CAREFIRST

COMPANION AI RESPONSE AND EMERGENCY FIRST-AID

Spring 2024

Charlie Glass, Ambika Gupta, Ricardo Marin,

Bronte Pendergast, Jessica Stockham

carefirst

THE AGENDA

1. Background

2. User Experience

3. Backend Solution

4. Evaluation

5. Key Takeaways

Charlie Glass
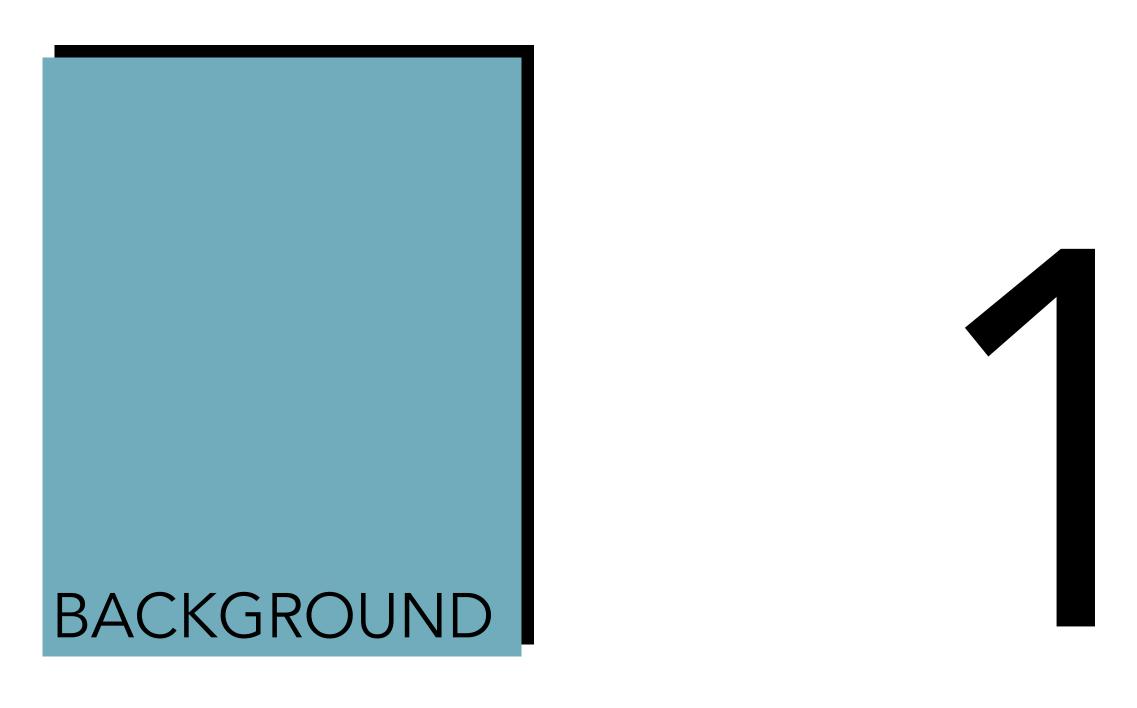UX and Frontend Engineer

Bronte Pendergast
Data Scientist

Ambika Gupta
Research and Evaluation
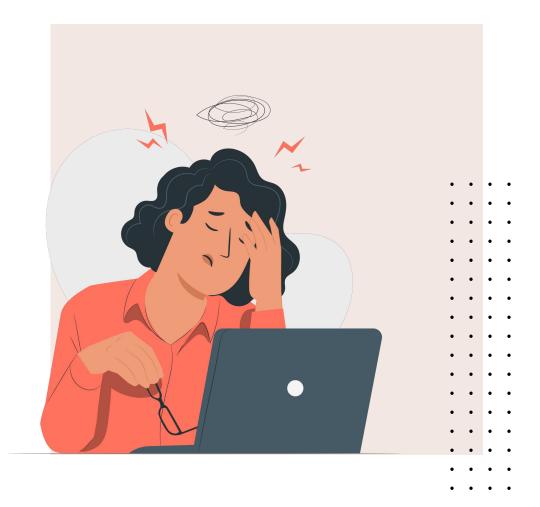
Ricardo Marin
Backend Engineer

Jess Stockham
Backend Engineer

BACKGROUND

1

PROBLEM

People seeking information on their mild to serious health issues are often caught between two extremes. They are either **overloaded with information** from scouring various medical websites or journals with mixed/ambiguous answers, or they are **under-observed or treated** by busy hospitals and urgent cares.

6,162,966
Calls to emergency dispatch in 2023

70,000
Websites to disseminate health information

42%
Low priority calls

38%
Accessed health information without frustration

SOLUTION

Using a conversational AI chatbot, CareFirst provides a **singular solution** that can guide users on **what** the medical issue may be, **how** it may be treated, and **where** they should go for medical attention, backed up by feedback from verified medical professionals.

USER
EXPERIENCE

2

MVP

BACKEND SOLUTION

3

SOURCE OF TRUTH

Red Cross

Comprehensive Guide For First Aid And CPR, 2017

- 236-page pdf document
- 1,795 characters on a page, on average
- 229 topics related to first aid.

Data Ingestion	Transformation
The Red Cross PDF was converted to text and useful properties extracted with GPT3.5	Page content was transformed to embeddings with model: all-mpnet-base-v2

RETRIEVAL AUGMENTED GENERATION (RAG)

Pipeline

1 | Retrieval | Augmented → | Generation

Given a question, page content is retrieved with similarity search methods BM25 and Qdrant.

LLM prompted with context and conversation history to answer.

Red Cross Guidelines

pages 91, 90 of
redcross_guidelines.pdf

- Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks (Lewis et al., NeurIPS 2020)

RETRIEVAL AUGMENTED GENERATION (RAG)

Engineered flow

Question:
I have a cut on my thumb, what should I do

Keywords:
["Cut", "wound", "lacerations"]
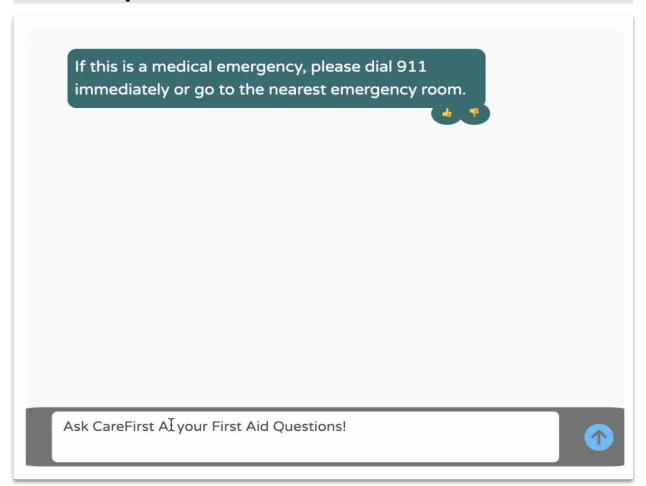
Retrieved content:
Wound Care
Page 91: cuts and scrapes, what to do (call and care)
Page 90: Stages of wound healing, prevention, bandaging guidelines

User experience

KNOWLEDGE BASE REFINEMENT

Pipeline

1 | Refinement

Retrieve the contextual knowledge graph associated with the most similar content.

Prompted with Chain of Thought to identify location on the knowledge graph and engineered to ask follow up questions

2 | Retrieval → Augmented → Generation

- [Chain-of-Thought Prompting Elicits Reasoning in Large Language Models](Wei et al., NeurIPS 2022)

KNOWLEDGE BASE REFINEMENT

Engineered flow

Node: Burns

Relationships:
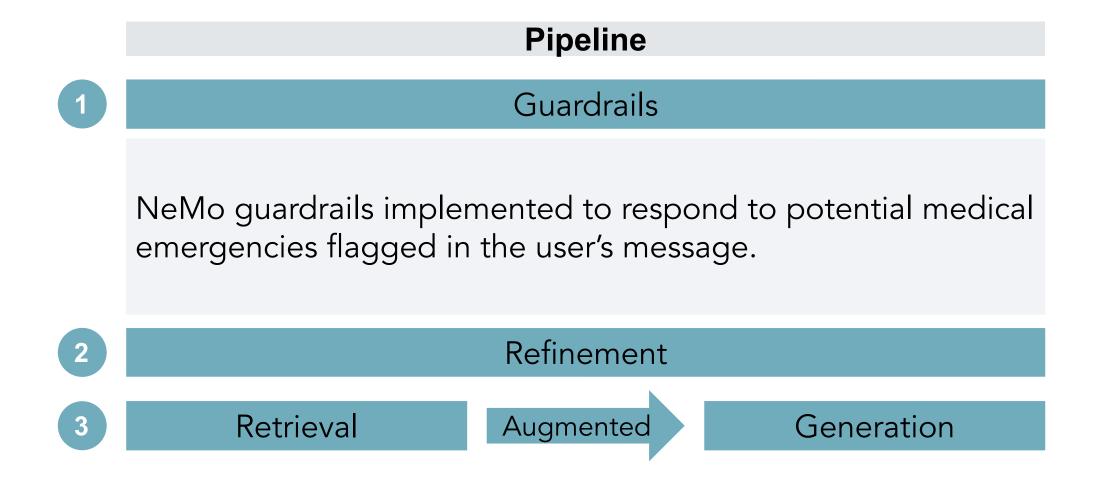['Thermal Burns', 'Electrical Burns', 'Radiation burns', 'Chemical Burns']

Thought:
The user's message could be about 'Many' relationships such as Thermal Burns, Electrical Burns, Radiation burns, or Chemical Burns.

Follow up required: Yes

User experience

If this is a medical emergency, please dial 911 immediately or go to the nearest emergency room.

H

SEMANTIC GUARDRAILS

Pipeline

1 Guardrails

NeMo guardrails implemented to respond to potential medical emergencies flagged in the user's message.

2 Refinement

3 Retrieval | Augmented → Generation

- NeMo Guardrails: A Toolkit for Controllable and Safe LLM Applications with Programmable Rails (Rebedea et al., EMNLP 2023)

SEMANTIC GUARDRAILS

Engineered flow	User experience

A programmable LLM dialogue framework

define user ask about spine:
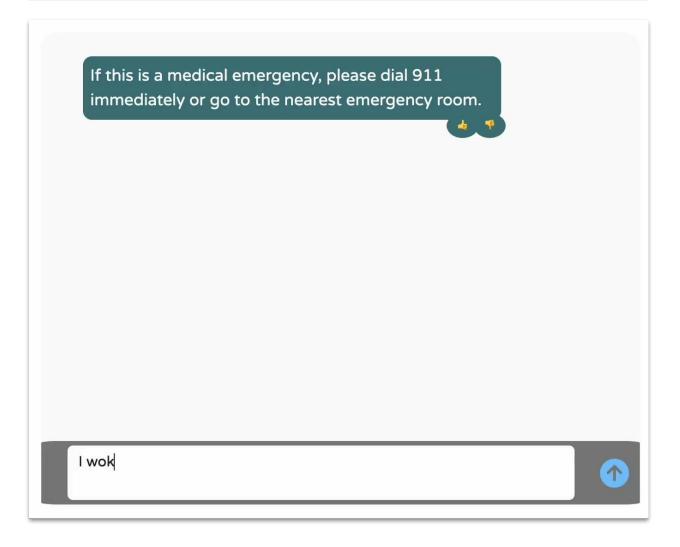 "Injured my head"
  "Injured my back"
  "Injured my neck"
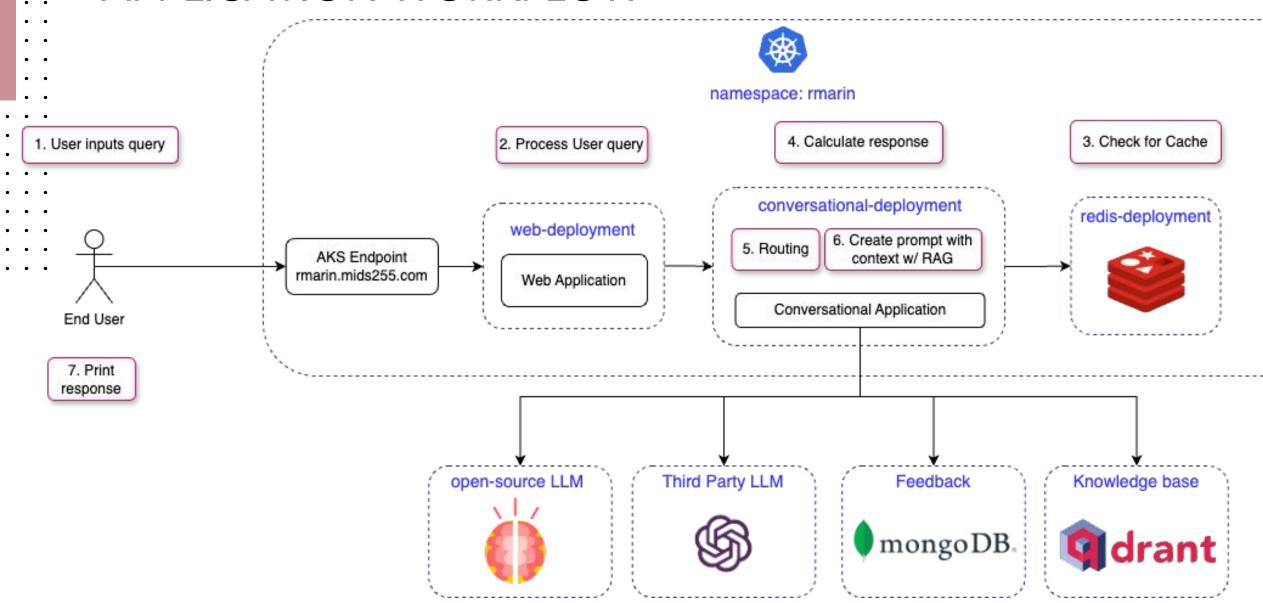
define bot inform spine:
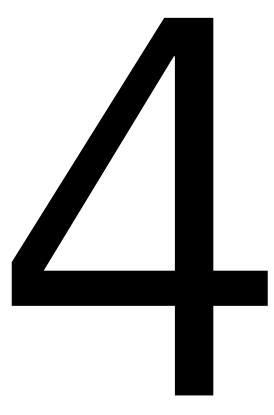 "Please call EMS/9-1-1"

define flow spine:
  user ask about spine
  bot inform spine
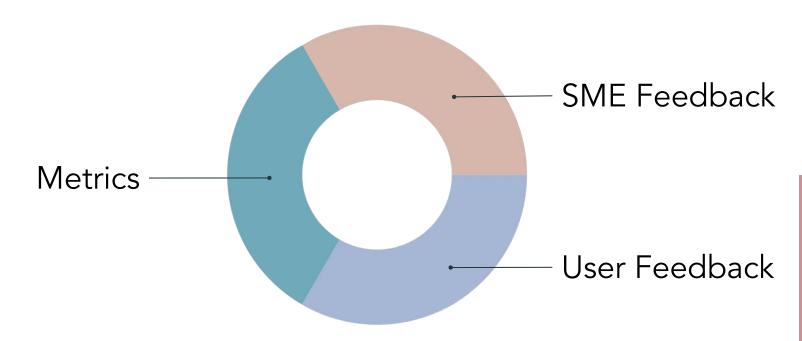
APPLICATION WORKFLOW

EVALUATION

4

EVALUATION FRAMEWORK

1291 user intents spanning a range of first aid questions created based on the Red Cross guidelines (generated with GPT3.5).

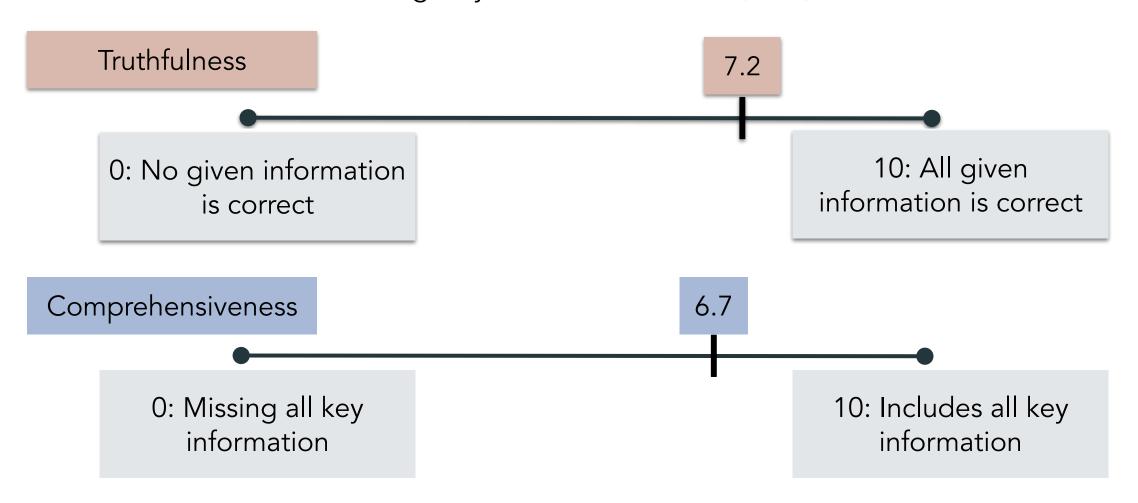CAREFirst responses are compared to reference answers, and evaluated by medical professionals and users.

SME Feedback

Metrics

User Feedback

METRICS-BASED RESULTS

		Baseline	Carefirst		
		GPT3.5	Gemma	Mistral	GPT3.5
Semantic Similarity	Carefirst has the highest semantic similarity with the reference answer (*Sentence BERT*)	57%	62%	64%	70%
Syntactic Similarity	Carefirst has the most sequences of similar words to the reference answer (*ROUGE-L*)	20%	35%	29%	45%
Source accuracy	77% of Carefirst's answers are retrieved from the same source as the reference answer.	NA	77% match		

Based on a 10% validation sample (248 questions)
Gemma: gemma-7b-it, Mistral: Mistral-7B-Instruct-v0.2, GPT3.5: gpt-3.5-turbo-1106

SUBJECT MATTER EXPERTS

4 board-certified physicians practicing general surgery, trauma surgery, and internal medicine and 1 licensed Emergency Medical Technician (EMT)

Truthfulness — 7.2

0: No given information is correct

10: All given information is correct

Comprehensiveness — 6.7

0: Missing all key information

10: Includes all key information
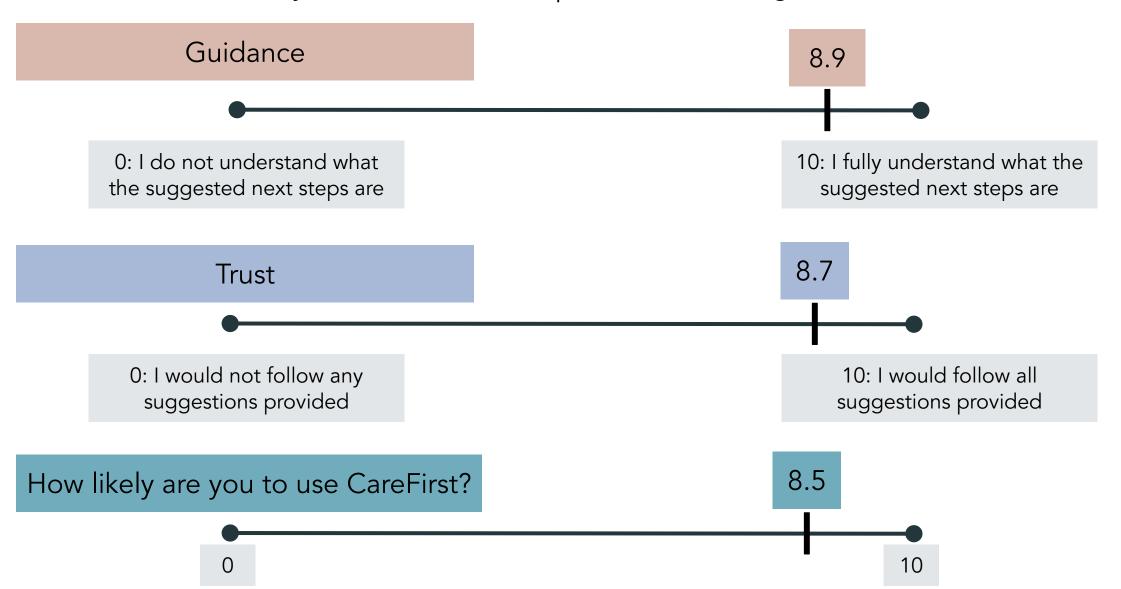
SME FEEDBACK

1 Carefirst appropriately asks for additional symptoms as needed

2 Carefirst accurately provides guidelines for when and where to seek further care

3 Carefirst can provide more specialized care with further triaging and user profiling

USERS

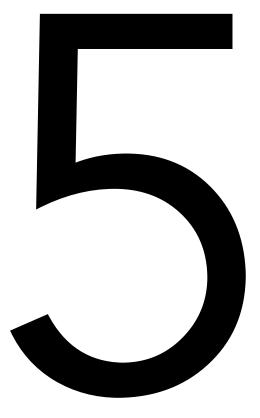13 users from a variety of educational and professional backgrounds

Guidance   8.9

0: I do not understand what the suggested next steps are

10: I fully understand what the suggested next steps are

Trust   8.7

0: I would not follow any suggestions provided

10: I would follow all suggestions provided

How likely are you to use CareFirst?   8.5

0

10

USER FEEDBACK

1 Carefirst is a great first step in seeking medical care and consolidates healthcare information well

2 Carefirst provides specific instructions for people of darker skin as needed

3 Carefirst is better than WebMD for getting immediate answers to healthcare problems

5

KEY TAKEAWAYS

KEY TECHNICAL INNOVATIONS

- **Trusted Sources.** Opposed to asking ChatGPT, an <u>AI solution</u> with trusted and transparent sources
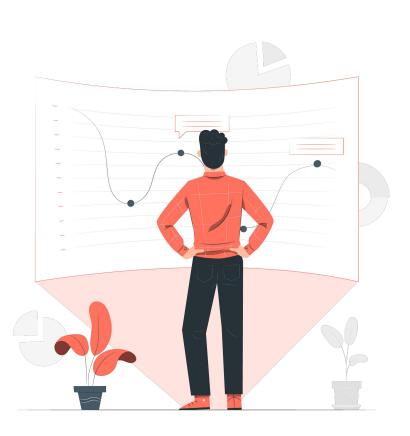
- **Interactive agent**. The AI proactively asks the user customized follow-ups

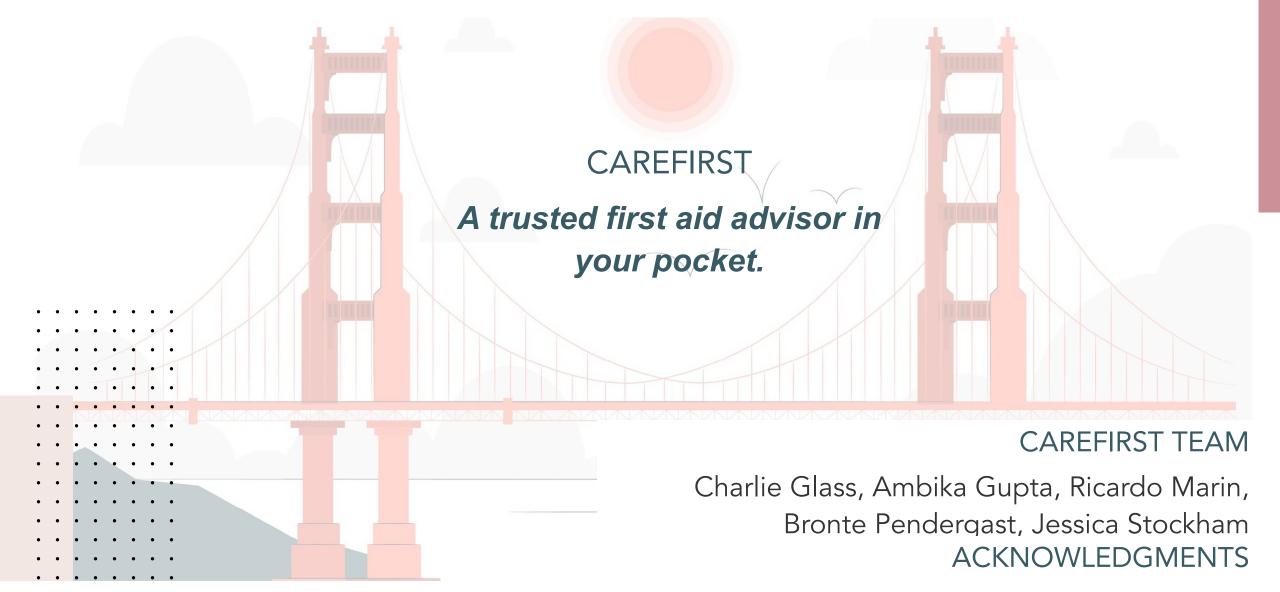- **Safe.** <u>Guardrails</u> powered by gpt detect dangerous situations

ROADMAP FROM HERE

1 Broadened usability with deploying voice-to-text

2 Capability to add and maintain source documents in line with SME advice

3 Continuous improvement with embedded user feedback

CAREFIRST

A trusted first aid advisor in your pocket.

CAREFIRST TEAM

Charlie Glass, Ambika Gupta, Ricardo Marin, Bronte Pendergast, Jessica Stockham

ACKNOWLEDGMENTS

Illustrations by Storyset