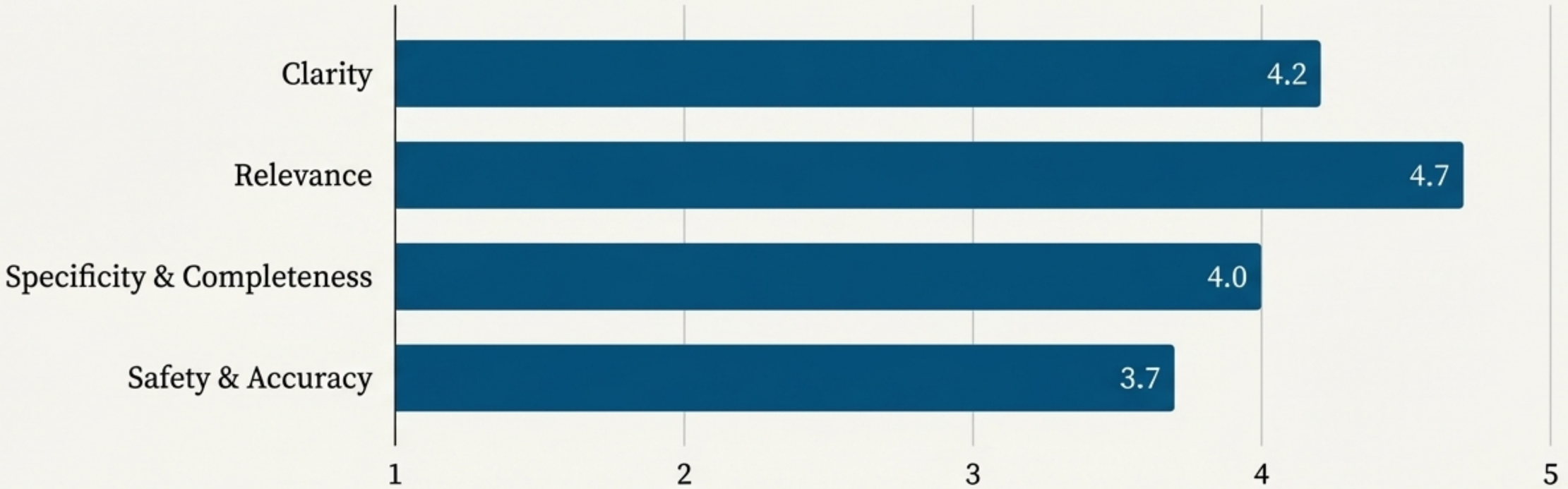
- The LLM was explicitly instructed to generate answers *only* from the provided text chunks.
- A diverse selection of chunks was used to capture a wide range of situations across the emergency response domain.

3. Manual Expert Review

- A random sample of 200 QA pairs was evaluated by 3 independent medical professionals.

Expert Validation Confirms High Levels of Clarity and Relevance

Three medical professionals scored a random sample of 200 QA pairs on a 1-5 scale across four key criteria.



Key Takeaway: The dataset scored exceptionally high on relevance, confirming its direct applicability to first aid scenarios. The lower (but still strong) safety score highlights the need for cautious handling and further refinement.

Responsible Use Requires Acknowledging the Dataset's Limitations and Safety Flags.

Core Limitations

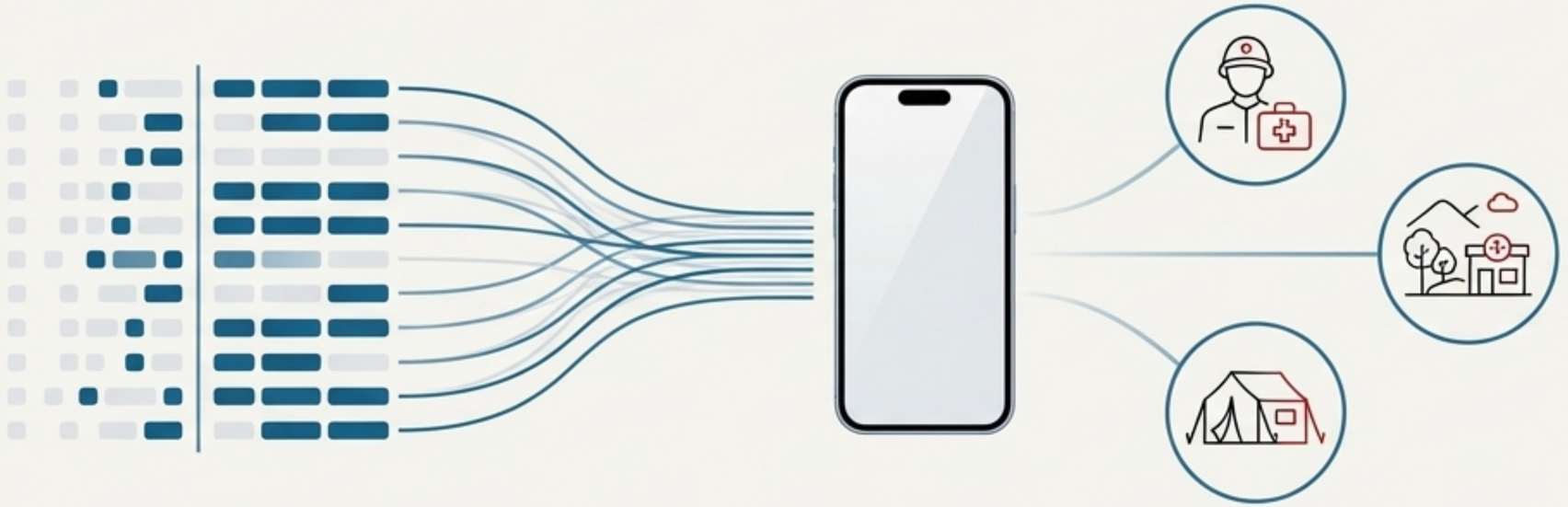
- **Not a Medical Substitute:** Designed to support, not replace, professional medical care.
- **Emergency-Only Focus:** Not suitable for general, non-urgent health advice.
- **Synthetic Generation:** May not cover all edge cases and requires careful handling.

Examples of Expert-Flagged Instructions

Topic	Issue
Tick Removal	Recommendation to use a freezing product conflicts with standard guidance (direct removal with
Marine Stings	Recommending vinegar for stingrays, when hot water is the correct treatment.
Choking (CPR)	Mentions blind finger sweeps, which are not advised as they can push objects deeper.

**Users should handle flagged topics with caution. The full list of flagged QA pairs is provided in the paper's appendix.*

FirstAidQA Unlocks the Potential for Fast, Reliable, On-Device Medical Guidance



This dataset is a foundational tool for building the next generation of AI for emergency response. It directly enables:

- **Fine-tuning of SLMs**: Creating small, efficient models that can run on low-tier devices without an internet connection.
- **Instruction Tuning**: Developing models that can provide clear, step-by-step procedural guidance in stressful situations.
- **Offline-Capable Systems**: Empowering first responders and civilians in rural areas, disaster zones, and other resource-constrained environments with life-saving information.

We Invite the Research Community to Build the Future of Emergency Response AI

FirstAidQA is publicly available to advance research on safety-critical and resource-constrained AI. We envision a world where this work leads to models that empower individuals with real-time medical consultation, saving lives where expert help cannot reach instantly.

Explore and download the dataset on Hugging Face

huggingface.co/datasets/i-am-mushfiq/FirstAidQA